

PAPER

The Use of ChatGPT in Task-Based ESP Learning at University: Does it Make a Difference?

Iryna Lytovchenko¹(✉),
Yuliana Lavrysh¹, Oksana
Synekop¹, Valentyna
Lukianenko¹, Oksana
Chugai¹, Iryna Shastko²

¹National Technical University
of Ukraine "Igor Sikorsky
Kyiv Polytechnic Institute",
Kyiv, Ukraine

²Ivano-Frankivsk National
Technical University of Oil and
Gas, Ivano-Frankivsk, Ukraine

irinalyt@ukr.net

ABSTRACT

This study investigates the integration of ChatGPT into task-based learning (TBL) environments for English for specific purposes (ESP) at a technical university in Ukraine. Conducted with 27 bachelor students, the research employs a mixed-methods approach to evaluate the effectiveness of ChatGPT in enhancing students' writing task performance (creating a sustainable campus initiative proposal). The use of a quantitative research method permitted comparing the quality of students' writing between two TBL groups: the one that used ChatGPT to complete the tasks and the one that did not use this tool. The qualitative method was used to survey students and explore their attitudes towards using ChatGPT to complete the tasks. Results indicate that students using ChatGPT showed the same level of achievement as those who did not use it in generating ideas and structuring initial drafts but struggled with the deeper aspects of writing, such as communicative achievement and creativity. Also, the ChatGPT's ability to enhance language accuracy was evident, yet not superior to traditional methods. Students generally perceive ChatGPT positively, appreciating its role in facilitating the writing process and providing immediate assistance. However, they also acknowledge the necessity for critical evaluation and human refinement of artificial intelligence (AI)-generated texts to ensure the quality and originality of their work. This study concludes that while AI tools, such as ChatGPT, can complement traditional teaching methods, they cannot replace human evaluation and expertise.

KEYWORDS

English for specific purposes (ESP), task-based learning, artificial intelligence (AI), ChatGPT, technical university, writing skill

1 INTRODUCTION

Task-based learning (TBL) is an innovative approach in language education based on learning through tasks that are similar to real-world situations. This approach has had an enormous influence on education since the 1980s, when it "inspired a generation of language teachers seeking to engage productively with Communicative

Lytovchenko, I., Lavrysh, Y., Synekop, O., Lukianenko, V., Chugai, O., Shastko, I. (2025). The Use of ChatGPT in Task-Based ESP Learning at University: Does it Make a Difference? *International Journal of Interactive Mobile Technologies (IJIM)*, 19(2), pp. 4–22. <https://doi.org/10.3991/ijim.v19i02.51115>

Article submitted 2024-07-13. Revision uploaded 2024-11-01. Final acceptance 2024-11-01.

© 2025 by the authors of this article. Published under CC-BY.

Language Teaching” [1, p. xi]. Since that time, it has been representing a new look at language teaching methodology, assessment, and syllabus design. As the popularity of TBL is rising, there is a growing number of theoretical research that investigate its conceptual foundations and classroom-based studies in which teachers examine the quality and quantity of learners’ interaction in the process of task completion and explore better conditions for learning through the use of tasks. However, as pointed out by Willis and Willis [2], it is important that teachers expand their studies and test the principles and procedures of TBL through critical observation (often self-observation) in classroom practice.

In recent years, integrating artificial intelligence (AI) into education has opened new possibilities for enhancing traditional teaching methods. ChatGPT represents one such innovation, offering students immediate feedback and assistance during language learning tasks. However, despite the promise of AI in education, its impact on TBL, particularly in English for specific purposes (ESP) courses, remains underexplored. In technical universities, where ESP is essential for engineering students, understanding how AI tools such as ChatGPT influence student outcomes is especially critical.

Given this gap in the study, we studied new possibilities for TBL practices that have emerged with the integration of AI into this approach. The research questions we sought to address were: 1) Does ChatGPT enhance the writing performance of ESP students in a TBL environment? 2) How do students perceive the role of ChatGPT in completing academic writing tasks? By investigating these questions, we aim to contribute to the ongoing conversation about AI’s role in language education and to provide findings about its potential advantages and limitations when integrated into TBL for ESP learners.

2 RELATED WORKS

This review explores the foundational theories and recent studies on TBL and its integration into ESP instruction, focusing on how AI, specifically ChatGPT, is beginning to influence language learning and writing tasks. The review highlights key contributions to TBL and its effectiveness in language acquisition, engagement, and task completion. It further examines the limited but emerging research on AI’s role in ESP learning and writing tasks, pointing to the gap in understanding how AI tools such as ChatGPT can be systematically integrated into TBL frameworks to enhance language skills.

2.1 Task-based learning: origins and core concepts

Task-based learning originates from the communicative approach and is focused on language learning through authentic communication in the process of completing tasks [1]. A pedagogical task, as defined by Nunan [3] is “a piece of classroom work that involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is focused on mobilizing their grammatical knowledge in order to express meaning, and in which the intention is to convey meaning rather than to manipulate form” (p. 4). Willis and Willis [2] point out that, contrary to form-based approaches, TBL “involves the specification not of a sequence of language items, but of a sequence of communicative tasks to be carried out in the target language” (p. 173). At the heart of a communicative task is

the exchange of meanings. Therefore, in the TBL approach, the language development is realized through the language use. Studying the language form is secondary. However, it is not neglected [1]. As students work with meaningful content, they naturally take the time to decide which sentence structures to use to express their messages best [4].

The framework of the task cycle typically consists of three stages: the pre-task stage, the main task stage, and the post-task stage. Of these three, only the main task stage is mandatory. However, pre-task and post-task activities are also critical for the successful completion of the task because they help learners to prepare for the main task (pre-task stage) and reflect on their task performance or do focus-on-form activities (post-task stage) [1], [2], [5].

2.2 Classifications and characteristics of tasks in task-based learning

Different authors suggest different classifications of tasks. For example, Prabhu [6] classifies tasks into three broad types that require: (1) Information gap filling (transferring information from one person to another); (2) Reasoning gap filling (deriving new information by means of inference, deduction, practical argumentation, comprehension of relationships/patterns); and (3) Opinion gap filling (determination and articulation of a personal preference, feeling, attitude in response to a given situation) (pp. 46–47). Ellis [7] categorizes tasks as focused and unfocused. The design of the unfocused tasks implies that learners should communicate generally. Focused tasks suggest that the learners should use particular linguistic forms in communication. Willis's [8] classification includes six types of tasks: listing, ordering and sorting, comparing, problem-solving, sharing personal experiences, and creative tasks. These classifications highlight the versatility of TBL, allowing it to address various aspects of communication and language use.

According to Prabhu [6], a task is successful if it has the following characteristics: (1) a clear outcome, (2) clear criteria for success, (3) a balance of predictability and unpredictability, (4) content that allows learners to draw on their background knowledge, and (5) content that generates interest and sustained engagement. These attributes ensure that learners remain motivated and are challenged appropriately, contributing to linguistic and cognitive development.

2.3 Effectiveness of TBL in language learning and ESP instruction

Research into the use of TBL in language teaching has demonstrated the efficacy of this approach in language acquisition, fluency development, and learner engagement. TBL has been found efficient in promoting both linguistic and pragmatic competencies, as learners are given opportunities to use authentic language and discuss meaning in communicative tasks [1]. Also, it has been reported that TBL promotes motivation and confidence in learners as they successfully complete the task [3]. These findings underscore TBL's value in promoting active, communicative learning, which aligns with modern educational goals.

In the context of our research, of particular relevance are studies of TBL integration into ESP instruction. Lytovchenko et al. [9] compared the outcomes of grammar acquisition in two groups of ESP students at a technical university: one taught using the traditional PPP (presentation, practice, and production) approach and

the other using TBL. Better achievement in grammar tests and higher motivation were fixed in the TBL group. Positive results were also found when using the TBL approach to teach ESP vocabulary [10], [11]. Students in TBL groups learned vocabulary faster and easier. They were better prepared to apply it in real-life professional contexts while improving their communication skills and self-confidence in the professional setting.

2.4 The role of AI in language learning

A novel approach to language learning and teaching has emerged with the integration of AI, particularly ChatGPT. Among the most important advantages of ChatGPT is its ability to give learners instant feedback, language support, and assist in producing authentic language texts. Research on AI integration in language education has shown its potential to promote language production and improve language skills and soft skills, including critical thinking, research, creativity, and collaboration [12], [13], [14].

As shown by empirical studies, AI tools, such as chatbots, provide personalized learning experiences, accelerate language acquisition, and improve student outcomes in areas such as vocabulary and grammar [15], [16]. Researchers also found AI's positive role in improving pronunciation and enhancing engagement through interactive chatbots [17]. AI tools have been found particularly useful for writing tasks, acting as a comprehensive writing assistant, from idea generation to proofreading, improving content quality and organization [18]. The ability to perform tasks such as answering questions and guiding discussions further extends its utility in education. AI chatbots are also increasingly used in language learning to simulate real-life interactions, providing immediate feedback and allowing learners to practice language skills autonomously. This has proven beneficial in improving both linguistic and cultural understanding [19].

2.5 Artificial intelligence integration into TBL for ESP writing

Integrating AI into task-based ESP learning represents a cutting-edge approach in language education aimed at using this advanced technology to improve learning achievement, develop language skills, and support language acquisition. Particularly relevant to our research are studies on the role of AI, especially ChatGPT, in various writing tasks. The introduction of ChatGPT in education is the most debated due to its enormous potential for academic writing, long essays, short stories, poems, and even letters [20]. According to Imran and Almusharraf [12], the most impressive role of ChatGPT is its use as a writing assistant.

This chatbot can perform various tasks, including answering questions, writing, coding, and guiding through discussions [21]. ChatGPT's important role in accomplishing writing tasks in science and academia is globally recognized [22]. Based on the study of current literature, Imran and Almusharraf [12] summarized ChatGPT's role in writing tasks as "a complete package from ideas generation to final proofreading and editing of writing material" (p. 2). Thus, integrating AI tools such as ChatGPT into ESP TBL opens new perspectives for language learning, significantly improving language skills and transforming writing tasks.

However, despite the growing recognition of AI's role in language education, particularly in writing tasks, there is a notable gap in research concerning its integration into task-based ESP learning. While studies have demonstrated the effectiveness of TBL and highlighted the advantages of AI in supporting language production, the combined use of TBL and AI in ESP instruction has not, to our knowledge, been investigated. This study aims to address this gap by exploring how ChatGPT can enhance writing performance and support task completion in a TBL setting, highlighting the potential for AI to transform traditional ESP learning.

3 METHODS AND MATERIALS

3.1 Research design

To assess the impact of ChatGPT integration in TBL, we employed a mixed research design that combined quantitative and qualitative research methods. Using a quantitative research method permitted the comparison of the quality of students' writing task performance between two TBL groups: one that used ChatGPT to complete the tasks and one that did not use this tool. The qualitative method was used in the survey of students to explore their attitudes towards using ChatGPT to complete the tasks.

3.2 Participants

The study involved 27 bachelor's degree students of the National Technical University of Ukraine, "Igor Sikorsky Kyiv Polytechnic Institute" (Kyiv, Ukraine). As part of the ESP course, they studied the topic "Citizen Science," which included writing a sustainable campus initiative proposal. The students studied online. Their English language proficiency level was B1 on the CEFR scale.

The sampling technique employed was purposive sampling, as participants were selected based on specific criteria: their enrollment in the university's ESP course and their B1 English proficiency. This ensured that the study targeted a specific population relevant to the research objective.

The students were informed about the anonymity and confidentiality of data collection and voluntarily agreed to participate in the experiment. They studied in two academic groups: the experimental group (13 students) using the TBL method with the integration of ChatGPT and the control group (14 students) using the TBL method without the integration of ChatGPT. Each of the groups was further divided into three teams that performed all the tasks independently in breakout rooms in Zoom.

3.3 Materials and procedure

The students studied the topic "Citizen Science" in three lessons (each 90 minutes long) using the TBL method. They worked in teams in breakout rooms (three teams in the experimental group and three teams in the control group) on the zoom platform. Each team independently carried out the following tasks in a google document: 1) listing, 2) ordering and sorting, 3) problem solving, 4) sharing personal experience, 5) comparing, and 6) creative tasks [8] (refer to Table 1).

Table 1. The topic of TBL lessons: sustainable campus initiative proposal

Task	Procedure	Inter-Action
Introduction to topic	– The teacher asks the students to read an example of a proposal and define the goals of the proposed project (See Ex. 8A, p. 46 in [23]).	T-S
	– The teacher asks the students questions about the purpose of each section of the proposal (See Ex. 8A, p. 46 in [23]).	T-S
Task 1: Listing	– The teacher gives the students the background information about the task they are going to perform: the university they are studying at is going green and taking steps to become more environmentally sustainable. The university's administration has announced the Sustainable Campus Initiative and asked each faculty and department to contribute proposals that may help achieve the goal. As engineering students, their task is to create a proposal that would suggest a sustainable solution to any engineering issue on campus related to energy efficiency, waste reduction, or water conservation.	T-S
	– Students work in small teams to gather ideas about how to promote sustainability on the university campus by researching and brainstorming. They also write down the list of the ideas they brainstormed in a Google document. The teacher acts as a facilitator and helps students to conduct the research.	S-S
Task 2: Ordering and sorting	– In the same teams, the students rank the brainstormed ideas about the ways of making the university campus more sustainable in a meaningful order: <ul style="list-style-type: none"> • ideas related to energy efficiency; • ideas related to waste reduction; • ideas related to water conservation. 	S-S
Task 3: Problem solving	– In the same teams, students choose two ideas from the list they created and analyze possible solutions to implement each on campus. The teacher provides feedback on content and form.	S-S
Task 4: Sharing personal experience	The students share their ideas and ways of their implementation with the whole class, providing as many details as possible.	S-S
Task 5: Comparing and contrasting	By pointing out the similarities and differences the students compare the ideas and ways of their implementation suggested in their team with those indicated in other teams.	S-S
Task 6: Creative task	In the same teams, students collaborate to plan and write the Sustainable Campus Initiative Proposal (200–250 words) using the Grant Proposal (See Ex. 8A, p. 46 in [23]). <i>Speakout</i> (Intermediate Students' Book), Second edition. Pearson Education Limited) as a model.	S-S

The experimental group performed the tasks with the employment of ChatGPT. The control group performed the tasks using the internet resources except for AI tools. On completing the tasks, six sustainable campus initiative proposals (three in the experimental group and three in the control group) were assessed by the teacher. Additionally, the experimental group students were surveyed

about the utility of ChatGPT in their work on the tasks. For the data gathering, we utilized a questionnaire that was developed based on a review of existing literature and adapted to this specific context. It consisted of 7 Likert scale questions, ranging from strongly agree (1) to strongly disagree (5), and 1 open-ended question. The Likert scale questions measured students' perceptions of ChatGPT's utility in aspects such as ease of use, contribution to task completion, and quality improvement. The open-ended question provided evidence of the specific functions ChatGPT served in the writing process: "When writing the proposal, what specific purposes did you employ ChatGPT for?"

3.4 Data analysis

The paper uses the following key data analysis methods: rubric-based evaluation, independent samples t-test, and descriptive statistics for analyzing survey responses.

To analyze the results of the creative task (writing a sustainable campus initiative proposal), we used rubric-based evaluation. The rubric included predefined criteria for assessing the quality of the proposals:

- Content (How effectively did the students address the topic and task, i.e., did they do what was asked?);
- Communicative achievement (did the students use the correct text type? How clever were their ideas? Did they explain their ideas clearly? Did they use a proper register?);
- Creativity (How unique and original were the students' ideas?);
- Organization (How logical and ordered was the students' writing?);
- Language (What was the range and accuracy of vocabulary and grammar use?).

Each criterion was scored on a scale from 1 to 5, which provided a structured approach to evaluating the proposals.

Since the proposals are qualitative outputs (written texts), a rubric-based evaluation provides a structured and objective way to quantify the quality across multiple dimensions. It allows consistent scoring across different groups (experimental and control) and teams, ensuring that subjective elements of writing are evaluated in a standardized way.

An independent samples t-test (using statistical software (<http://www.statisticslectures.com/calculators/ttest2/>)) was used to compare the quality of proposals between the two groups. This parametric test is appropriate for comparing the means of two independent groups when the dependent variables are continuous (scored on a scale from 1 to 5). The t-test helps establish whether the differences in writing performance between the two groups are statistically significant, which aligns with the study's objective of comparing the effect of ChatGPT use.

Microsoft excel and descriptive statistics, including median and interquartile range (IQR), were used to analyze responses to Likert scale-type questions, which measured the students' attitudes towards using ChatGPT to complete the tasks and write the proposal. Descriptive statistics such as the median are appropriate for summarizing the central tendency of the Likert scale data. At the same time, IQR helps understand the spread of responses, which allows us to assess how the students' attitudes vary.

4 RESULTS

In our study we compared the quality of writing a sustainable campus initiative proposal in two TBL groups (experimental and control ones) studying ESP at university. Writing the proposal was a creative task. It was the final task that followed listing, ordering, problem-solving, sharing personal experience, and comparing (refer to Table 1). The experimental group used ChatGPT to complete the tasks. The control group used different internet resources except for AI tools.

The results of our study showed that the overall quality of the sustainable campus initiative proposals written in the control group (which did not use ChatGPT) was better than in the experimental group (which used ChatGPT) (refer to Table 2).

Table 2. Evaluation scores for sustainable campus initiative proposals in the control and experimental groups according to assessment criteria

Assessment Criteria	Control Group A (Number of Scores)	Experimental Group B (Number of Scores)
Content		
Team 1	4	2
Team 2	5	3
Team 3	4	4
Communicative achievement		
Team 1	5	2
Team 2	5	3
Team 3	4	4
Creativity		
Team 1	5	2
Team 2	5	4
Team 3	5	4
Organization		
Team 1	5	3
Team 2	5	4
Team 3	5	5
Language		
Team 1	5	3
Team 2	4	4
Team 3	5	5
Total scores		
Team 1	24	12
Team 2	24	18
Team 3	23	22

As shown by Table 3, the difference in total scores for the writing task between the control group A and the experimental group B was significant ($t(4) = 2.1653$ and $p < .05$). This means that, contrary to our expectation, the overall quality of the proposals that were written without the use of AI tools was higher than the quality of the proposals written with the employment of ChatGPT.

Table 3. Comparison of the total scores for the writing task in the control group A and the experimental group B

Descriptive Statistics			
	Mean	Standard Deviation	n
Group A	23.6667	0.5774	3
Group B	17.3333	5.0332	3
Independent Samples t-Test			
t-Statistic	2.1653	Result	
Degrees of Freedom	4	Reject the null hypothesis	
Critical Value	2.1318	Conclusion	
95% Confidence Interval	[2918, 97487]	Group A is significantly different from Group B, $t(4) = 2.1653$, $p < .05$. We are 95% confident that the mean difference lies between 2.918 and 9.7487.	

The comparison of the sustainable campus initiative proposals based on each assessment criterion revealed the following results. The assessment of the sustainable campus initiative proposals by the content, organization, and language criteria showed no significant difference in the quality between the control group A and the experimental group B (Tables 4–6). The observed difference in means could be attributed to random variation rather than a true difference.

The scores for the content criterion in both groups, which appear quite similar ($t(4) = 2$, $p > .05$), suggest that the use of ChatGPT in the experimental group did not notably enhance or diminish the students' ability to address the content requirements of the task (Table 4). Both groups demonstrated a comparable understanding and inclusion of relevant information in their sustainable campus initiative proposals.

Table 4. Comparison of the scores assigned for the content criterion in control group A and experimental group B

Descriptive Statistics			
	Mean	Standard Deviation	n
Group A	4.3333	0.5774	3
Group B	3	1	3
Independent Samples t-Test			
t-Statistic	2	Result	
Degrees of Freedom	4	Do not reject the null hypothesis	
Critical Value	2.1318	Conclusion	
95% Confidence Interval	[0.5549, 2.1118]	Group A is not significantly different from Group B, $t(4) = 2$, $p > .05$.	

The absence of any significant difference in the scores assigned for the organization criterion between control group A and experimental group B ($t(4) = 1.7321$, $p > .05$) indicates that both groups demonstrated similar abilities in structuring and organizing their sustainable campus initiative proposals (refer to Table 5). The similarity in the scores could suggest that the use of ChatGPT did not have a pronounced effect on the organizational aspects of the task.

Table 5. Comparison of the scores assigned for the organization criterion in control group A and experimental group B

Descriptive Statistics			
	Mean	Standard Deviation	n
Group A	5	0	3
Group B	4	1	3
Independent Samples t-Test			
t-Statistic	1.7321	Result	
Degrees of Freedom	4	Do not reject the null hypothesis	
Critical Value	2.1318	Conclusion	
95% Confidence Interval	[0.3259, 1.6741]	Group A is not significantly different from Group B, $t(4) = 1.7321$, $p > .05$.	

The insignificant difference in scores for the language criterion between control group A and experimental group B ($t(4) = 1$, $p > .05$) demonstrates that the students in both groups performed similarly when it came to grammar, vocabulary, and overall language accuracy in their sustainable campus initiative proposals (refer to Table 6).

Table 6. Comparison of the scores assigned for the language criterion in control group A and experimental group B

Descriptive Statistics			
	Mean	Standard Deviation	n
Group A	4.6667	0.5774	3
Group B	4	1	3
Independent Samples t-Test			
t-Statistic	1	Result	
Degrees of Freedom	4	Do not reject the null hypothesis	
Critical Value	2.1318	Conclusion	
95% Confidence Interval	[-0.1118, 1.4451]	Group A is not significantly different from Group B, $t(4) = 1$, $p > .05$.	

The assessment of the sustainable campus initiative proposals by the rest two criteria, communicative achievement and Creativity, quite unexpectedly, showed statistically significantly higher results in Control Group A, where ChatGPT was not used (refer to Tables 7–8).

The significant difference in scores for the communicative achievement criterion ($t(4) = 2.5$, $p < .05$) (refer to Table 7) between the control group A and experimental

group B, with the control group outperforming the experimental group, suggests that the students who did not use ChatGPT were more effective in achieving the communicative goals of the task, such as addressing the target audience appropriately and conveying ideas more clearly and persuasively.

Table 7. Comparison of the scores assigned for the communicative achievement criterion in control group A and experimental group B

Descriptive Statistics			
	Mean	Standard Deviation	n
Group A	4.6667	0.5774	3
Group B	3	1	3
Independent Samples t-Test			
t-Statistic	2.5	Result	
Degrees of Freedom	4	Reject the null hypothesis	
Critical Value	2.1318	Conclusion	
95% Confidence Interval	[0.8882, 2.4451]	Group A is significantly different from Group B, $t(4) = 2.5$, $p < .05$. We are 95% confident that the mean difference lies between 0.8882 and 2.4451.	

Another significant difference in scores (for the Creativity Criterion with the control group once again outperforming the experimental group) ($t(4) = 2.5$, $p < .05$) suggests that the students in control group A demonstrated higher creativity in their proposals, as reflected in their ability to generate original and innovative ideas (refer to Table 8).

Table 8. Comparison of the scores assigned for the creativity criterion in control group A and experimental group B

Descriptive Statistics			
	Mean	Standard Deviation	n
Group A	5	0	3
Group B	3.3333	1.1547	3
Independent Samples t-Test			
t-Statistic	2.5	Result	
Degrees of Freedom	4	Reject the null hypothesis	
Critical Value	2.1318	Conclusion	
95% Confidence Interval	[0.8882, 2.4451]	Group A is significantly different from Group B, $t(4) = 2.5$, $p < .05$. We are 95% confident that the mean difference lies between 0.8882 and 2.4451.	

The survey of the students of experimental group B conducted after the experimental learning and dealing with the usability of ChatGPT for completing the tasks showed a generally positive opinion of the respondents on the usefulness of this tool (refer to Table 9). Generally, the students demonstrated a positive perception of the functionality of ChatGPT across different tasks. There was a strong agreement among them that this tool aids in brainstorming ideas, problem-solving, and finding

information (Median = 2). Most respondents (73.3%) also found ChatGPT useful for composing a written text. At the same time, a quarter of them (26.7%) gave “hard to answer” responses to this question, which suggests that some participants may need further exploration of the functionality of this tool for this particular task.

Table 9. Experimental group students’ opinion on the usefulness of ChatGPT for completing the tasks

Statement	Strongly Agree N (%)	Agree N (%)	Hard to Answer N (%)	Disagree N (%)	Strongly Disagree N (%)	Median	IQR
1. ChatGPT is useful for brainstorming ideas	4 (26.7%)	8 (53.3%)	1 (6.7%)	2 (13.3%)	0 (0%)	2	0.5
2. ChatGPT is useful for problem-solving	3 (20%)	9 (60%)	1 (6.7%)	2 (13.3%)	0 (0%)	2	0
3. ChatGPT is useful for composing text	3 (20%)	8 (53.3%)	4 (26.7%)	0 (0%)	0 (0%)	2	0.5
4. ChatGPT helps find information for writing	5 (33.3%)	9 (60%)	1 (6.7%)	0 (0%)	0 (0%)	2	1
5. ChatGPT helps organize written content	1 (6.7%)	6 (40%)	5 (33.3%)	2 (13.3%)	1 (6.7%)	3	1
6. ChatGPT helps format, edit, and refine style	3 (20%)	3 (20%)	4 (26.7%)	3 (20%)	2 (13.3%)	3	2
7. ChatGPT output needs evaluation and revision	7 (46.7%)	5 (33.3%)	2 (13.3%)	1 (6.7%)	0 (0%)	2	1

Even more uncertainty was demonstrated about ChatGPT’s abilities to organize content, format the text, and edit the language and style (Median = 3). However, there was a general consensus on the need to evaluate and revise AI-generated text critically, underlining the importance of human supervision and judgment in the use of AI tools in writing (Median = 2).

The analysis of the students’ answers to the open-ended question (“When writing the proposal, what specific purposes did you use Chat GPT for?”) showed that they used it for linking parts of the text in a logical way; making a quick search of information and finding answers to questions; generating action plans; formulating ideas; planning and developing projects on conservation of energy and protection of the environment, etc. As can be seen, the students used ChatGPT for different purposes and found it helpful in various aspects of writing the Sustainable Campus Initiative Proposal, from generating ideas and searching for information to planning and improving projects. However, besides the numerous benefits of this tool, some students pointed out that it may produce an overload of information that needs to be filtered.

5 DISCUSSION

Our study aimed to evaluate the effectiveness of integrating ChatGPT in teaching ESP at university using the TBL method and to find out students’ perceptions of using this AI tool to complete the tasks.

We found the results of our research quite unexpected because they showed that the overall quality of the sustainable campus initiative proposals in the control group, which did not use ChatGPT, was higher than the quality of the proposals created in the experimental group that used this tool. In our opinion, this suggests that

despite its ability to assist students in different stages of completing the tasks, the use of ChatGPT may not always result in a final product of the highest quality.

Contrary to expectations, the experimental group students had significantly lower grades for communicative achievement and creativity than the control group students. This finding is important, as it suggests that relying on AI tools may not enhance certain communicative aspects of writing, particularly those requiring a nuanced understanding of audience and purpose. ChatGPT's automated text generation might not have fully captured the communicative intent that students achieved through their independent writing. Thus, the control group's higher scores indicate that human-driven writing could better tailor communication to specific contexts. The findings also suggest that AI-generated text might have been more formulaic or less inventive, limiting the students' ability to express their own creative ideas. The control group, working independently without AI assistance, may have had more freedom to explore unique approaches to the task, resulting in higher creativity scores.

The assessment of the students' sustainable campus initiative proposals by the content, organization, and language criteria was also somewhat surprising since it did not reveal any significant differences between the groups. These results could indicate that AI tools such as ChatGPT may not necessarily impact the depth of content creation, especially when the task requires specific knowledge or thoughtful reflection. The students in the control group, relying on non-AI Internet resources, performed just as well as those using AI. It is possible that the task's content requirements demanded more critical thinking and subject matter understanding, areas where AI assistance may be less effective compared to human-driven research.

This result also suggests that AI tools such as ChatGPT might not significantly improve the ability to organize written content. Instead, the students in both groups seem to have relied on their pre-existing skills and knowledge, which resulted in comparable scores for this criterion. The lack of statistical significance implies that any observed differences could be due to random variation rather than a true impact of ChatGPT usage. The findings also imply that the use of ChatGPT did not provide a clear advantage in improving the language quality of the proposals. The control group students who did not use AI tools were just as capable of producing well-written texts as those using ChatGPT. This might indicate that while AI tools can assist with language-related tasks, the students' own proficiency played a more decisive role in the overall quality of language use.

Thus, the analysis of the quality of writing in the experimental group showed that it varied across different criteria (content, organization, and language were of higher quality than communicative achievement and creativity), which might be explained by the fact that ChatGPT's efficiency depends on the educational context and purposes for which it is used. It also seems logical to assume that the difference in the quality of various aspects of students' writing depended on how well they had mastered ChatGPT and how they utilized it.

Our findings support the work of other researchers who found ChatGPT useful for educational contexts similar to those presented in this study. For example, Kovačević [24] utilized ChatGPT in ESP instruction to create instructional materials, such as generating textual content. Imran and Almusharraf [12], in their systematic review of the literature, revealed that, at present, ChatGPT is most extensively used as a writing tool or writing assistant at higher education levels. There are also similarities between the present study and previous research, which anticipated that AI tools might significantly influence writing tasks in the future but, at the same time, recognized that they have quite a limited output at present [25], [26], [27], [28].

The results of the survey conducted in the experimental group showed generally positive attitudes of students to using ChatGPT for completing tasks.

ChatGPT was found particularly useful for brainstorming ideas, problem-solving, searching for information, and composing written text, which reflects its potential to promote cognitive and research processes in learning. However, the students expressed uncertainty regarding ChatGPT's efficiency in organizing content, formatting, editing language, and refining style. This shows that while highly estimating ChatGPT's ability to provide significant initial input, students still realize that the final product needs to be critically approached, improved, and polished by a human. The general consensus on the importance of critical evaluation and revision of text generated by AI emphasizes the need for students to develop writing and critical thinking skills to use AI tools with maximum benefit.

Similar findings were reported by Shoufan [29], who surveyed senior computer engineering students after completing a learning activity using ChatGPT and found that overall, they highly valued the capabilities of this tool but expressed concerns regarding the accuracy of ChatGPT's responses. The same concerns were expressed by the university students participating in the study by Xu and Jumaat [30] that explored ChatGPT's applications in English academic writing. The students noted that prior knowledge is required to use this tool effectively, and it is crucial to scrutinize the quality of AI-generated texts.

Our findings also support previous research that found ChatGPT a useful collaborative tool for receiving feedback [31]–[32], customizing search results, editing the language, improving grammar, structure, and coherence of the written text, generating the content, producing initial drafts, and brainstorming ideas [12]. Still, it cannot substitute or be equated with a human writer in terms of various skills and knowledge.

This aligns with the observations of Fuchs [33], who emphasized that students in higher education should develop competence and acquire higher-order thinking skills (e.g., critical thinking or problem-solving) and warned them against overreliance on technology. The accuracy of AI models is impacted by the quality of the data they are trained on. However, it also depends on the quality of the input provided by students. If their query is not clear, concise, and relevant, the system might not understand and produce an inaccurate response. Higher-order thinking skills will help students articulate their prompts in a way that can be easily understood by the system and critically evaluate the accuracy and relevance of the responses generated by the system. To address these challenges, universities should use AI models to supplement human interaction and not replace it.

The results of our research have certain implications for integrating ChatGPT or other AI tools into education. First of all, we believe that AI cannot substitute traditional methods or human evaluation, so its role in learning can be defined as complementary. To use AI tools with maximum advantage, students should be taught to critically evaluate and refine AI-generated text.

Since the efficiency of ChatGPT varies across different aspects of writing, it is useful to create instructional strategies focused on developing particular aspects. For example, teachers may focus on improving students' communicative achievement and creativity in writing through the use of both AI tools and traditional methods, keeping the balance between the use of AI technology and traditional pedagogical approaches.

Finally, the positive attitudes of students towards ChatGPT mean that students are open to using AI technologies, which is in line with the empirical study by Lavidas, Voulgari, Papadakis, Athanassopoulos, Anastasiou, Filippidi, and Karacapilidis [34] that explored factors influencing students' intention to use AI. Therefore, integrating AI tools into the learning process and further investigating their potential in TBL can foster students' engagement and innovation in education.

Several factors in our study suggest both strengths and potential limitations regarding reliability. The tasks performed by both groups followed a well-defined structure, which contributes positively to reliability. The sequence of tasks (listing, ordering, problem-solving, sharing personal experience, and comparing) was consistent across groups. Both the control and experimental groups were evaluated using clear criteria (content, communicative achievement, creativity, organization, and language). This consistency in evaluation supports the reliability of the study outcomes.

The experimental and control groups consisted of three teams each, ensuring a balanced comparison. However, the small number of participants (three teams in each group) challenges reliability. With such a small sample size, random factors may make the results susceptible to variability.

Using t-tests to determine statistical significance makes findings more quantitative and reliable. However, as some criteria did not show statistically significant differences (e.g., content, organization, and language), it suggests that the results for these dimensions may not be consistent or reliable enough to draw firm conclusions.

The survey of students in the experimental group revealed general consistency in their opinions about ChatGPT's usefulness for specific tasks (brainstorming, problem-solving, and finding information). However, the variance in responses regarding writing organization and style (as reflected by the high interquartile range) suggests a degree of unreliability in the perceived effectiveness of ChatGPT for specific writing tasks and indicates that further exploration is needed. These findings highlight the importance of human intervention in evaluating AI-generated text, which aligns with the observed lower performance in creativity and communicative achievement.

The validity of this study's findings can be evaluated across several dimensions, including construct validity, internal validity, external validity, and ecological validity. The study has a clear construct validity, as the evaluation criteria (content, communicative achievement, creativity, organization, and language) are relevant and comprehensive measures of proposal quality. These criteria align with standard practices for evaluating writing tasks in academic settings, ensuring the study measures the right aspects of writing quality.

The study's internal validity is based on its controlled experimental design, where only the use of the AI tool (ChatGPT) was varied between the two groups. The use of ChatGPT as the experimental factor makes it likely that observed differences in proposal quality can be attributed to this tool, particularly for criteria like communicative achievement and creativity, where the control group performed significantly better. However, the internal validity may be affected by confounding factors such as students' prior familiarity or experience with ChatGPT, which could influence how effectively they used the tool.

The generalizability of the findings is more limited due to the small sample size and context-specific nature of the task (writing a sustainable campus initiative proposal). The small number of participants and the focus on a specific task in a university ESP class may limit the ability to generalize these findings to other academic settings or tasks.

However, the study involved a real-world task (writing a sustainable campus initiative proposal), which enhances ecological validity. The results are more likely to be relevant and applicable to similar academic and professional contexts.

Thus, although the study provides valuable insights into the use of ChatGPT for creative writing tasks such as the sustainable campus initiative proposal, there are limitations in both reliability and validity. Reliability is somewhat compromised by the small sample size, which increases the likelihood of variability and random

effects, particularly for criteria where no significant differences were found. Validity is generally strong in terms of construct and internal validity, as the assessment criteria are relevant and the experimental design isolates the effect of ChatGPT. However, limitations in external validity (due to the specific context and small sample) and potential confounding factors (e.g., prior experience with AI tools) suggest that the findings should be interpreted with caution.

Despite the mixed results from the current study, the use of ChatGPT in TBL for ESP education has promising potential for enhancing various aspects of the learning process. When properly integrated and used with adequate guidance, AI tools can serve as valuable learning aids, helping students develop essential skills for academic and professional success. As the technology continues to evolve, further research into the most effective methods of AI-assisted learning could yield even greater benefits for students and educators alike.

6 CONCLUSIONS

In this study we aimed to evaluate the integration of ChatGPT in TBL in a university ESP course. The findings suggest that while ChatGPT offers support comparable to non-AI tools for brainstorming ideas, problem-solving, generating content, composing written texts, and language enhancement, it is less effective for other aspects of writing. In the control group, where ChatGPT was not used, the overall quality of writing was higher, especially in terms of communicative achievement and creativity, which highlights the need for human intervention and refinement when using AI tools for academic purposes.

Students generally perceive ChatGPT positively, appreciating its role in facilitating the writing process and providing immediate assistance. However, they also acknowledge the necessity for critical evaluation and human refinement of AI-generated texts to ensure the quality and originality of their work.

The significance of this empirical study is that it broadens our understanding of the role of ChatGPT in task-based language learning. While ChatGPT is a useful tool for improving different aspects of writing, it does not replace the critical thinking and creativity of a human. This emphasizes the need for a balanced approach that combines AI tools with traditional pedagogical methods to achieve maximum results in learning.

Since this study was limited in terms of the sample size and duration, further research is necessary to explore larger and more diverse samples in a variety of academic and professional tasks for extended periods of time to establish the long-term impacts of ChatGPT on student learning outcomes.

7 REFERENCES

- [1] R. Ellis, P. Skehan, S. Li, N. Shintani, and C. Lambert, *Task-Based Language Teaching: Theory and Practice*. Cambridge: Cambridge University Press, 2020. <https://doi.org/10.1017/9781108643689>
- [2] D. Willis and J. Willis, "Task-based language learning," in *The Cambridge Guide to Teaching English to Speakers of Other Languages*, R. Carter and D. Nunan, Eds., Cambridge: Cambridge University Press, 2009, pp. 173–179. <https://doi.org/10.1017/CBO9780511667206.026>

- [3] D. Nunan, *Task-Based Language Teaching*. Cambridge: Cambridge University Press, 2004. <https://doi.org/10.1017/CBO9780511667336>
- [4] I. Lytovchenko, "How to make upper-level university classes more interactive," *English Teaching Forum*, no. 2, pp. 24–29, 2009. [Online]. Available: https://americanenglish.state.gov/files/ae/resource_files/09-47-2-e.pdf
- [5] J. M. Herrin, "Six (6) Tasks for Task-based Learning (TBL)," 2009. [Online]. Available: <https://nmtesol.org/wp-content/uploads/2009/04/6-tasks-for-tbl-part-copy.doc>
- [6] N. S. Prabhu, *Second Language Pedagogy*. Oxford: Oxford University Press, 1987.
- [7] R. Ellis, "Task-based language teaching: Sorting out the misunderstandings," *International Journal of Applied Linguistics*, vol. 19, no. 3, pp. 221–246, 2009. <https://doi.org/10.1111/j.1473-4192.2009.00231.x>
- [8] D. Willis, *A Framework for Task-Based Learning*. Harlow, Essex: Longman, 1996.
- [9] I. Lytovchenko, Y. Lavrysh, V. Lukianenko, and O. Ogienko, "How to teach grammar to adult ESP learners at technical university more communicatively: Task-based approach," *Multidisciplinary Journal for Education, Social and Technological Sciences*, vol. 7, no. 1, pp. 54–71, 2020. <https://doi.org/10.4995/muse.2020.12419>
- [10] N. Nychkalo, W. Jinba, L. Lukianova, N. Paziura, and N. Muranova, "Use of task-based approach in teaching vocabulary to business English learners at university," *Advanced Education*, vol. 7, no. 16, pp. 98–103, 2020. <https://doi.org/10.20535/2410-8286.215117>
- [11] I. Lytovchenko et al., "Online problem-based learning: Possibilities for engineering vocabulary acquisition in ESP course at technical university," *International Journal of Information and Education Technology*, vol. 12, no. 9, pp. 905–911, 2022. <https://doi.org/10.18178/ijiet.2022.12.9.1700>
- [12] M. Imran and N. Almusharraf, "Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature," *Contemporary Educational Technology*, vol. 15, no. 4, p. ep464, 2023. <https://doi.org/10.30935/cedtech/13605>
- [13] S. Fiialka, Z. Kornieva, and T. Honcharuk, "ChatGPT in Ukrainian education: Problems and prospects," *International Journal of Emerging Technologies in Learning*, vol. 18, no. 17, pp. 236–250, 2023. <https://doi.org/10.3991/ijet.v18i17.42215>
- [14] O. Synekop, I. Lytovchenko, Y. Lavrysh, and V. Lukianenko, "Use of Chat GPT in English for engineering classes: Are students' and teachers' views on its opportunities and challenges similar?" *International Journal of Interactive Mobile Technologies (ijIM)*, vol. 18, no. 3, pp. 129–146, 2024. <https://doi.org/10.3991/ijim.v18i03.45025>
- [15] R. Vall and F. Araya, "Exploring the benefits and challenges of AI-language learning tools," *International Journal of Social Sciences and Humanities Invention*, vol. 10, no. 1, pp. 7569–7576, 2023. <https://doi.org/10.18535/ijsshi/v10i01.02>
- [16] S. Athanassopoulos, P. Manoli, M. Gouvi, K. Lavidas, and V. Komis, "The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom," *Advances in Mobile Learning Educational Research*, vol. 3, no. 2, pp. 818–824, 2023. <https://doi.org/10.25082/AMLER.2023.02.009>
- [17] J. Son, N. Ružić, and A. Philpott, "Artificial intelligence technologies and applications for language learning and teaching," *Journal of China Computer-Assisted Language Learning*, 2023. <https://doi.org/10.1515/jccall-2023-0015>
- [18] U. Widiati, D. Rusdin, I. Indrawati, and N. Govender, "The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective," *Cogent Education*, vol. 10, no. 2, 2023. <https://doi.org/10.1080/2331186X.2023.2236469>
- [19] K. Mageira, D. Pittou, A. Papasalouros, K. Kotis, P. Zangogianni, and A. Daradoumis, "Educational AI Chatbots for content and language integrated learning," *Applied Sciences*, vol. 12, no. 7, p. 3239, 2022. <https://doi.org/10.3390/app12073239>

- [20] I. Dergaa, K. Chamari, P. Zmijewski, and H. B. Saad, "From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing," *Biology of Sport*, vol. 40, no. 2, pp. 615–622, 2023. <https://doi.org/10.5114/biolsport.2023.125623>
- [21] B. D. Lund and T. Wang, "Chatting about ChatGPT: How may AI and GPT impact academia and libraries?" *Library Hi Tech News*, 2023. <https://doi.org/10.2139/ssrn.4333415>
- [22] M. Sallam, "ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns," *Healthcare*, vol. 11, no. 6, p. 887, 2023. <https://doi.org/10.3390/healthcare11060887>
- [23] A. Clare and J. J. Wilson, *Speakout. Intermediate Plus. Students' Book*, 2nd ed. Harlow, Essex: Pearson Education Limited, 2018.
- [24] D. Kovačević, "Use of ChatGPT in ESP teaching process," in *Proc. of the 2023 22nd International Symposium INFOTEH JAHORINA (INFOTEH)*, 2023, pp. 1–5. <https://doi.org/10.1109/INFOTEH57020.2023.10094133>
- [25] A. A. Alattar and M. M. McDowell, "The rise of AI coauthors: Navigating the future of scientific writing with ChatGPT," *Journal of Neurosurgery*, vol. 139, no. 5, pp. 1485–1486, 2023. <https://doi.org/10.3171/2023.3.JNS23254>
- [26] S. Biswas, "ChatGPT and the future of medical writing," *Radiology*, vol. 307, no. 2, p. 223312, 2023. <https://doi.org/10.1148/radiol.223312>
- [27] J. Wen and W. Wang, "The future of ChatGPT in academic research and publishing: A commentary for clinical and translational medicine," *Clinical and Translational Medicine*, vol. 13, no. 3, p. e1207, 2023. <https://doi.org/10.1002/ctm2.1207>
- [28] A. Zimmerman, "A ghostwriter for the masses: ChatGPT and the future of writing," *Annals of Surgical Oncology*, vol. 30, pp. 3170–3173, 2023. <https://doi.org/10.1245/s10434-023-13436-0>
- [29] A. Shoufan, "Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey," *IEEE Access*, vol. 11, pp. 38805–38818, 2023. <https://doi.org/10.1109/ACCESS.2023.3268224>
- [30] T. Xu and N. F. Jumaat, "ChatGPT-empowered writing strategies in EFL students' academic writing: Calibre, challenges and chances," *International Journal of Interactive Mobile Technologies (ijIM)*, vol. 18, no. 15, pp. 95–114, 2024. <https://doi.org/10.3991/ijim.v18i15.49219>
- [31] M. Jamshed, A. S. M. M. Ahmed, M. Sarfaraj, and W. U. Warda, "The impact of ChatGPT on English language learners' writing skills: An assessment of AI feedback on mobile," *International Journal of Interactive Mobile Technologies (ijIM)*, vol. 18, no. 19, pp. 18–36, 2024. <https://doi.org/10.3991/ijim.v18i19.50361>
- [32] M. Imran, N. Almusharraf, M. S. Abdellatif, and M. Y. Abbasova, "Artificial intelligence in higher education: Enhancing learning systems and transforming educational paradigms," *International Journal of Interactive Mobile Technologies (ijIM)*, vol. 18, no. 18, pp. 34–48, 2024. <https://doi.org/10.3991/ijim.v18i18.49143>
- [33] K. Fuchs, "Exploring the opportunities and challenges of NLP models in higher education: Is ChatGPT a blessing or a curse?" *Frontiers in Education*, vol. 8, p. 1166682, 2023. <https://doi.org/10.3389/feduc.2023.1166682>
- [34] K. Lavidas *et al.*, "Determinants of humanities and social sciences students' intentions to use artificial intelligence applications for academic purposes," *Information*, vol. 15, no. 6, p. 314, 2024. <https://doi.org/10.20944/preprints202404.0750.v1>

8 AUTHORS

Iryna Lytovchenko is a ScD in Education and Professor at the Department of English for Engineering No. 2, Faculty of Linguistics, National Technical University

of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute.” Her scientific interests include ESP teaching and learning, adult education, andragogy, and corporate education (E-mail: irinalyt@ukr.net).

Yuliana Lavrysh is a ScD in Education, Professor and Head of the Department of English for Engineering No. 2, Faculty of Linguistics, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute.” Her scientific interests include digital didactics, ESP teaching and learning, adult education, and personalized and autonomous learning (E-mail: lavu@ukr.net).

Oksana Synekop is a ScD in Education and Professor at the Department of English for Engineering No. 2, Faculty of Linguistics, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute.” Her scientific interests include ESP teaching and learning, differentiated instruction, communicative competence formation, learner and group autonomy, and ICT in education (E-mail: oksana.synekop@gmail.com).

Valentyna Lukianenko is an Associate Professor at the Department of English for Engineering No. 2, Faculty of Linguistics, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute.” Her scientific interests include ESP teaching and learning, innovative technologies in contemporary education, and student and teacher self-efficacy (E-mail: vv.lukian@gmail.com).

Oksana Chugai is an Associate Professor at the Department of English for Engineering No. 2, Faculty of Linguistics, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute.” Her scientific interests include emergency teaching, social-emotional learning, gamification, and cooperative learning (E-mail: OChugai@meta.ua).

Iryna Shastko is an Associate Professor at the Department of Social Sciences at Ivano-Frankivsk National Technical University of Oil and Gas, where she specializes in the fields of educational psychology, the psychology of management, and the ambivalence of motherhood. With a PhD in Psychological Sciences, Dr. Shastko’s research focuses on the complex psychological dynamics within educational and managerial contexts, as well as the nuanced experiences of motherhood (E-mail: iryna.shastko@nung.edu.ua).