Journal of Educational Technology and Innovation
https://jeti.thewsu.org/index.php/cieti

Article

A Corpus-Based Analysis of Verb Collocations in Human and AI-Generated IELTS Writing

Meirong Du [1], Min Lu [2], Yi Dai [1,*] and Fan Wang [2]

[1] School of Education, City University of Macau, Macau, China
[2] University International College, Macau University of Science and Technology, Macau, China
* Correspondence: yidai@cityu.edu.mo

How To Cite: Du, M., Lu, M., Dai, Y., & Wang, F. (2025). A Corpus-Based Analysis of Verb Collocations in Human and AI-Generated IELTS Writing. *Journal of Educational Technology and Innovation*, *7*(2), 67–80. https://doi.org/10.61414/mzmmrq74

Received: 8 May 2025
Revised: 20 June 2025
Accepted: 27 June 2025
Published: 30 June 2025

Abstract: In this study, it aims at examining the differences between human-generated and AI-generated texts in IELTS Writing Task 2. It especially focuses on lexical resourcefulness, grammatical accuracy, and contextual appropriateness. We analyzed 20 essays, including 10 human written ones by Chinese university students who have achieved an IELTS writing score ranging from 5.5 to 6.0, and 10 ChatGPT-4 Turbo-generated ones, using a mixed-methods approach, through corpus-based tools (NLTK, SpaCy, AntConc) and qualitative content analysis. Results showed that AI texts exhibited superior grammatical accuracy (0.4%–3% error rates for AI vs. 20–26% for university students) but higher lexical repetition (17.2% to 23.25% for AI vs. 17.68% for university students) and weaker contextual adaptability (3.33/10–3.69/10 for AI vs. 3.23/10 to 4.14/10 for university students). While AI's grammatical precision supports its utility as a corrective tool, human writers outperformed AI in lexical diversity and task-specific nuance. The findings advocate for a hybrid pedagogical model that leverages AI's strengths in error detection while retaining human instruction for advanced lexical and contextual skills. Limitations include the small corpus and single-AI-model focus, suggesting future research with diverse datasets and longitudinal designs.

Keywords: AI-generated writing; IELTS Writing Task 2; verb collocations

1. Introduction

In recent years, the Artificial Intelligence (AI) has been integrated into language education, which has garnered significant attention, particularly in the context of high-stakes language assessments such as the International English Language Testing System. As AI technologies' wide application in English writing, there is a growing need to understand the differences between human-generated and AI-generated texts, especially in terms of lexical resourcefulness and grammatical accuracy. In this context, this study aims to examine the differences in verb collocations, lexical repetition, grammatical accuracy, and contextual appropriateness between AI-generated and human-written IELTS writing tasks.

The importance of lexical resources in IELTS writing has been repeatedly emphasized in the past decades and it has been officially noted that a rich and varied vocabulary is crucial for achieving high scores (IDP IELTS, n.d., Luo et al., 2023). Lexical resourcefulness not only reflects a candidate's vocabulary size but also their ability to clearly and coherently express complex ideas (Li & Schmitt, 2009). However, the assessment of lexical resources in IELTS writing faces challenges due to its subjective nature and the potential for variability among examiners (Arefsadr, Babaii, & Hashemi, 2022). Besides, previous studies have shown that IELTS candidates often struggle with lexical repetition (Pearson, 2019). To finish the writing task within the limited time, they tend to use simple verb forms, or inappropriately use reporting verbs, which can negatively impact their testing scores.

The advent of AI has introduced new possibilities and challenges in language education (Rahmouni, 2025). AI-generated texts have been shown to excel in grammatical accuracy and coherence but may lack the contextual nuances and creativity of writing pieces, (Jasim & Awqati, 2025), including IELTS writings. However, a cross-sectional study by Jahan, Arif, and Mustafa (2024), conducted over one month, found that with the support of ChatGPT, a well-known AI tool, 10 university students majoring in English showed significant improvements in their grammatical basics. Although AI-generated texts are increasingly used in language learning, research comparing them to human-written texts, particularly in terms of lexical diversity and contextual appropriateness within the IELTS writing context, remains scarce.

Recent studies in AI and language education have begun to examine both the advantages and challenges of integrating AI into language learning and instruction. For example, Hafeez, Ajmail, and Zulfiqar (2025) state that AI tools have impelled the creation and implementation of effective teaching resources, eventually contributing to improved educational quality. This study demonstrates AI's strengths in delivering personalized learning plans. Similarly, Ilieva, Yankova, Ruseva, & Kabaivanov (2025) proposed a new framework for generative AI-supported assessment in higher education in their study, which explains how teaching staff can design adaptive, AI-based tasks and provide feedback.

In the context of IELTS writing, a study by Samsidar, Fauzi, Karuppiah, and Patak (2022) found there is a positive correlation between lexical resourcefulness and writing performance, that is candidates who have a wide range of vocabulary are more competent in using it accurately and appropriately. However, the study also noted that students' overall writing performance will be negatively impacted if they face challenges in maintaining lexical diversity and contextual appropriateness. With the support of AI, students' lexical development and grammatical accuracy can be improved. However, AI's limitations in capturing the complexity and diversity of human language have addressed the need for targeted pedagogical strategies.

This study seeks to fill the gap in the literature by conducting a detailed comparative analysis of human and AI-generated IELTS writing samples from three aspects: lexical repetition, grammatical accuracy, and contextual appropriateness. On this basis, this study will provide insights into the specific differences between human and AI-generated texts and their implications for IELTS writing assessment.

Specifically, this study seeks to address the following research questions:

RQ1. In what ways do human-written and AI-generated texts differ when it comes to the diversity, accuracy, and contextual use of verb collocations?

RQ2. How might these distinctions influence how the 'Lexical Resource' criterion is evaluated in IELTS writing assessments?

By shedding light on these differences, this study aims to add valuable insights to the ongoing conversation about AI's role in language education and support the development of practical teaching strategies that effectively incorporate AI into language learning environments.

2. Literature Review

2.1. Lexical Resource in IELTS Writing

Lexical resources are embodied as "web components that implement a minimal set of predefined programming interfaces" (Piasecki et al., 2018). Vocabulary use plays an instrumental role in reflecting a candidate's linguistic competence within the IELTS writing assessment, particularly under the criterion known as "Lexical Resource". As outlined in the official IELTS Writing Band Descriptors (2024), this component evaluates how effectively candidates utilize vocabulary in terms of variety, precision, and appropriateness. Candidates aiming for higher band scores are expected not only to demonstrate lexical range, including the use of less frequent expressions, but also to apply vocabulary naturally, with fewer repetitions. Moreover, lexical proficiency involves accurate word choice that conveys intended meaning clearly while minimizing grammatical and semantic errors (Saadatara et al., 2022).

Within the IELTS writing assessment, the "Lexical Resource" criterion plays a central role in evaluating a candidate's language proficiency.Through English writing, the candidates' breadth of words and their ability to precisely use words can be measured; besides, their ability to express and organize ideas clearly in English can be evaluated. As Allami et al. (2025c) have noted that there is a strong positive relationship exists between lexical richness and writing performance, candidates who possess a broad and accurate vocabulary are more capable of making arguments and expressing complex thoughts. As a result, the overall coherence and effectiveness of their writing were promoted (Alavi & Masjedlou, 2017).

Although "Lexical Resource" is a critical component in the IELTS writing rubric, its evaluation is not without complications. One of the key difficulties lies in the subjective nature of the scoring process, as examiners may differ in their interpretations of what constitutes sufficient lexical range and accuracy (Eckes, 2008). In addition, many test-takers struggle with improper word choice, register mismatches, and grammatical inaccuracies, all of which can lead to their average performance (L. Q. Nguyen & Van Le, 2022). Worsening the situation, research has shown that a considerable number of candidates tend to depend on overly simplistic vocabulary and formulaic language, which limits their opportunity to demonstrate more advanced lexical proficiency (Amirian et al., 2016).

2.2. Current State of Verb Collocations in IELTS Writing by Human Students

Verb collocations hold comparable weight to other lexical features in IELTS writing, as they reflect a candidate's command of vocabulary and syntactic accuracy. These combinations, typically involving verbs paired with nouns, adjectives, or prepositions, are fundamental to producing natural and fluent writing, which is crucial for achieving high scores in IELTS writing tasks. Nonetheless, research indicates that learners often struggle with the appropriate use of such collocations, a challenge that contributes to reduced writing performance (Estaji & Hashemi, 2022).

A major issue in IELTS writing is the inconsistent use of verb tenses. Many candidates find it difficult to keep their tenses consistent across the essay. This often causes confusion for readers and contributes to lower scores. Moreover, subject-verb disagreement is common. The misuse of modal verbs is also frequently seen. These grammatical problems reduce the overall clarity of the writing. Another issue is the overuse of simple verb forms throughout the essay. Due to limited vocabulary, many students rely heavily on basic verbs. This repetition can negatively impact their score in the "Lexical Resource" category of the IELTS writing assessment. Beyond that, many candidates misuse conditional structures. Such errors make it harder to express complex ideas effectively.

Reporting verbs form a crucial subset of verb collocations, especially in academic writing like IELTS Writing Task 2. These verbs help introduce and discuss others' ideas. Using them correctly can improve an essay's coherence and persuasiveness. However, research shows that many candidates find it difficult to choose and use the right reporting verbs (A. S. Mohammed et al., 2020). For instance, many overuse general verbs such as "say" or "think", instead of alternative verbs like "argue", "claim", or "suggest".

Educators and researchers recommend several teaching strategies to tackle these issues. First, verb collocations and reporting words are explicitly explained and given special attention in English writing lessons; besides, a variety of synonymous verbs and their correct collocations are introduced in classes to improve students' ability to use vocabulary. Beyond that, educators purposely design targeted practice activities and feedback sessions to help students accurately understand verbs and their usage in sentences. Furthermore, guidance on verb tense and subject-verb agreement is provided to boost grammatical correctness (Suraprajit, 2021). For example, students are encouraged to consult resources like the AMA Manual of Style or the EASE Guidelines. These materials teach proper verb tense usage in academic writing.

To begin with, direct teaching of verb collocations and reporting verbs is of great value. Learners need exposure to a broad selection of synonym verbs along with their correct combinations. This helps develop their lexical flexibility. Beyond that, structured exercises and follow-up feedback are used to strengthen students' knowledge of these expressions. These tasks aim to improve both accuracy and appropriateness in students' use of verbs. Students also receive clear guidance on grammar rules, particularly verb tense consistency and subject-verb agreement (Suraprajit, 2021).

In summary, mastering verb collocations is essential for success in IELTS writing. Both learners and instructors need to recognize its importance. Common issues, like inconsistent verb tense, frequent use of basic verbs, and incorrect application of reporting verbs, must be tackled directly. Addressing these problems can boost students' vocabulary use and improve grammatical precision (Ahmed & Mohammed, 2024). As a result, their overall writing performance on the IELTS exam is likely to improve.

2.3. Comparison of Human and AI Language Use

The wide and in-depth use of artificial intelligence (AI) is reshaping language education. This shift has sparked growing interest in how AI-generated language compares with human writing (Karataş et al., 2024). Recent research has focused on tools like automated writing assessment and natural language processing. It is because these tools are changing how students learn and teachers instruct. Comparing AI and human output helps identify what AI does well and where it falls short. Understanding this is key to building better teaching strategies in language classrooms. Earlier studies have pointed out both overlaps and distinctions in how language is used by humans and machines. A consistent observation is that AI tends to produce texts that are grammatically sound

and logically structured (Maeda, 2024). However, it often falls short in capturing clear meanings in certain contexts and delivering creative expressions. For example, AI-generated writing may miss the richness found in idioms or cultural references. However, machines can generate content far more quickly than humans. In some educational settings, this effectiveness is valuable.

AI tools present both advantages and drawbacks. They, on one hand, offer quick feedback and tailored learning, which can support language development (Park, 2019). On the other hand, the overuse of AI tools have caused concerns among some scholars. They thought heavily relying on AI might reduce chances for real communication and genuine language practice. From the perspective of content quality, texts produced by AI often lack the richness and flexibility compared with human language. As a result, AI-generated texts are quite rigid and formulated (Perdana et al., 2021).

A blended model has been recommended to make the most of AI while reducing its defects (Oved & Alt, 2025). As this method mixes AI tools with conventional classroom teaching, it allows students to benefit from instant feedback and personalized support. At the same time, it ensures they still use language in real contexts and get direct and quick help from teachers. To well use this mixed method, students are encouraged to learn AI technology. But it is worth emphasizing that AI is just a support tool for learning. This approach makes instant feedback and personalized self-learning possible. At the same time, it ensures students practice authentic language use and receive direct support from human teachers. Additionally, teaching students about AI is crucial. They need guidance on how to use AI tools properly to aid their language learning instead of replacing it. This approach can amplify its strengths in providing instant feedback and personalized learning while ensuring that students engage in authentic language use and receive personalized guidance from human instructors. Furthermore, it is important to provide students with AI knowledge and instruct them to appropriately use AI tools to support their language learning rather than replacing it.

The comparison of human and AI language use reveals both potential benefits and challenges in the context of language education. While AI can significantly enhance language learning through personalized feedback and efficient text generation, it is essential to address its limitations in handling context-specific nuances and promoting authentic language use (Zhu & Wang, 2025). Future research should focus on developing more sophisticated AI models capable of capturing the complexity of human language and exploring effective pedagogical strategies for integrating AI into language education.

2.4. Student Acceptance and Feedback on AI Assistants vs. Human Teachers

The wide use of AI in language learning has prompted researchers and scholars to pay attention to relevant topics. Specifically, they have begun to investigate compared to real-world teachers, how students view and react to AI in their studies. Rienties et al. (2025) focused on students' acceptance, preferences, and feedback in language learning settings. Through studies like this, both the strengths and the drawbacks of using AI in the classroom have been disclosed.

It has been found that most students prefer a mixed approach of AI and in-person instructors compared with entirely relying on AI tools or human teachers (Peng & Wan, 2023). For example, in a study conducted by Lingoda, researchers have found that students, regardless of their gender, tend to favor a mix of AI and human tutoring. Male students, in particular, showed higher preference for this blended learning model. In contrast, only a small group of participants, namely 28% of males and 30% of females, chose to rely solely on AI in their leaning. These results indicate that although AI is a valuable tool in education, human engagement still plays an equally important role (Peng & Wan, 2023). And when it refers to students' perception of the role of AI tools and human instructors in their learning, previous studies showed that a majority of investigated students trust human teachers more, especially when they encounter complex tasks (Dwivedi et al., 2023). According to Kim et al. (2020), many students saw AI tools as less reliable and less helpful in giving feedback than their teachers. Even so, students still recognized certain strengths of AI, such as its quick response to questions and its extensive knowledge reserve.

Some students indicate that in certain scenarios, AI tools are quite effective; while others argue that AI tools are less effective in understanding implications of some cultural contexts and meeting learners' emotional needs in learning. Unlike human teachers, AI cannot offer empathy or help establish personal connections. In Escalante et al. (2023b)'s study, they found that teachers' feedback, especially the positive one, can inspire students whereas they cannot be motivated by AI even though they receive positive feedback from AI. Even so, learners valued the quick and consistent responses from AI. While AI is useful for fast and consistent support, it cannot fully replace the depth and personal touch provided by human educators.

Both the learning experience and the final outcome hold equal importance. Hence, the learning outcomes of AI tools and human teachers have caught researchers' attention. In earlier studies, it has been indicated that using

both AI and human instruction together can lead to better outcomes than relying on one alone. Yang (2025d) noted that when teachers and AI tools work together, students become more engaged and take a more active role in learning. These studies have proved that AI serves as a useful tool in practical classroom environments.

The respective strengths of AI tools and human teachers have made students believe that a mixed approach of AI and human teachers is more beneficial in their learning. As mentioned previously, AI offers fast, tailored, knowledgeable feedback; while human teachers bring emotional support, inspire in-depth thinking, and provide more enlightening interaction. Hence, it is necessary to explore how to better combine both elements in the future. In this way, the learning outcomes can be improved and the varied needs of students can be met (Spence et al., 2024).

2.5. Research Gaps

The use of artificial intelligence (AI) in language instruction has drawn increasing attention from the academic community. But a number of urgent problems are still unsolved. First, research on the linguistic distinctions between AI-generated writings and human-written essays, such as verb collocations, is conspicuously lacking, especially when it comes to the "Lexical Resource" criterion in the IELTS writing assessment. According to previous research (Hyland, 2020; Veerappan & Sulaiman, 2012), AI-generated texts frequently seem formulaic and uncreative in comparison to the lexical variety and complexity displayed by human writers.

Second, there are three primary limitations to the current research. First, there isn't a solid foundation for comparing texts produced by AI and humans in a methodical manner. Second, there are currently no longitudinal studies examining the long-term impacts of AI-assisted learning on students' lexical development. Third, nothing is known about the pedagogical shifts and ethical issues raised by AI in high-stakes language tests. It's important to remember that while AI can generate papers that are grammatically correct, it still has trouble with semantic subtleties and creative expression in certain situations, two things that are essential for getting good IELTS writing scores.

Future research should concentrate on three areas in light of these gaps: creating rigorous methodologies for comparing writing produced by humans and artificial intelligence, planning long-term empirical investigations, and carrying out a thorough assessment of the societal effects of AI-assisted language training. These initiatives will advance knowledge of the usefulness and possible effects of AI in language instruction.

3. Methodology

3.1. Research Design

This study employs a quantitative approach to provide a comprehensive understanding of the differences between human and AI-generated IELTS writing samples. This design allows for a detailed examination of lexical resourcefulness, particularly focusing on verb collocations, and their impact on IELTS writing scores. To capture both the measurable differences in lexical usage and the nuanced qualitative aspects of writing, the mixed-methods approach is chosen in this study (Creswell, 1994).

3.2. Data Collection

In this study, a corpus of 20 IELTS Writing Task 2 mock test essays was included. The student-written samples were collected from a language training institution in Zhuhai, a city in southeastern China. The sample includes 10 essays written by university students ranging from first to third year. 5 of the essays received an IELTS writing band score of 5.5, while the remaining 5 were rated at band 6.0 in writing. Each essay contains approximately 500 words, with a total corpus of 5,000 words. The 10 essays were deliberately selected from a pool of 30 responses to the same writing question tested over the past year. This selection ensured a balanced dataset, with equal representation of essays at both the 5.5 and 6.0 band levels. Besides, the purposive sampling method was also utilized to pick the credible IELTS training institution in Zhuhai through examining its official IELTS institution code issued by British Council and the recommendation from IELTS experts. In this way, the reliability and credibility of the writing samples have been guaranteed.

As students at this training center regularly took IELTS mock exams, and their essays were previously included in the human-written dataset, the relevance and adequacy of data have been ensured. The selection criteria assures the diversity of proficiency levels and writing styles, providing a representative sample of typical IELTS candidates. To create a comparable AI-generated dataset, we input each of the 10 IELTS writing samples, originally produced by 10 selected students, into ChatGPT-4o Turbo, a state-of-the-art AI text generation model. Together with each writing sample, we provided the same IELTS writing question along with the target band score that matched the given sample's score, ensuring alignment between the AI-generated and human-written responses. When ChatGPT generated works, IELTS experts, three teachers with at least five-year experience in

teaching IELTS writing, at this training center assessed whether the bands are consistent with the target band, attempting to keep the data from human and AI sources as consistent as possible.To make sure that the AI-generated replies could be fairly compared with those created by university students, researchers used the same IELTS writing prompts and goal bands. This method made it possible to evaluate the AI's performance in comparison to human norms with greater accuracy. The AI's ability to meet the same band criteria as those specified by IELTS examiners was confirmed by the IELTS experts' assessment of the AI-generated works, which offered important insights into the AI's potential and skills for English writing assessment.

Ethical approval was obtained from the institution's review board, and informed consent was secured from all participants. The data collection process ensured anonymity and confidentiality, with all identifying information removed from the essays. The use of AI-generated texts was also reviewed to ensure compliance with ethical guidelines, particularly regarding the generation of content that could be used for educational purposes (Bender et al., 2021).

3.3. Data Analysis

Stage 1: Verb Collocation Extraction

The first step was extracting verb collocations from writings produced by AI and by humans. This was accomplished by combining two potent natural language processing packages, SpaCy and Natural Language Toolkit (NLTK). Potential verb collocations (nouns, verbs, and prepositions) were found using the BigramCollocationFinder from NLTK and the dependency parsing capabilities of SpaCy. Tokenisation, part-of-speech tagging, and filtering to eliminate non-lexical parts were among the data pre-processing procedures used in the extraction process (Bird et al., 2009).

Stage 2: Corpus Comparison Analysis

AntConc 4.2.0, a popular corpus analysis tool, was used to examine the retrieved collocations (Anthony, 2022). Table 1 displays baseline data that describe the study's corpus. The frequency and distribution of verb collocations in texts produced by AI and humans were compared using AntConc's keyness analysis and collocates network properties.The collocates network displayed the co-occurrence patterns, and the keyness analysis revealed notable variations in collocation utilization. This analysis brought to light possible variations in the two data sets' lexical resources.

Table 1. Details of corpus.

	Writing Band Score	Words Tokens	Mean Length of per Texts
From Human	5.5	1350	270
	6.0	1369	273.8
From AI	5.5	1508	301.6
	6.0	1541	308.2

Stage 3: Qualitative Content Analysis

To investigate the context-specific subtleties and inventiveness of verb collocations, a qualitative analysis was carried out. In particular, template-based repeats in AI-generated texts and novel collocations specific to human-generated texts were found. Additionally, the study identifies contextual and cultural inconsistencies in AI-generated information, which may affect how lexical resourcefulness is viewed. Beyond just frequency counts, this qualitative method offered more profound insights into the variations in vocabulary usage (Biber et al., 2020).

4. Results

4.1. Lexical Repetition

At IELTS writing bands 5.5 and 6.0, the research showed clear differences in lexical recurrence between texts created by humans and AI. The average lexical repetition rate for first- to third-year students at Band 5.5 was 17.68%, with significant individual variation. For example, the lexical repetition for Student 3 is 28% and the students used 'save money' for four times and develop habits for three times (see Figure 1). Such elevated repetition rates suggest their constrained lexical resources, which may limit their ability to effectively articulate complex ideas (Liu & Shi, 2022). In contrast, students at Band 6 demonstrated a significantly lower repetition (average: 12.8%), with a high-performing student achieving 0% repetition (e.g., Student 9) (see Figure 2). This

aligns with Hyland's (2020) framework, where lexical diversity serves as a key marker of advanced language proficiency.

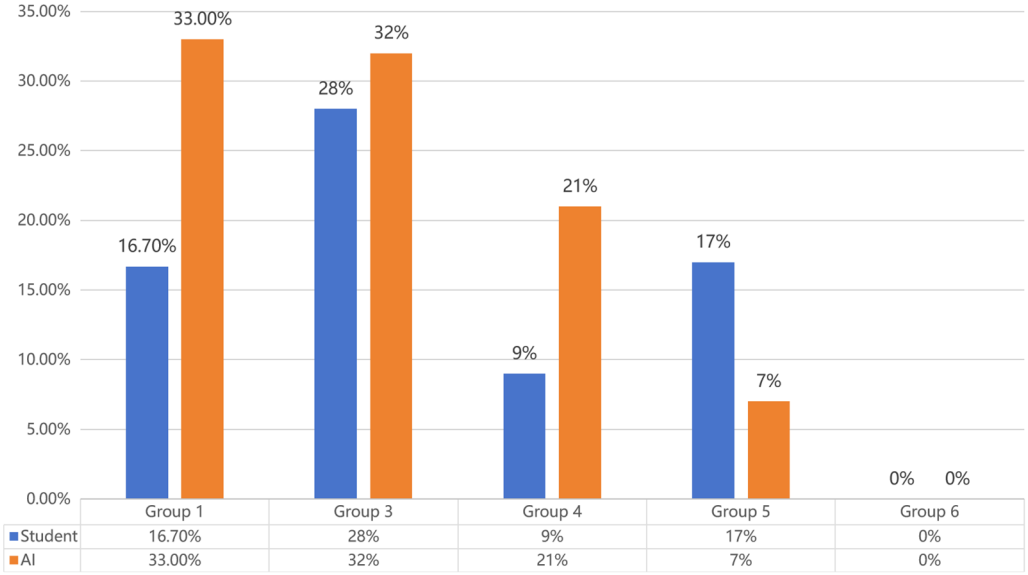

Figure 1. Group Repetition Rate for Students at Band 5.5.

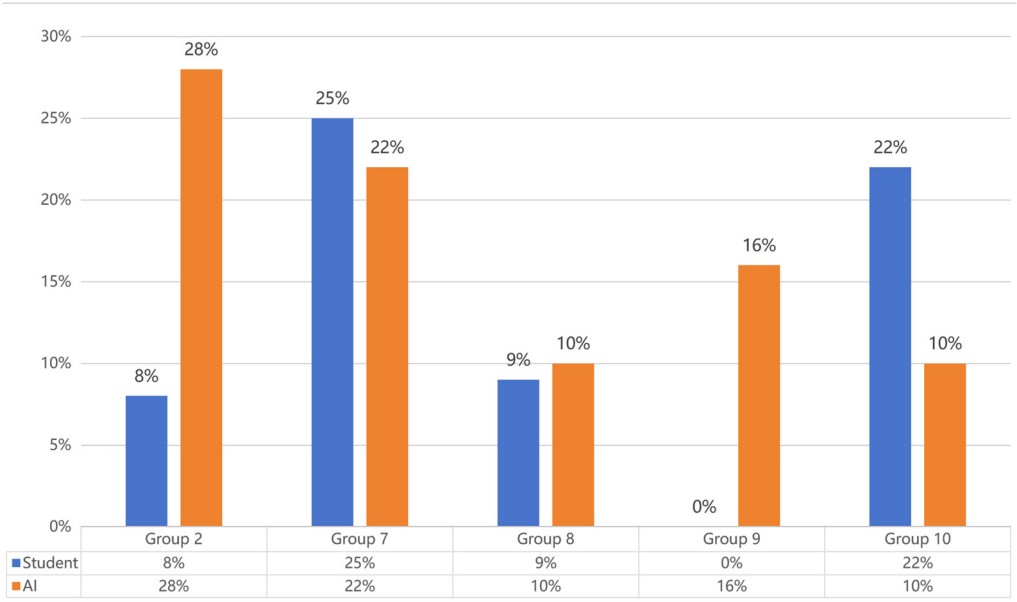

Figure 2. Group Repetition Rate for Students at Band 6.0.

AI outputs are basically consistent with this band-based progression but with critical limitations. For Band 5.5 prompts, AI-generated texts averaged 23.25% lexical repetition (see Figure 1), a rate higher than both Band 5.5 and 6.0 students, which reflected formulaic language generation. While AI's repetition rate reduced for Band 6.0 tasks (17.2%) (see Figure 2), it remained systematically higher than human-produced texts at the same level. This persistent gap exhibited AI's tendency toward standardized and formulated output, lacking the adaptive lexical variation characteristic of human writers (Hyland, 2020). Both human-written and AI-generated texts utilized common verb collocations such as "increase in" and "solve the problem", indicating a shared understanding of basic verb collocational patterns. However, human students, especially those at Band 6.0, showed greater verb diversity and complexity by using more sophisticated and context-specific collocations such as "mitigate the impact" and "facilitate the transition".

The findings highlight that compared with AI, university students showed superior lexical diversity, particularly those at Band 6.0, where human writers have a leveraged nuanced vocabulary to meet IELTS assessment criteria. AI's comparatively higher repetition rates, even when simulating higher English writing proficiency levels, reveal fundamental limitations in replicating human-like lexical creativity. As Brown and

Hilgers (2021) emphasize, natural language variation is essential for high-scoring academic writing, a domain where AI's current algorithmic constraints become apparent. These results suggest that while AI can produce grammatically correct texts, its utility for high-stakes language assessment may be bounded by its inability to authentically emulate human lexical flexibility.

4.2. Grammatical Mismatch Rate

The result of the analysis of grammatical mismatch rates revealed that there was stark contrasts between human students and AI-generated texts across proficiency levels in English writing. Students at Band 5.5 exhibited an average grammatical error rate of 26.4%, with substantial individual variation (range: 2%–37%; e.g., Student 3: 2% vs. Student 5: 37%) (see Figure 3). For instance, student 5 has trouble in identifying the difference of verb and gerund verb and there are five similar errors in this point. Such pronounced variability in foundational grammar skills, particularly in subject-verb agreement and tense consistency, aligns with McNamara and Knoch's (2020) findings that grammatical accuracy significantly predicts writing clarity and coherence in high-stakes assessments. Students at Band 6 demonstrated marked improvement (average: 19.8%), though persistent variability (e.g., Student 8: 19%) (see Figure 4) suggests that even higher-proficiency learners require continued grammatical reinforcement.

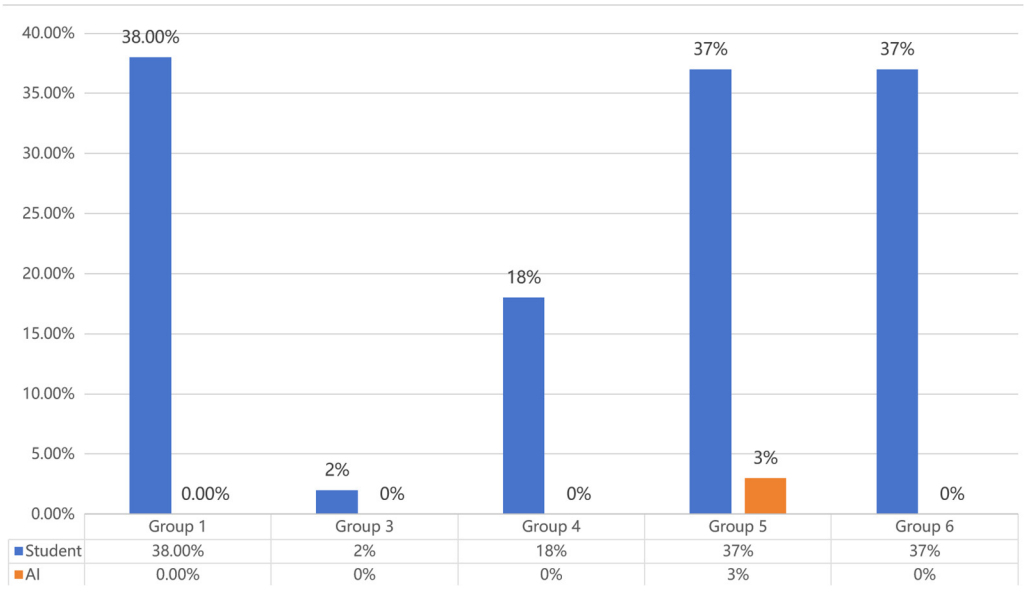

Figure 3. Grammatical Mismatch Rate for Students at Band 5.5.

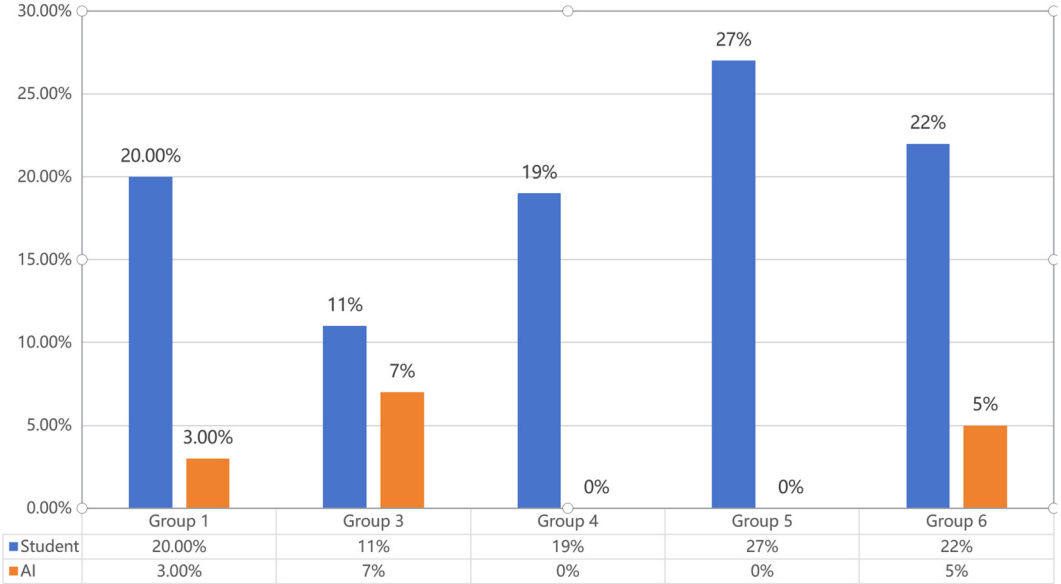

Figure 4. Grammatical Mismatch Rate for Students at Band 6.0.

In stark contrast, AI outputs maintained near-perfect grammatical precision across both bands (Band 5.5: 0.6%; Band 6.0: 3% (see Figures 3 and 4), even students at Band 5.5 had a lower mismatch rate than those at Band 6.0. This grammatically consistency reflects AI's algorithmic strength in syntactic rule application, aligning with Brown and Hilgers' (2021) assertion that AI tools excel in error detection and formal language generation.

The findings demonstrate AI's potential as a grammar-correction tool, particularly for learners who face the greatest grammatical challenges. However, the persistent human-AI gap in contextual grammatical adaptation (e.g., students' errors often involving discourse-level tense shifts rather than isolated syntax) suggests that AI's utility may be more effective when paired with targeted pedagogy. As McNamara and Knoch (2020) stated, systematic grammar instruction that focuses on high-frequency error patterns remains valuable in bridging proficiency gaps, particularly in transitional bands like 5.5 where grammatical instability largely impacts communicative effectiveness.

4.3. Contextual Appropriateness

The analysis of contextual appropriateness revealed nuanced differences between human-created and AI-generated English texts. Students at Band 5.5 achieved an average score of 4.14/10, with significant variability (range: 2.73–5.19). Notably, some students (e.g., Student 6: 5.19) (see Figure 5) exhibited strong contextual adaptation despite higher grammatical error rates. This phenomenon suggested that pragmatic competence can compensate for syntactic weaknesses in certain cases. Althought students at Band 6 did not show any better performance in the contextual appropriateness rate, the (average: 3.23/10, range: 2–5), the gap between the two groups of students was not significant (see Figure 6), which indicates that higher-proficiency English writers may not necessarily possess sound contextual knowledge.

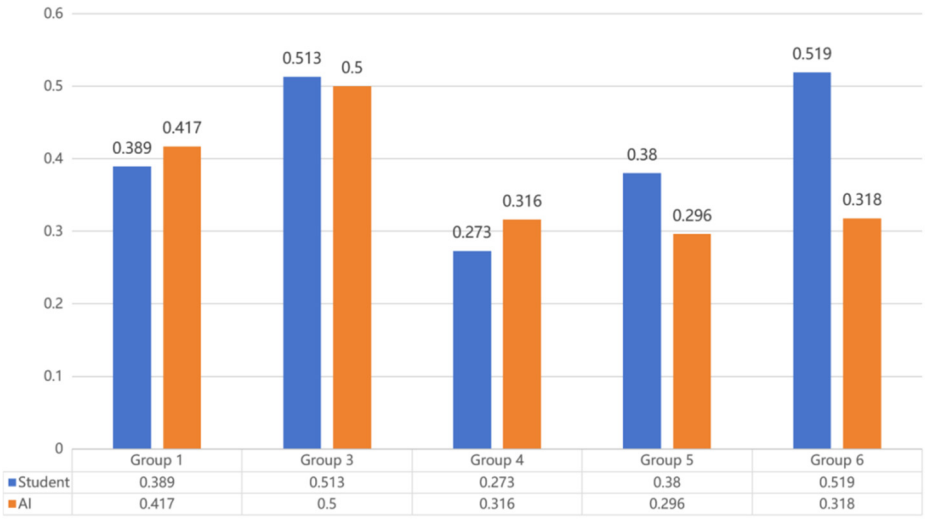

Figure 5. Contextual Appropriateness Rate for Students at Band 5.5.

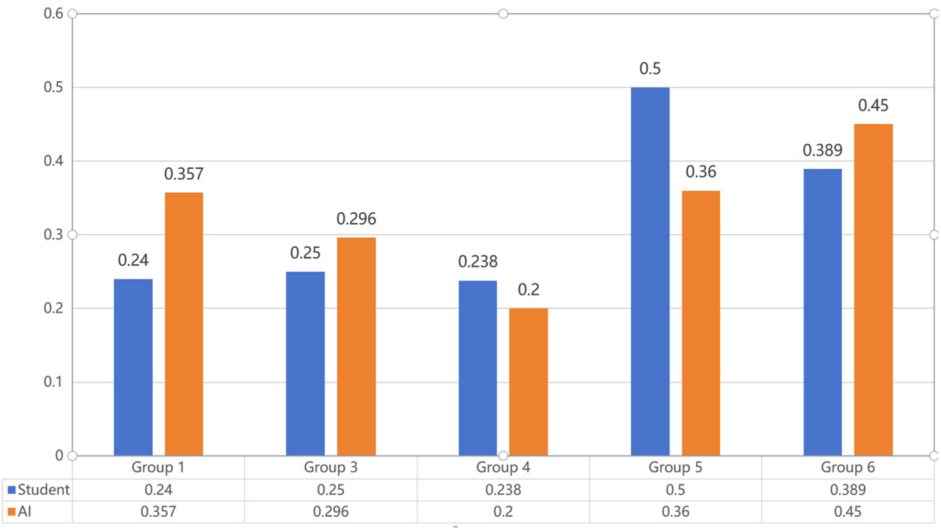

Figure 6. Contextual Appropriateness Rate for Students at Band 6.0.

AI-generated texts exhibited marginally higher and stable scores (Band 5.5: 3.69/10; Band 6.0: 3.33/10), with narrower ranges (Band 5.5: 2.96–5.0; Band 6: 2–4.5). While texts created by students at Band 5.5 slightly outperformed AI-generated texts on average and there was just a tiny gap between AI-generated texts and human-created texts for students at Band 6.0, it cannot fully demonstrate the contextual appropriateness of human-produced English writing texts excel that generated by AI, further study is required. But it, to some extent, can verify the finding by Hyland (2020) that AI struggled with culturally embedded idioms and register shifts.

From these findings, it can be inferred that university students are more competent to navigate complex rhetorical contexts, where task alignment scores peaked. AI's weaker performance reflects its current limitations in processing situational nuances, as its pattern-based generation often prioritizes grammatical correctness over discursive coherence (Read, 2021). This supports the argument made by Hyland (2020) that contextual fitness in academic writing cannot be ensured just by lexical sophistication. According to the findings, students, particularly those moving up from Band 5.5, would benefit from specific instruction in audience-aware writing techniques, and AI tools should be used in conjunction with genre-based instruction (e.g., teaching context-specific collocations) to close the adaptability gap.

5. Discussion

In the context of IELTS writing, the current study sought to compare the lexical repetition, grammatical accuracy, and contextual appropriateness of papers produced by students and AI. The results show notable differences between the two groups, which have consequences for pedagogy and language assessment.

Lexical diversity was higher among human students, especially those in Band 6.0, than in AI-generated texts, according to the analysis of lexical recurrence rates. This is consistent with earlier studies showing that lexical richness is a crucial indicator of advanced language competency (Lei & Yang, 2020). The increased rates of repetition seen in texts produced by AI point to a potential constraint in their capacity to build a diverse and contextually relevant lexicon. This result is in line with research that has shown how AI lacks natural language diversity and has a propensity to produce formulaic language (Amirjalili et al., 2024). As it demonstrates the writer's lexical inventiveness and capacity to express complex concepts, the ability to employ a broad variety of words without repetition is essential for high-scoring IELTS writing (Lei & Yang, 2020).

Additionally, the study discovered that, especially in Band 5.5, AI-generated writings had noticeably lower rates of grammatical errors than those produced by human pupils. This outcome is in line with other research showing that AI systems are excellent at producing documents that are cohesive and grammatically correct (Khalifa & Albadawy, 2024). The promise of AI-generated texts as grammar-correction tools is suggested by their high grammatical accuracy, especially for students who struggle with grammar (Tengler & Brandhofer, 2025). However, the requirement for focused grammar teaching to enhance writing competency is highlighted by the ongoing variation in grammatical accuracy among human pupils (Kim et al., 2021). This emphasises how crucial it is to combine AI technologies with conventional teaching techniques in order to take use of their advantages in error detection and meet the particular requirements of each student (Amirjalili et al., 2024).

In contrast to AI systems, students have shown flexibility in intricate rhetorical circumstances, according to the contextual appropriateness analysis. This result is in line with studies showing AI has trouble understanding cultural allusions and context-specific subtleties, which are essential for writing well on the IELTS (Lei & Yang, 2020). A crucial component of lexical resourcefulness is the capacity to match words and expressions to the particular demands of a writing assignment since it shows how well the writer understands the intended audience and goal (Khalifa & Albadawy, 2024). According to the study, although AI is capable of producing writings that are grammatically correct, it might not have the strategic adaptability needed to reach high levels of contextual appropriateness.In order to improve students' contextual flexibility, this emphasizes the necessity of genre-based education and specific training in audience-aware writing techniques (Amirjalili et al., 2024).

The results show that lexical diversity, repetition, grammatical precision, and contextual appropriateness all have a substantial impact on lexical resource scoring in IELTS writing. Lexical repetition scores were lower for students in Band 6.0. This is consistent with earlier studies that found a favourable relationship between writing performance and lexical resourcefulness (Lei & Yang, 2020). Although AI outperformed human students in real-world situations due to limitations in lexical richness and contextual appropriateness, its strong grammatical accuracy justified its scores. According to this, artificial intelligence (AI) may not be able to completely recreate the richness and diversity of human language needed for high-stakes language assessment, even though it can be a useful tool for grammatical correction (Dwivedi et al., 2023).

In conclusion, human students—especially those in Band 6.0—show an edge in lexical diversity, which is essential for writing well on the IELTS. Although AI-generated texts are quite good at grammar, they have trouble

with lexical variety and subtleties of context. This demonstrates AI's potential as a grammar-correction tool while also highlighting how beneficial it is for English writing learners to combine AI with conventional teaching techniques. Future studies should concentrate on creating increasingly complex AI models that can more accurately represent the complexities of human language and investigate successful pedagogical approaches to incorporating AI into language instruction. However, to give a thorough grasp of AI's involvement in language acquisition, longitudinal studies monitoring the long-term effects of AI on students' English lexical resourcefulness and English writing proficiency are required (Tengler & Brandhofer, 2025).

6. Conclusions

This study provides empirical evidence of the distinct lexical and grammatical patterns between students-produced and AI-generated texts in IELTS writing tasks. Our findings demonstrate that while AI systems excel in grammatical accuracy (averaging 0.6–3% error rates vs. 20–26% for humans), they consistently underperform in lexical diversity (23% repetition at Band 5.5) and contextual appropriateness (3.69/10 at Band 5.5) compared to human writers. These results align with prior research on AI' s limitations in natural language generation (Supriyono et al., 2024) while extending our understanding of specific verb collocation challenges in high-stakes assessment contexts.The pedagogical implications are twofold. First, AI shows great potentials as a grammar-correction tool, particularly for learners who struggle with foundational syntax. Second, human instructors remain indispensable for developing advanced lexical resourcefulness and contextual adaptability: skills that correlate strongly with higher IELTS bands (Amirjalili et al., 2024). This supports calls for a "hybrid pedagogy" model where AI plays a supportive role in education (Jafry & Vorstermans, 2024).

There are various limitations to this study that should be taken into account. First off, there are just 20 texts from IELTS test-takers at a single Chinese college, making the corpus scope somewhat limited. This restricts the findings' applicability to a larger population and emphasises the necessity of more extensive, multicohort datasets to take linguistic and cultural variations in writing styles into consideration (Lei & Yang, 2020). For a more thorough understanding of AI's effects on lexical resourcefulness across a range of demographics, future research should include longitudinal data from several test centres. Second, the results cannot be applied to other AI tools due to the study's exclusive focus on ChatGPT-4 Turbo. Comparative research across several large language models (LLMs) is crucial, particularly in light of the quick development of AI technology that could close existing performance disparities (Holzinger et al., 2024b). Last but not least, quantitative assessments of verb collocations were the main component of the assessment criteria used in this study. Qualitative dimensions need more in-depth research. In order to provide a more comprehensive picture of AI's potential and constraints in language assessment, further research should be planned to investigate these dimensions using techniques like think-aloud protocols or eye-tracking investigations.

Notwithstanding these drawbacks, our research adds to the continuing discussions on AI's place in language instruction. Continuous benchmarking against human writing norms is still necessary as LLMs are more complex in order to guarantee ethical AI incorporation and reliable assessment procedures (Holzinger et al., 2024b). Future studies should concentrate on creating increasingly complex AI models that can more accurately represent the complexities of human language and investigate successful pedagogical approaches to incorporating AI into language instruction. Furthermore, to give a thorough grasp of AI's involvement in language acquisition, longitudinal studies monitoring the long-term effects of the technology on students' lexical resourcefulness and writing skill are required (Kim et al., 2021). In conclusion, AI cannot completely replace the complexity and diversity of human language needed for high-stakes language evaluation, even though it has great promise as a grammar-correction tool and for personalized feedback. High-scoring IELTS writing requires significant lexical resourcefulness and contextual adaptability, both of which are developed by human instructors. To optimize learning results and meet the specific requirements of varied learners, a hybrid pedagogy approach that capitalizes on the advantages of both AI and human instruction is advised.

Author Contributions

M.D.: writing—original draft preparation; M.L.: visualization, data curation; Y.D.: supervision, validation; F.W.: data curation, formal analysis. All authors have read and agreed to the published version of the manuscript.

Funding

This paper is supported by the Macao Science and Technology Development Fund (FDCT) (No.0071/2023/RIB3), Joint Research Funding Program between the Macau Science and Technology

Development Fund (FDCT) and the Department of Science and Technology of Guangdong Province (FDCT-GDST) (No.0003-2024-AGJ).

Institutional Review Board Statement

Ethical review and approval were waived for this study, due to the nature of the research involving only the analysis of anonymized, pre-existing IELTS writing samples and AI-generated texts, which do not pose any risk to participants. The study adhered to all applicable ethical guidelines for research involving human participants, including data confidentiality and anonymity.

Data Availability Statement

The raw data supporting the findings of this study are available from the corresponding author upon reasonable request. The data will be retained securely and will be accessible to qualified researchers who agree to protect the confidentiality of the participants and comply with all relevant ethical guidelines. Requests for data access should be directed to M24092100349@cityu.edu.mo.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

Ahmed, R., & Mohammed, F. (2024). Lexical collocational errors made by efl high school students. *Govara Zanistîn Mirovayetî Ya-zankoya Zaxo*, *12*(4), 888–894. https://doi.org/10.26436/hjuoz.2024.12.4.1384.

Allami, H., Najari, B., & Tajeddin, Z. (2025c). The impact of sociocultural theory-informed instruction on learners' IELTS writing: Task response, grammar, vocabulary, coherence and cohesion. *Asian-Pacific Journal of Second and Foreign Language Education*, *10*(1), 7. https://doi.org/10.1186/s40862-024-00310-z.

Amirian, S. M. R., Pourfarhad, M., & Nafchi, A. M. (2016). Teachers' Assessment Literacy and Its Correlation with IELTS Students' Achievement in Writing Skill. *Theory and Practice in Language Studies*, *6*(5), 994. https://doi.org/10.17507/tpls.0605.12.

Amirjalili, F., Neysani, M., & Nikbakht, A. (2024). Exploring the boundaries of authorship: A comparative analysis of AI-generated text and human academic writing in English literature. *Frontiers in Education*, *9*, 1347421. https://doi.org/10.3389/feduc.2024.1347421.

Anthony, L. (2022). *AntConc* (Version 4.2.0) [computer software]. Waseda University. http://www.laurenceanthony.net/software.

Arefsadr, S., Babaii, E., & Hashemi, M. R. (2022). Why IELTS candidates score low in writing: Investigating the effects of test design and scoring criteria on test-takers' grades in IELTS and World Englishes essay writing tests. *Iranian Journal of Language Teaching Research*, *12*, 145–159. https://doi.org/10.22034/ijlt.2022.157131.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March 3–10). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? [Conference session]. 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623), Virtual. https://doi.org/10.1145/3442188.3445922.

Biber, D., Reppen, R., Staples, S., & Egbert, J. (2020). Exploring the longitudinal development of grammatical complexity in the disciplinary writing of L2-English university students. *International Journal of Learner Corpus Research*, *6*(1), 38–71. https://doi.org/10.1075/ijlcr.18007.bib.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media, Inc.

Creswell, J. W. (1994). *Research Design: Qualitative*, *quantitative*, *and mixed methods approaches*. http://www.revistacomunicacion.org/pdf/n3/resenas/research_design_qualitative_quantitative_and_mixed_methods_approaches.pdf.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642.

Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, *25*(2), 155–185. https://doi.org/10.1177/0265532207086780.

Escalante, J., Pack, A., & Barrett, A. (2023b). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2.

Estaji, M., & Hashemi, M. (2022). Phraseological competence in IELTS academic writing task 2: Phraseological units and test-takers' perceptions and use. *Language Testing in Asia*, *12*(1), 34. https://doi.org/10.1186/s40468-022-00180-7.

Hafeez, M., Ajmal, F., & Zulfiqar, Z. (2025). *Artificial intelligence in instructional pedagogy: Potential benefits and future implications* (pp. 143–174). ISTES BOOKS.

Holzinger, A., Zatloukal, K., & Müller, H. (2024b). Is Human Oversight to AI Systems still possible? *New Biotechnology*, *85*, 59–62. https://doi.org/10.1016/j.nbt.2024.12.003.

Hyland, K. (2020). Lexical density and academic writing: A corpus-based study. *Applied Linguistics*, *41*(3), 425–445. https://doi.org/10.1093/applin/amz012.

IDP IELTS. (n.d.). *L is for lexical resources*. https://ielts.idp.com/prepare/article-l-is-for-lexical-resource.

Ilieva, G., Yankova, T., Ruseva, M., & Kabaivanov, S. (2025). A framework for generative AI-driven assessment in higher education. *Information*, *16*(6), 472. https://doi.org/10.3390/info16060472.

Jafry, A., & Vorstermans, J. (2024). Evolving intersections: AI, disability, and academic integrity in higher education. *New Directions for Teaching and Learning*, 1–14. https://doi.org/10.1002/tl.20629.

Jahan, D., Arif, B., & Mustafa, U. (2024). Enhancing practical English grammar skills through AI: A study on the impact of ChatGPT-assisted feedback on student writing. *Contemporary Journal of Social Science Review*, *2*, 217–229.

Jasim, Y.A., & Awqati, A.J. (2025). Distinguishing human creativity from AI-generated literary texts. *Journal of Prospective Researches*, *25*(2), 40–47. https://doi.org/10.61704/pr.494.

Karataş, F., Abedi, F. Y., Gunyel, F. O., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, *29*(15), 19343–19366. https://doi.org/10.1007/s10639-024-12574-6.

Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, *5*, 100145. https://doi.org/10.1016/j.cmpbup.2024.100145.

Kim, J., Merrill, K., Xu, K., & Sellnow, D. D. (2020). My Teacher is a Machine: Understanding Students' perceptions of AI teaching assistants in Online Education. *International Journal of Human-Computer Interaction*, *36*(20), 1902–1911. https://doi.org/10.1080/10447318.2020.1801227.

Kim, Y. G., Yang, D., Reyes, M., & Connor, C. (2021). Writing instruction improves students' writing skills differentially depending on focal instruction and children: A meta-analysis for primary grade students. *Educational Research Review*, *34*, 100408. https://doi.org/10.1016/j.edurev.2021.100408.

Lei, S., & Yang, R. (2020). Lexical richness in research articles: Corpus-based comparative study among advanced Chinese learners of English, English native beginner students and experts. *Journal of English for Academic Purposes*, *47*, 100894. https://doi.org/10.1016/j.jeap.2020.100894.

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, *9*(2), 117–141. https://www.researchgate.net/publication/248545547.

Luo, X., Zhang, T., Wang, Y., & Liu, C. (2023). A Study on the Impact of A spherical videobased virtual reality on Teaching English Writing Guided by Experiential Learning Circle Theory. *Journal of Educational Technology and Innovation*, *5*(2), 69–76. https://doi.org/10.61414/jeti.v5i2.125.

Maeda, H. (2024). Field-Testing Multiple-Choice questions with AI examinees: English Grammar items. *Educational and Psychological Measurement*, *85*(2), 221–244. https://doi.org/10.1177/00131644241281053.

Mohammed, A. S., Mujiyanto, J., & Faridi, A. (2020). Libyan studentsâ€TM syntactic and semantic problems in using phrasal verbs to write English texts. *English Education Journal*, *10*(4), 449–455. https://doi.org/10.15294/eej.v10i4.39230.

Nguyen, L. Q., & Van Le, H. (2022). Improving L2 learners' IELTS task 2 writing: The role of model essays and noticing hypothesis. *Language Testing in Asia*, *12*(1), 58. https://doi.org/10.1186/s40468-022-00206-0.

Oved, O., & Alt, D. (2025). Teachers' technological pedagogical content knowledge (TPACK) as a precursor to their perceived adopting of educational AI tools for teaching purposes. *Education and Information Technologies*. https://doi.org/10.1007/s10639-025-13371-5.

Park, J. (2019). An AI-based English Grammar Checker vs. Human Raters in Evaluating EFL Learners' Writing. *Multimedia-Assisted Language Learning*, *22*(1), 112–131. https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART002449524.

Pearson, W. S. (2019). 'Remark or retake'? A study of candidate performance in IELTS and perceptions towards test failure. *Language Testing in Asia*, *9*(1), 17. https://doi.org/10.1186/s40468-019-0093-8.

Peng, Z., & Wan, Y. (2023). Human vs. AI: Exploring students' preferences between human and AI TA and the effect of social anxiety and problem complexity. *Education and Information Technologies*, *29*(1), 1217–1246. https://doi.org/10.1007/s10639-023-12374-4.

Perdana, I., Manullang, S. O., & Masri, F. A. (2021). Effectiveness of online Grammarly application in improving academic writing: Review of experts experience. *International Journal of Social Sciences*, *4*(1), 122–130. https://doi.org/10.31295/ijss.v4n1.1444.

Piasecki, M., Walkowiak, T., Rudnicka, E., & Bond, F. (2018). Lexical platform—The first step towards user-centred integration of lexical resources. *Cognitive Studies 18*, 1–20. https://doi.org/10.11649/cs.1811.

Rahmouni, K. (2025). Exploring the use of ChatGPT in teaching Arabic case endings: Effectiveness, challenges and recommendations. *Journal of Educational Technology and Innovation*, *6*(4), 1–20. https://doi.org/10.61414/jeti.v6i4.198.

Rienties, B., Tessarolo, F., Coughlan, E., Coughlan, T., & Domingue, J. (2025). Students' perceptions of AI Digital Assistants (AIDAs): Should institutions invest in their own AIDAs? *Applied Sciences*, *15*(8), 4279. https://doi.org/10.3390/app15084279.

Saadatara, A., Kiany, G., & Talebzadeh, H. (2022). Bundles to beat the band in high-stakes tests: Pedagogical applications of an exploratory investigation of lexical bundles across band scores of the IELTS writing component. *Journal of English for Academic Purposes*, *61*, 101208. https://doi.org/10.1016/j.jeap.2022.101208.

Samsidar, A., Fauzi, R., Karuppiah, S., & Patak, A. (2022). An investigation of vocabulary knowledge on students' writing performance. *International Journal of Humanities and Innovation (IJHI)*, *5*(4), 124–133. https://doi.org/10.33750/ijhi.v5i4.162.

Spence, P. R., Kaufmann, R., Lachlan, K. A., Lin, X., & Spates, S. A. (2024). Examining perceptions and outcomes of AI versus human course assistant discussions in the online classroom. *Communication Education*, *73*(2), 121–142. https://doi.org/10.1080/03634523.2024.2308832.

Supriyono, N., Wibawa, A. P., Suyono, N., & Kurniawan, F. (2024). Advancements in natural Language Processing: Implications, challenges, and future directions. *Telematics and Informatics Reports*, *16*, 100173. https://doi.org/10.1016/j.teler.2024.100173.

Suraprajit, P. (2021). An analysis of errors in English essays written by Thai Non-English major students. *Theory and Practice in Language Studies*, *11*(1), 12. https://doi.org/10.17507/tpls.1101.02.

Tengler, K., & Brandhofer, G. (2025). Exploring the difference and quality of AI-generated versus human-written texts. *Discover Education*, *4*(1), 113. https://doi.org/10.1007/s44217-025-00529-z.

Veerappan, V. A., & Sulaiman, T. (2012). A Review on IELTS Writing Test, Its Test Results and Inter Rater Reliability. *Theory and Practice in Language Studies*, *2*(1), 138–143. https://doi.org/10.4304/tpls.2.1.138-143.

Yang, M. (2025d). Artificial intelligence versus human teacher assistants and language learners' progress in learning and retention of complex sentences. *Current Psychology*, *44*, 6292–6304. https://doi.org/10.1007/s12144-025-07488-6.

Zhu, M., & Wang, C. (2025). A systematic review of research on AI in language education: Current status and future implications. *Language Learning & Technology*, *29*(1), 1–29. https://hdl.handle.net/10125/73606.