VNU Journal of Foreign Studies

Journal homepage: https://jfs.ulis.vnu.edu.vn/

CORPUS-BASED COMPARISON OF CHATGPT AND STUDENT WRITING ON SPEAKING PROMPTS IN VIETNAMESE EFL CONTEXTS

Nguyen Thi Tuyet Mai*

Vietnam Military Medical University, No.160 Phung Hung, Ha Dong, Hanoi, Vietnam

Received 16 April 2025
Revised 04 September 2025; Accepted 11 December 2025

Abstract: This study examines written texts produced by Vietnamese EFL students and ChatGPT in response to ten common speaking prompts. Although initially designed for oral practice, the prompts were adapted into short writing tasks, ensuring comparability with ChatGPT's written outputs. A mixed-methods approach combined quantitative corpus profiling (type–token ratio, collocational range, discourse markers) with qualitative coding of syntax, lexis, and pragmatics. The analysis revealed that ChatGPT texts displayed greater lexical diversity (TTR 0.70 vs. 0.63), more natural collocations, and a wider range of cohesive devices. In contrast, student texts relied heavily on high-frequency verbs and basic connectors. Pragmatic contrasts also emerged: ChatGPT often employed hedges and polite expressions, while student writing tended to be abrupt or overly direct. These findings suggest that ChatGPT can enrich learners' exposure to varied vocabulary, collocational accuracy, and pragmatic awareness. However, its use should remain pedagogically guided to avoid over-reliance. The study underscores the importance of teacher mediation and aligns its findings with second language acquisition (SLA) perspectives on input, noticing, and authenticity.

Keywords: ChatGPT, second language acquisition (SLA), EFL learners, pragmatic awareness

* Corresponding author.
Email address: tuyetmai2611@gmail.com
https://doi.org/10.63023/2525-2445/jfs.ulis.5490

SO SÁNH DỰA TRÊN NGỮ LIỆU CÁC VĂN BẢN VIẾT CỦA CHATGPT VÀ SINH VIÊN VIỆT NAM HỌC TIẾNG ANH QUA CÁC CHỦ ĐỀ NÓI

Nguyễn Thị Tuyết Mai

Học viện Quân y, Số 160 Phùng Hưng, Hà Đông, Hà Nội, Việt Nam

Nhận bài ngày 16 tháng 4 năm 2025
Chỉnh sửa ngày 04 tháng 9 năm 2025; Chấp nhận đăng ngày 11 tháng 12 năm 2025

Tóm tắt: Nghiên cứu này xem xét các văn bản viết của sinh viên Việt Nam học tiếng Anh và các văn bản do ChatGPT tạo ra dựa trên 10 chủ đề nói thông dụng. Mặc dù được thiết kế cho thực hành nói, các chủ đề này đã được chuyển thành bài viết ngắn nhằm đảm bảo tính so sánh với văn bản do ChatGPT tạo ra. Nghiên cứu sử dụng phương pháp hỗn hợp, kết hợp phân tích ngữ liệu định lượng (tỉ lệ loại - từ, phạm vi kết hợp từ, từ nối diễn ngôn) với mã hóa định tính về cú pháp, từ vựng và dụng học. Kết quả cho thấy văn bản của ChatGPT có sự đa dạng từ vựng cao hơn (TTR 0,70 so với 0,63), sử dụng kết hợp từ tự nhiên hơn và phạm vi phương tiện liên kết rộng hơn, trong khi văn bản của sinh viên chủ yếu dựa vào động từ thông dụng và từ nối cơ bản. Khác biệt về dụng học cũng xuất hiện: ChatGPT thường dùng cách diễn đạt giảm nhẹ và lịch sự, trong khi văn bản của sinh viên có xu hướng đột ngột hoặc quá trực tiếp. Những phát hiện này gợi ý rằng ChatGPT có thể làm giàu vốn từ, nâng cao độ chính xác của kết hợp từ và nhận thức về dụng học cho người học. Tuy nhiên, việc sử dụng cần có định hướng sư phạm để tránh lạm dụng. Nghiên cứu nhấn mạnh vai trò trung gian của giáo viên và gắn kết kết quả với các quan điểm trong thụ đắc ngôn ngữ thứ hai về đầu vào, sự chú ý và tính xác thực.

Từ khóa: ChatGPT, thụ đắc ngôn ngữ thứ hai, sinh viên học tiếng Anh, nhận thức dụng học

1. Introduction

In recent years, the rapid advancement of artificial intelligence (AI), particularly in the development of large language models (LLMs) such as ChatGPT, has significantly reshaped our relationship with language production, comprehension, and instruction. These models are no longer confined to offering predictive text suggestions or simple grammar corrections. Instead, they can produce coherent, contextually appropriate, and grammatically accurate discourse in real time, mimicking human-like communication across diverse topics and registers.

As these capabilities become more sophisticated and widely accessible, their entry into educational contexts, especially English as a Second Language (ESL) classrooms, has been swift and often uncritical. ChatGPT is a versatile learning companion, from assisting with vocabulary suggestions and writing feedback to simulating dialogue partners and generating entire essays. However, the speed of this integration has outpaced the depth of our understanding of its linguistic and pedagogical consequences. While many educators and learners welcome its convenience, several linguistic and theoretical questions remain largely unexplored.

Beneath the surface of efficiency and fluency lie deeper concerns rooted in applied linguistics and second language acquisition (SLA). Scholars such as Krashen (1982) have long argued that language learning is not simply about exposure or imitation but about interaction, internalisation, and progressive development. For input to be practical, it must be comprehensible and situated just beyond the learner's current level, a notion encapsulated in Krashen's "i+1" principle. While ChatGPT can generate well-formed, native-like sentences,

there is uncertainty about whether it supports the type of linguistic struggle and form-focused attention required for accurate acquisition. If learners bypass the productive struggle by relying too heavily on ChatGPT-generated responses, they may miss opportunities to engage deeply with grammar, vocabulary, and meaning.

There is also the issue of authenticity, a long-standing concern in communicative language teaching. Widdowson (1998) emphasised that authenticity is not a fixed feature of texts but a dynamic quality co-constructed through social interaction and shared context. ChatGPT, however, produces language without human intent, negotiation, or situated meaning, raising questions about whether its output can truly support authentic communicative competence. This is particularly relevant in ESL settings that prioritise interactive tasks, role plays, and learner-centred discourse, where spontaneity and negotiation of meaning are key to development.

Against this backdrop, the current study aims to investigate the linguistic and pedagogical value of ChatGPT-generated texts within the context of SLA. Drawing on foundational theories in applied linguistics, it seeks to determine whether ChatGPT can offer input that is not only grammatically accurate but also contextually appropriate, interactionally relevant, and pedagogically meaningful for language learners.

Drawing on SLA frameworks, the comparison between ChatGPT and learner texts enables an evaluation of whether ChatGPT-generated input provides comprehensible, authentic-like, and noticeable linguistic features (Krashen, 1982; Schmidt, 1990; Widdowson, 1998) that may support learning, and how this contrasts with actual learner output.

Specifically, the study has three core objectives:

(1) to analyse the syntactic, lexical, and pragmatic features of ChatGPT-generated texts;

(2) to compare these features with learner-produced texts in similar tasks;

(3) to assess how ChatGPT's language aligns with key SLA theories, including Krashen's Input Hypothesis (1982), Schmidt's Noticing Hypothesis (1990), and authenticity frameworks by Widdowson (1998) and Gilmore (2007).

Building on these analyses, the paper also outlines pedagogical implications for integrating ChatGPT into EFL writing instruction, highlighting the importance of teacher mediation and learner reflection.

2. Literature Review

2.1. Foundational Theories in Second Language Acquisition (SLA)

Second Language Acquisition (SLA) is shaped by multiple cognitive, social, and contextual factors. Three influential frameworks, Krashen’s Input Hypothesis, Long’s Interaction Hypothesis, and Schmidt’s Noticing Hypothesis, form the theoretical basis for this study.

Krashen’s Input Hypothesis (1982) posits that acquisition occurs when learners receive comprehensible input slightly beyond their current level (i+1) and engage with language in meaningful contexts rather than through explicit grammar instruction.

Long’s Interaction Hypothesis (1996) emphasises the role of interaction and negotiation of meaning in helping learners notice language forms and refine their interlanguage, with feedback and clarification driving deeper engagement.

Schmidt’s Noticing Hypothesis (1990) argues that conscious attention to new linguistic forms is essential for acquisition, as subconscious exposure alone is insufficient.

In the context of AI, these theories raise the question of whether tools like ChatGPT can

provide input and interaction that genuinely support SLA. While ChatGPT can produce accurate, contextually appropriate language, its limited capacity for authentic communicative intent and active engagement poses challenges. Studies (Zhai et al., 2024; Khlaif, 2024; Xia et al., 2024) caution against overreliance on AI, noting risks to learner agency, critical thinking, and interaction quality. These insights suggest that ChatGPT should be used as a supplement within pedagogically mediated environments that align with comprehensibility, interaction, and awareness principles.

These theories directly guide the present study: Krashen's Input Hypothesis frames ChatGPT as potential i+1 input; Long's Interaction Hypothesis highlights what may be missing in AI-produced texts; and Schmidt's Noticing Hypothesis supports examining whether learners can attend to salient linguistic forms in ChatGPT output.

2.2. Authentic Language

Authentic language, central to communicative and task-based approaches, generally refers to language reflecting real-life communication with a genuine purpose in socially situated contexts.

This section clarifies how authenticity, as theorised in SLA, can be used as a benchmark to evaluate whether ChatGPT texts meet pedagogical standards of real-life communication.

Widdowson (1998) argued that authenticity is not inherent in a text but "a quality conferred on language by its use in appropriate contextual conditions," challenging teachers to focus on interactional context rather than text origin. Gilmore (2007) emphasised that authentic texts, produced by and for native speakers, often contain idiomatic expressions, real-world references, and unpredictable structures absent from simplified materials, offering richer input for learners.

Authenticity also involves interactional features such as negotiation of meaning, turn-taking, clarification, and co-construction (Ishida, 2006). These are often absent in machine-generated dialogues, which tend to be linear and context-insensitive. While ChatGPT can produce fluent, idiomatic sentences, its lack of genuine communicative purpose, responsiveness, and situated interaction raises questions about whether it meets the pedagogical standards of authenticity envisioned in SLA theory.

2.3. The Application of ChatGPT in Language Education

Integrating AI tools such as ChatGPT into language education has gained growing attention. Studies highlight both pedagogical opportunities and concerns in second language (L2) classrooms.

One key advantage is ChatGPT's ability to support language proficiency through interactive, contextualized practice. Kohnke et al. (2023) show that ChatGPT can act as a conversational partner, enabling simulated dialogues that enhance fluency and provide immediate feedback on vocabulary, grammar, and discourse organization, elements essential for communicative competence. Another benefit is personalization: by adapting output to learner input, ChatGPT allows differentiated scaffolding for various proficiency levels. Halaweh (2023) notes that this adaptability can reduce anxiety, boost motivation, and create more inclusive learning environments.

However, scholars also warn against overreliance on AI-generated feedback. Without critical engagement, learners may accept suggestions uncritically, hindering independent linguistic judgment (Halaweh, 2023). Academic integrity is another concern, as generating accurate, well-structured text raises questions about authorship and originality. Halaweh (2023)

stresses the need for clear policies and ethical guidelines for AI use.

These issues highlight the importance of teacher preparedness. As Kohnke et al. (2023) argue, effective integration depends on aligning AI use with curriculum goals and guiding learners to apply it purposefully. Teachers lacking digital literacy may fail to maximize benefits or mitigate risks. Nguyen (2024) adds that public discourse framing of AI, through metaphors like "AI as a human" or "AI as a force of nature", can shape learner perceptions, especially in EFL contexts.

Overall, literature presents a balanced view: ChatGPT can foster engagement, personalization, and communicative practice, but must be embedded thoughtfully into instructional design and supported by pedagogical expertise.

2.4. Previous Studies

- Vietnamese Studies on ChatGPT in Higher Education

Recent Vietnamese studies have begun to examine the practical implications of ChatGPT in higher education, highlighting its potential and challenges in teaching and learning contexts.

Nguyen (2024) emphasises ChatGPT's value in fostering self-learning and digital competence among young learners. His study shows that ChatGPT encourages proactive learning, helping students access and process information independently.

Building on this, Nguyen and Phan (2023) explore the tool's application in teaching, research, and administration at the university level. ChatGPT can enhance lesson preparation, support academic writing, and improve institutional efficiency. However, they also call for clear ethical guidelines in its use.

Taking a broader perspective, Duong (2023) views ChatGPT as a bridge to innovation in AI-era education. Her work outlines opportunities and obstacles in implementation, and she proposes training and curricular integration as key strategies for effective adoption.

These studies establish an important local foundation for understanding how ChatGPT can be responsibly integrated into Vietnamese higher education settings. International studies have also highlighted pedagogical benefits and concerns. Mahapatra (2024) reported improvements in ESL students' writing proficiency through ChatGPT-assisted feedback, while Yıldız (2024) observed enhanced speaking self-efficacy in EFL learners. By contrast, Abbas and Khan (2024) warned against the negative consequences of overreliance, including procrastination and reduced retention.

However, unlike these studies, the present research directly compares learner and AI-generated texts on identical prompts, addressing the lack of empirical contrastive evidence.

- ChatGPT in Higher Education in the World

In recent years, integrating generative AI tools like ChatGPT into education has attracted substantial scholarly interest, particularly in language learning and academic performance. Several empirical studies have highlighted the pedagogical benefits and potential concerns surrounding its application.

Mahapatra (2024) conducted a mixed-methods study exploring how ChatGPT affects the academic writing skills of ESL learners. Her findings revealed that ChatGPT significantly improved students' writing proficiency and confidence as a formative feedback tool, primarily through its instant, personalized responses.

Similarly, Yıldız (2024) examined the impact of ChatGPT on EFL learners' speaking self-efficacy. Through a controlled experiment, she found that students engaging in ChatGPT-

integrated speaking tasks developed greater confidence and fluency than those who followed traditional instruction. These two studies suggest ChatGPT may enhance receptive and productive language skills when thoughtfully integrated into pedagogy.

However, expanding the lens beyond language learning, Abbas and Khan (2024) investigated generative AI usage among university students more broadly. Her study identified academic pressure and time constraints as key drivers behind the adoption of ChatGPT. Nonetheless, over-reliance on the tool was linked to procrastination, memory retention issues, and reduced academic performance. This suggests that while ChatGPT can be helpful, it may also have unintended cognitive consequences if not accompanied by digital literacy and critical awareness.

Together, these studies underscore the transformative potential of ChatGPT in higher education, especially in supporting language-related outcomes. At the same time, they caution against uncritical adoption, emphasizing the need for structured guidance and responsible implementation.

2.5. Research Gaps

Despite increasing attention from Vietnamese and international scholars, current research on ChatGPT in education still reveals several key gaps.

In Vietnam, studies by Nguyen (2024)**,** Nguyen and Phan (2023)**,** and Duong (2023) have focused mainly on theoretical discussions and general observations. However, empirical data assessing ChatGPT's impact on student outcomes or engagement remains limited. Moreover, although ethical concerns are acknowledged, such as the need for usage guidelines, they are not thoroughly investigated.

In contrast, international research such as Mahapatra (2024) and Yıldız (2024) highlights ChatGPT's potential to improve writing and speaking skills. However, similar studies in Vietnamese language classrooms are still lacking. Additionally, Abbas and Khan (2024) point to risks like overreliance and cognitive disengagement, which have not been addressed in the Vietnamese context.

Finally, little is known about how Vietnamese educators perceive and adapt to ChatGPT, especially in traditionally teacher-centred environments.

While prior research explores ChatGPT’s role in higher education, little is known about how its output compares linguistically with learner writing in specific EFL contexts. This study addresses this gap by systematically analysing both sources to reveal convergences and divergences in lexical, syntactic, and pragmatic features.

3. Methodology

This study adopts a mixed methods design, combining quantitative corpus profiling with qualitative manual analysis to examine the linguistic characteristics and pedagogical relevance of ChatGPT-generated texts compared with texts produced by Vietnamese EFL learners. The focus lies not on learning outcomes but on the linguistic nature of the output, analyzed through both automated and manual methods, under SLA theory framework.

3.1. Research Design and Dataset

The research employs a textual analysis approach involving two data sources:

- 20 texts written by Vietnamese university-level EFL learners
- 20 texts generated by ChatGPT (GPT-4)

All texts were written in response to the same 10 common speaking topics, with two samples per topic per source, resulting in 40 short texts (each approximately 80–120 words). The topics include familiar themes such as hometown, a hobby, favourite season, an admired person, a living place (countryside or the city), a memorable trip, favourite type of food, technology, daily routine, and a teacher. These prompts, adapted from the course syllabus, were chosen for their familiarity, moderate complexity, and communicative relevance, aligning with B1–B1+ competence and offering i+1 input as described in SLA theory. Although initially designed for speaking practice, all data here were produced in written form.

Student responses were created during writing-to-speak tasks, where learners drafted short paragraphs to prepare for oral performance. This approach ensured that the dataset consisted entirely of written texts, directly comparable to ChatGPT's written outputs. Participants were recruited from intact classes through convenience sampling, and all eligible students in the target cohort participated. Tasks were completed in class within 15 minutes under teacher supervision, without access to dictionaries, online resources, or AI tools. Prompts were displayed on screen, instructions read aloud, and students submitted their handwritten paragraphs at the end of the activity. All student texts were anonymised before analysis.

ChatGPT texts were generated using the same prompts, with instructions to produce natural, B1-level English paragraphs. This combined approach follows corpus-assisted discourse studies, integrating quantitative lexical analysis to identify salient patterns with qualitative interpretation to capture syntactic, lexical, and pragmatic functions, ensuring both breadth and depth in addressing the research questions.

3.2. Participants

The participants were second-year non-English major undergraduates at a Vietnamese university, aged 19–20, with an estimated B1–B1+ CEFR proficiency. All were enrolled in a four-skill English program with placement testing to ensure similar proficiency levels.

From a class of about 40, 10 students were selected based on full attendance, completion of all in-class writing-to-speak tasks under identical conditions, and consistent B1–B1+ performance in prior assignments. Each produced two written responses on assigned speaking topics, totalling 20 texts. Prior to data collection, participants were briefed on the study's purpose and consent procedures. Texts were anonymised, scanned, and stored securely. The analysis combined AntConc with manual coding, with two raters ensuring consistency through standardised procedures.

3.3. Analytical Approach

This study integrates two complementary methods within a dual-method approach to achieve both a broad quantitative overview and in-depth qualitative insight.

- *Corpus-Based Analysis with AntConc*

The software AntConc (Anthony, 2023) was used to process both datasets and extract key lexical features. Specifically:

+ Type-Token Ratio (TTR) was calculated to assess lexical diversity.

+ The Collocates function identified frequent collocations, focusing on topic-specific terms ("city," "food," "teacher," etc.).

+ Discourse markers and idiomatic expressions were traced using the *Concordance* function to examine how learners and ChatGPT manage cohesion and interaction.

First, AntConc was used to calculate TTR, identify collocations, and trace discourse markers. Next, based on these results, representative samples were selected for manual coding. Codes were grouped into three themes: syntax (complexity, subordination, passive), lexis (collocations, idiomaticity, appropriateness), and pragmatics (register, politeness, coherence). Two raters cross-checked coding to ensure reliability.

The qualitative element lies in this interpretive manual analysis, where patterns detected by the corpus tools were examined in depth for their communicative appropriateness, pragmatic functions, and alignment with SLA principles. This step moved beyond numerical counts to evaluate how language choices fulfilled communicative purposes and reflected learner competence.

3.4. Theoretical Framework

Findings were interpreted through the lens of SLA theory, including:

- Krashen's Input Hypothesis (1982): ChatGPT offers comprehensible input that often meets the i+1 requirement, providing a slightly challenging yet accessible language for learners.

- Schmidt's Noticing Hypothesis (1990): Its output includes salient linguistic features that can support learner noticing and awareness of form.

- Authenticity (Widdowson, 1998; Gilmore, 2007): AI-generated texts demonstrate a degree of contextual and social authenticity, though limitations remain due to the lack of genuine communicative intent.

RQ1 and 2 analyze lexical, syntactic, and pragmatic features; RQ3 relates these to i+1 and noticing; RQ4 considers classroom use via authenticity and pedagogic fit. This ensures the framework directly supports the study's aims.

Manual analysis interpreted corpus patterns in terms of communicative function and SLA principles. The final text was language-checked with ChatGPT (GPT-4) for grammar only; no AI was used for data analysis.

3.5. Ethical Considerations

All student texts were collected with informed consent and anonymized for confidentiality. AI-generated texts were created by the researcher under standardized conditions and used solely for academic research purposes.

The final manuscript was revised for grammar and clarity using ChatGPT (GPT-4, OpenAI), with the author maintaining complete control over content, interpretation, and quality. No AI-generated text was used in the findings or literature review sections.

4. Findings and Discussion

This section integrates results from two complementary methods: corpus-based analysis, which provides quantitative measures such as lexical diversity, collocational frequency, and syntactic variety, and manual linguistic analysis, which offers qualitative insights into pragmatic appropriateness, register, and communicative function. All corpus statistics are reported descriptively to indicate pattern direction; given the small, controlled sample, no inferential claims are made. By presenting these results together, the study highlights both measurable linguistic contrasts and their functional implications.

Table 1 summarizes key differences in lexical diversity, collocational usage, discourse markers, and pragmatic appropriateness between the two data sets to provide a more straightforward overview of the linguistic contrasts.

Table 1

Comparative Analysis of Linguistic Features in ChatGPT vs. Student Texts

Feature	ChatGPT (avg)	Students (avg)	Examples
TTR (Lexical Diversity)	Higher lexical variety	More repetition of words	e.g. "various activities" (GPT) vs. repeated "go, go, go" (S)
Collocations	Frequent and natural	Awkward or literal	"family gathering" (GPT) vs. "play the sport" (S)
Discourse Markers	Advanced and varied	Basic and repetitive	"on the other hand" (GPT) vs. "and/but" (S)
Pragmatic Appropriateness	Consistently appropriate	Sometimes abrupt/direct	"You might want to consider…" (GPT) vs. "I tell you to…" (S)

The comparative analysis between student-generated texts and those produced by ChatGPT (GPT-4) revealed substantial differences in linguistic features, particularly in lexical diversity, syntactic complexity, and discourse cohesion. These findings underscore GPT's advanced linguistic modelling capacity under controlled conditions while simultaneously prompting deeper pedagogical reflection on the role and limitations of generative AI in developing second language competence. The differences observed are quantitative and qualitative, with implications that span vocabulary range, fluency, form-function alignment, and communicative appropriateness.

4.1. Lexical Diversity and TTR

One of the most prominent distinctions lies in lexical diversity, measured by the type–token ratio (TTR). On average, ChatGPT texts recorded a TTR of 0.70, compared with 0.63 in student texts, indicating a broader range of vocabulary and less repetition. This quantitative difference highlights how ChatGPT tends to vary word choice across responses, while learners frequently reused a limited set of high-frequency verbs such as *go, like, have,* and *do.*

Table 2

Comparison of Lexical Diversity (TTR) in Student vs. ChatGPT Texts

Group	Average TTR	Common Pattern Observed
Student texts	0.63	Frequent repetition of high-frequency verbs
ChatGPT texts	0.70	Wider lexical variety and fewer repeated forms

Beyond numerical evidence, the manual coding revealed two clear themes:

- *Theme 1:* Over-reliance on simple verbs (Students). Many student texts recycled basic verbs in different contexts, producing expressions that lacked variety, such as *"I go to play football and I go to the park with my friends".*

- *Theme 2:* Broader lexical repertoire (ChatGPT). ChatGPT texts typically included synonyms and more precise verbs, which contributed to fluency and stylistic variation. For example, *"I enjoy participating in various activities such as hiking or playing football with my classmates."*

Taken together, these findings suggest that ChatGPT writing provides richer lexical input that may support learners in noticing new vocabulary choices, whereas student writing demonstrates developmental tendencies such as reliance on a narrow verb set. This aligns with Krashen's (1982) view that comprehensible input at the i+1 level is necessary for acquisition, and supports Schmidt's (1990) argument that noticing novel forms is crucial for vocabulary development. Similar benefits of ChatGPT use for enhancing lexical resources have been observed in recent studies on academic writing (Mahapatra, 2024) and in Vietnamese higher

education contexts (Nguyen & Phan, 2023; Nguyen, 2024).

4.2. Collocational Range and Lexical Naturalness

Beyond lexical diversity, collocational accuracy and idiomaticity emerged as another key area of contrast. ChatGPT texts demonstrated a wider mastery of lexical bundles and greater naturalness in word combinations, whereas student writing often relied on literal or awkward pairings.

Table 3

Comparison of Collocational Use in Student vs. ChatGPT Texts

Group	Average collocations per 100 words	Typical Pattern	Examples
Student texts	7–9	Literal or awkward pairings	"play the sport," "do homework hard"
ChatGPT texts	15–18	Native-like and idiomatic combinations	"digital learning tools," "effective communication," "family gathering"

As shown in Table 3, ChatGPT texts contained considerably more accurate collocations than student texts. Since collocations occur fairly often, figures are reported per 100 words rather than per 1,000 to make the density across the two groups directly comparable. The manual coding further highlighted two main themes:

- *Theme 1:* Literal collocations in student writing. Learners often produced combinations that were grammatically correct but sounded unnatural. *Illustrations*: *"do homework hard," "play the sport", "eat delicious food."*

- *Theme 2:* Idiomatic and contextually appropriate collocations in ChatGPT texts. These texts frequently employed well-formed lexical bundles that enhanced cohesion and credibility. *Illustrations*: *"digital learning tools", "effective communication", "family gathering"*.

This gap in collocational awareness is pedagogically significant, as research (Gilmore, 2007) has emphasized the importance of exposing learners to authentic lexical bundles to build idiomatic competence and enhance fluency. Recognizing and using such collocations fluently marks an important step in the progression from intermediate to advanced language proficiency.

4.3. Syntactic and Discourse Complexity

From a structural standpoint, ChatGPT's syntactic range is broader and more sophisticated. ChatGPT frequently utilizes compound-complex sentence forms, passive voice, and subordinate clauses with embedded logical connectors. For example, constructions such as *"Although it is a small town, it offers many cultural activities"* and *"If I had the chance, I would move back immediately"* are common in GPT texts. These structures contribute to more nuanced, precise, and coherent arguments.

In comparison, student texts tend to rely heavily on simple and compound sentences, often with repetitive syntactic patterns. This limited structural variation affects readability and restricts the expression of more complex ideas. Furthermore, sentence transitions in student texts tend to be abrupt, relying primarily on basic connectors such as *and*, *so*,and *because*, with limited use of more advanced cohesive devices. In contrast, ChatGPT texts demonstrate a much wider array of discourse markers, including *moreover, on the other hand, consequently,* and *in contrast,* helping to create logical flow and textual unity.

Table 4

Comparison of Syntactic and Discourse Features in Student vs. ChatGPT Texts

Group	Complex/compound–complex sentences (%)	Range of discourse markers (number of types)	Typical examples
Student texts	15–18%	3–4	and, so, because
ChatGPT texts	35–40%	10–12	moreover, consequently, in contrast

As shown in Table 4, ChatGPT texts contained a significantly higher proportion of complex and compound–complex sentences (35–40%) than student texts (15–18%). They also employed a wider range of discourse markers (10–12 types compared with only 3–4), which enhanced logical flow and cohesion. The percentage of complex sentences was calculated relative to the total number of sentences, while the range of discourse markers reflects the number of distinct types observed in each dataset. Manual coding revealed two recurrent themes:

- *Theme 1:* Overuse of simple structures in student writing. Learners often chain short, paratactic sentences, limiting nuance, such as *"I like summer. I go swimming. I feel happy"*.

- *Theme 2:* Varied and cohesive structures in ChatGPT writing. Subordination and embedded connectors supported more apparent logic. For example: *"Although it is a small town, it offers many cultural activities, and if I had the chance, I would move back immediately"*.

These findings suggest that student writing reflects developmental reliance on simple syntax and limited cohesive devices, while ChatGPT writing models more advanced structural variety and discourse cohesion. From an SLA perspective, exposure to such structurally complex and cohesive input could support learners' noticing of new forms (Schmidt, 1990) and foster gradual development in syntactic competence.

4.4. Pragmatic Appropriateness and Register Control

Another area of divergence lies in pragmatic control and register appropriateness.

Quantitative analysis showed that approximately 45–50% of ChatGPT sentences contained hedging or polite expressions (e.g., *might, perhaps, I would recommend*), while fewer than 10% of student sentences used such forms. Conversely, over 60% of student texts included abrupt or overly direct expressions, compared with less than 15% in ChatGPT texts.

Table 5

Pragmatic and Register Features in Student vs. ChatGPT Texts

Group	Use of hedging/polite devices (%)	Use of abrupt/direct forms (%)	Examples
Student texts	<10%	~60%	*I write this email to you*; *I tell you to…*
ChatGPT texts	45–50%	<15%	*You might want to consider…*; *Perhaps you could…*

Manual coding revealed two recurrent themes:

- *Theme 1:* Directness and lack of softening in student writing. Learners often used abrupt statements or blunt advice without pragmatic mitigation, such as *"I write this email to you"*.

- *Theme 2:* Pragmatic softening and register awareness in ChatGPT texts. These texts

frequently employed hedging and polite strategies to align with communicative conventions. For example: *"You might want to consider joining a club"*.

This observation aligns with Widdowson's (1998) view that authenticity is context-dependent, and effective language use depends not merely on grammatical accuracy but on the appropriate deployment of language in socially situated interactions. It also resonates with Nguyen's (2024) metaphor-based analysis of how AI is framed in public discourse. By portraying AI tools like ChatGPT through human-like or naturalistic metaphors, media narratives may influence learners' perceptions of the tool's communicative authority and naturalness. As a result, students may assume that AI-generated responses are pragmatically flawless, leading to uncritical adoption without awareness of nuanced sociolinguistic norms.

Therefore, pedagogical approaches should emphasize what ChatGPT can produce and how and why it produces specific registers and tones. This includes helping learners reflect on pragmatic appropriateness about purpose, audience, and context and using AI output as a springboard for discussing politeness strategies, cultural conventions, and genre expectations. From an SLA perspective, the outputs demonstrate textual authenticity (accurate, idiomatic, cohesive) but lack interactional authenticity (Widdowson, 1998; Van Lier, 1996). Authentic-like input alone is insufficient; it should be combined with interactive tasks to support language acquisition fully.

4.5. Idiomaticity and Expressive Language

Interestingly, idioms and figurative language were rare in both sets of texts, likely due to the nature of the prompts and the learners' proficiency level.

Table 6

Idiomatic and Figurative Expressions in Student vs. ChatGPT Texts

Group	Average idiomatic/semi-fixed expressions per 1000 words	Examples
Student texts	0–1	(rare or absent)
ChatGPT texts	2–3	stay on track, take responsibility

As shown in Table 6, ChatGPT texts averaged about 2–3 semi-fixed idiomatic expressions per 1,000 words, while student texts contained almost none. Because idioms occurred very infrequently, the counts were normalised per 1,000 words to make the comparison meaningful. Manual coding revealed two themes:

- *Theme 1:* Scarcity of idiomaticity across both groups. Neither corpus relied heavily on idioms, reflecting the influence of task type and B1–B1+ proficiency level.

- *Theme 2:* Occasional idiomaticity in ChatGPT writing. While rare, ChatGPT occasionally used semi-fixed phrases that added subtle stylistic colour, such as *"stay on track", "take responsibility"*.

While idioms were not a primary feature, the presence of a few idiomatic phrases in ChatGPT texts suggests broader exposure to naturalistic input and the ability to generate contextually appropriate language patterns that resemble real-world usage. For learners, contrasting these outputs with their own writing could raise awareness of idiomatic expressions as an advanced feature of fluency. This finding is consistent with previous research showing that idiomaticity is closely linked to authenticity in language use (Gilmore, 2007) and typically emerges at higher proficiency levels. From an SLA perspective, the scarcity of idioms indicates that such features may fall beyond the immediate i+1 input accessible to B1–B1+ learners

(Krashen, 1982), and that explicit noticing (Schmidt, 1990) may be required for learners to integrate them effectively into their writing.

5. Pedagogical Implications and Limitations

The analysis points to several implications for English language teaching, especially with learners at the intermediate level. The generated texts are best viewed as a helpful supplement rather than a replacement for student writing. They provide exposure to language that is often just beyond what learners can produce on their own, creating opportunities to notice new words, structures, and pragmatic features.

One clear implication comes from the lexical findings. The richer vocabulary and more natural collocations found in the generated texts can be used in class activities that draw learners' attention to word choice. Simple tasks such as synonym matching, collocation awareness exercises, or short gap-fills can help students reflect on differences between their own writing and the more varied expressions in the model texts.

Syntactic variety is another area with clear classroom potential. Because the generated texts contain a higher proportion of complex sentences and a broader range of cohesive devices, they can serve as models for rewriting or expansion activities. Asking students to recast their own short sentences into more complex ones, or to insert appropriate discourse markers, can encourage a greater sense of structural flexibility and coherence.

Pragmatic appropriateness also offers fertile ground for instruction. The model texts frequently used hedging and polite expressions, whereas student writing tended to be abrupt or overly direct. Highlighting these contrasts through role-plays, peer review, or short reflective tasks could raise awareness of how tone and register shape communication. Even the occasional idiomatic expressions in the generated texts, while not frequent, provide a chance to discuss how figurative language adds nuance and fluency.

To make the most of these opportunities, teacher mediation is essential. The texts should not be presented as perfect models to copy but as resources to analyse, adapt, and sometimes even "humanise". Activities such as rewriting them to match the learner's own voice, or critiquing their style and tone, can promote both creativity and critical thinking while reducing the risk of over-reliance.

At the same time, these implications should be treated with caution. The study did not investigate how learners themselves respond to such texts or whether exposure leads to measurable gains in their writing. The focus was limited to B1–B1+ Vietnamese university students, so the findings may not extend to other levels or contexts. Moreover, the analysis was based solely on written data, without classroom observation or teacher perspectives. Future work would benefit from triangulating text analysis with learner perceptions and classroom practice to provide a fuller picture of how such materials can be integrated effectively.

6. Conclusion

This study compared texts written by Vietnamese EFL students with those produced by ChatGPT (GPT-4), focusing on vocabulary use, sentence structure, and discourse organisation. The generated texts showed a higher type-token ratio (0.70 vs. 0.63), more accurate and idiomatic collocations, and a wider range of discourse markers (10-12 types compared with 3-4 in student texts). These differences made them closer to model writing samples and reflected greater cohesion and stylistic variety.

The findings also connect to second language acquisition theories. ChatGPT's capacity to provide comprehensible yet slightly advanced input resonates with Krashen's Input Hypothesis, while the salience of new phrases and structures supports Schmidt's Noticing Hypothesis. The more cohesive and pragmatically appropriate style observed in the generated texts further reflects Widdowson's view of authenticity as socially meaningful communication. At the same time, the student texts revealed developmental patterns, such as over-reliance on simple verbs, limited connectors, and abrupt pragmatics, highlighting areas where targeted pedagogical support is needed.

Nevertheless, some limitations remain. The generated texts occasionally misrepresented tone, formality, or cultural nuance, particularly in more personal or affective contexts. For this reason, teacher mediation is essential. Classroom activities such as comparing student and generated texts, revising them to fit learners' voices, or analysing lexical and pragmatic choices can encourage critical engagement and reduce the risk of uncritical adoption.

In short, ChatGPT can provide valuable supplementary input when used with teacher guidance and learner reflection. Future research should investigate the longer-term effects of integrating generated texts into writing instruction and explore how different levels of teacher support influence outcomes. With careful design, such integration can enrich classroom practice while keeping the focus firmly on active learner production.

References

Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education, 21*(1), 1-22. https://doi.org/10.1186/s41239-024-00444-7

Anthony, L. (2023). *AntConc* (Version 4.2.4) [Computer software]. https://www.laurenceanthony.net/software/antconc/releases/AntConc424/help.pdf

Duong, T. L. (2023). Application of ChatGPT in promoting higher education teaching and learning in the era of Artificial Intelligence. *Journal of Science and Technology – Binh Duong University*, *6*(2), 153–160. https://doi.org/10.56097/binhduonguniversityjournalofscienceandtechnology.v6i2.160

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching, 40*(2), 97–118. https://doi.org/10.1017/S0261444807004144

Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology, 15*(2), ep421. https://doi.org/10.30935/cedtech/13036

Ishida, H. (2006). Learners' perception and interpretation of contextualization cues in spontaneous Japanese conversation: Back-channel cue "uun." *Journal of Pragmatics, 38*(11), 1943–1981. https://doi.org/10.1016/j.pragma.2005.08.004

Khlaif, Z. (2024). Rethinking educational assessment in the age of artificial intelligence: Insights from recent training workshops. In M. Sanmugam, B. Edwards, N. Mohd Barkhaya & Z. Khlaif (Eds.), *Fostering Inclusive Education With AI and Emerging Technologies* (pp. 131-144). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-7255-5.ch005

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537-550. https://doi.org/10.1177/00336882231162868.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press. http://www.sdkrashen.com/content/books/principles_and_practice.pdf

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press. https://www.scribd.com/document/253376703/Long-1996

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments, 11*(1), Article 9. https://doi.org/10.1186/s40561-024-00295-9

Nguyen, D. V., & Phan, T. T. (2023). Application of ChatGPT in teaching, research and management at the university level. *Scientific Journal of Tan Trao University, Social Science and Humanities*, *9*(3), 170–182. https://doi.org/10.51453/2354-1431/2023/974.

Nguyen, P. Q. (2024). The impact of ChatGPT on science and education. *Journal of Science - Hong Bang International University, 30*, 149-158. https://doi.org/10.59294/HIUJS.30.2024.653

Nguyen, T. M. (2024). Conceptual metaphors of artificial intelligence and AI development in *The Guardian* newspaper. *VNU Journal of Foreign Studies, 40*(4), 128-141. https://doi.org/10.63023/2525-2445/jfs.ulis.5345

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy and authenticity*. Longman. https://books.google.com.vn/books?id=cWq4AwAAQBAJ

Widdowson, H. G. (1998). Context, community, and authentic language. *TESOL Quarterly, 32*(4), 705–716. https://doi.org/10.2307/3588001

Xia, Q., Weng, X., Ouyang, F., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, *20*(68). https://doi.org/10.1186/s41239-024-00468-z

Yıldız, C. (2024). ChatGPT Integration in EFL Education: A Path to Enhanced Speaking Self-Efficacy. *Novitas-ROYAL (Research on Youth and Language)*, *18*(2), 167–182. https://files.eric.ed.gov/fulltext/EJ1446766.pdf

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*, 1-37. https://doi.org/10.1186/s40561-024-00316-7