ESTEEM Journal of English Study Programme
E-ISSN: 2622-2213
P-ISSN: 2622-9323

PROMPTED BAND LABELS AND LINGUISTIC PATTERNING IN AI-GENERATED IELTS WRITING

Wirdatul Khasanah[1], Herlinda Nur Fauziya[2], Dwiki Dermawan[3], Mulat Sarira[4]

[1,2,3,4]PSDKU Universitas Negeri Surabaya
E-mail: [1,2,3,4]wirdatulkhasanah@unesa.ac.id

Accepted :

10 February 2026

Published :

10 February 2026

Corresponding Author:
Wirdatul Khasanah

Email Corresponding :
wirdatulkhasanah@unesa.ac.id

ABSTRACT

This study examines the extent to which a generative AI model demonstrates consistency and patterned variation when responding to IELTS Writing Task 2 prompts differentiated by band-level instructions. Specifically, it investigates how Google Gemini adapts its written output when producing essays aligned with distinct proficiency targets. Using a controlled research design, the study analyses 24 AI-generated essays produced under two band conditions Band 6.0 and Band 7.5 while holding the task prompt and topic constant, with band instruction as the sole experimental variable. Lexical diversity was measured using the Measure of Textual Lexical Diversity (MTLD) and the Moving Average Type-Token Ratio (MATTR), and discourse organisation was examined through key argumentative markers. The findings reveal that Band 7.5 prompts consistently generate essays with higher lexical diversity and more cohesive discourse organisation than Band 6.0 prompts, despite comparable text lengths. These results suggest that generative AI encodes latent representations of writing proficiency that are activated by band-level cues and realised probabilistically rather than deterministically. By adopting a within-system, corpus-based approach rather than comparing human and AI texts, this study demonstrates that proficiency distinctions in AI-generated writing emerge through systematic yet non-fixed adaptations, with important implications for AI-assisted language assessment and instruction.

Keywords: Generative Artificial Intelligence, IELTS Writing Task 2, Band-Level Prompting, Lexical Diversity, Discourse Organization.

1. INTRODUCTION

Increasing access to generative artificial intelligence (AI) has affected academic writing including the process of preparing for the International English Language Testing System (IELTS). AI tools are now the predominant tool used by test takers, faculty, and the test preparer to compose model essays, help with testing, and give feedback because it offers a fast and consistent alternative for exam preparation (Andriani, 2025; Mazari, 2025; Rosdiana et al., 2024). Consequently, how text generated by generative AI constitutes written responses in assessment-based media has become an interesting focus on applied linguistics and language assessment research (Mei et al., 2025; El Maamri, 2025; Lin, 2025). However, the published literature on AI-produced IELTS and academic writing has focused on similarities and differences between human essays and the AI (Amirjalili et al., 2024; Jiang & Hyland, 2025; Zhang & Crosthwaite, 2025). This is where research has been conducted to measure characteristics such as lexical richness, grammatical integrity, cohesion, and quality of writing which often try to measure whether AI-generated texts are closer or worse than normal human text, or whether AI

texts might undermine the validity of assessment (Al Hosni, 2025; Huang et al., 2024; Torky & Ahmed, 2025).

Although these works are informative, their concern with human–AI comparisons serves to conceptualize AI as an alternative test student, rather than a system whose linguistic behaviour is something that can be systematically compared. Recent efforts in AI and natural language generation have shown that generative models are responsive to prompt formation (Sharma et al., 2025; Joshi et al., 2024; Fattah et al., 2023). Variations in the phrasing of prompts, task framing, or stylistic instruction also affect output length, lexical selection, and syntactic structure. Yet prompt sensitivity has seldom been investigated in standard language assessment genres, where the linguistic expectations are explicitly stated through band descriptors.

Band descriptors are used in IELTS Writing to establish expected levels of lexical resource, grammatical scale and accuracy but we know very little about generative AI's linguistic behaviour in response to IELTS band labels used as control signals in the prompt. To bridge this gap, the current work adopts a within-system analytical view which considers generative AI as a text-making system rather than human writers. Instead of contrasting AI-generated texts with human-authored text, the study investigates the linguistic differences in AI IELTS Writing Task 2 responses generated under band-label prompting. By controlling for topic, task type and essay length, the study further isolates the effect of IELTS band labels on the model's internal linguistic patterning. Specifically, this study investigates the lexical diversity, sentence-level features, and discourse marker usage in AI-generated essays produced for two target bands, which are Band 6.0 and Band 7.5.

In this analysis, the aim was to learn more about how generative AI reallocates lexical, syntactic, and discourse-level resources in response to band labels alone, without explicit linguistic instruction or not. In this way, the research makes a contribution to the ongoing literature on AI-generated writing: by changing the focus from the evaluative comparison to system-level linguistic behavior in evaluation writing contexts, it moves us toward research-backed assessments of AI use as a potential route and not an end.

2. LITERATURE REVIEW

The increasing use of generative AI in educational contexts has engendered increased scholarly consideration of AI-generated writing, particularly in the context of high-stakes language testing. In IELTS-related research, AI written texts have been frequently assessed as compared with human writing essays and analyzed for linguistic features like lexical richness, grammatical correctness, cohesion, and organization of the discourse (Xiaofan & Annamalai, 2025; Markauskaite et al., 2022; Soughati et al., 2025). These researchers typically inquire whether AI-generated writings will meet human level performance standards or introduce problems on assessment accuracy and academic honesty. Though this line of inquiry has yielded some fruitful results, its comparative outlook places AI as something that should be compared to other test takers as opposed to a linguistic system with internally varying outputs. A lack of attention to the heterogeneity between AI texts in different forms produced under different pedagogical conditions has led the academic community to devote relatively little attention to the effect of AI in particular (Al Hosni, 2025; Liu, 2025; Amin, 2023). This disconnect constrains our comprehension of how generative AI modifies language use within explicit restrictions.

Research in natural language generation has shown that prompt design is especially responsive to generative AI systems (Cappello et al., 2025; Jaya et al., 2025; Putri et al., 2023). Differences in the

framing, specification, and style of prompts have been found to affect output length, lexical selection and syntactic complexity, and discourse structure. Prompt sensitivity has thus emerged as an important construct when approaching how generative models behave to the user input. However most of the studies on prompt sensitivity have taken place in a general or creative writing context than in a standardized assessment genre (Révauger & Rident, 2025). Prompts in language assessment environments are more than simple instruction; implicit or explicit performative expectations and expectations are encoded in prompts as well. Notwithstanding these challenges, there are few studies that investigate how generative AI reacts linguistically when prompts specify intended proficiency levels or performance bands. This is an unexplored intersection between AI prompt engineering and applied linguistics.

Linguistic features and features of IELTS writing band descriptors

The band descriptions of the IELTS writing are expected performance levels in terms of task response, coherence and cohesion, lexical resource, and grammatical range and accuracy. Language studies on IELTS writing have tended to operationalize these descriptors by looking at such features as lexical diversity, sentence complexity, and cohesive devices used in learner writing. Band descriptors have been used in other studies mostly as evaluative methods in human-authored texts (Saed et al., 2021; Sherine et al., 2020). As instructional standards for text production, especially in AI-generated writing, their role in this regard is relatively neglected. Since generative AI systems can be explicitly asked to address certain band levels, understanding the variation of linguistic features between these prompts provides a different approach to operationalizing a band descriptor without invoking human performance standards.

Overall, previous studies point to three key shortcomings. First, as AI-generated writing on IELTS remains mainly based on human-AI comparison, within-system variation has not been significantly addressed. Second, prompt sensitivity has seldom been studied in assessment-based writing genres. Third, band descriptors of the IELTS have rarely been viewed as features in instruction influencing language production (Devitska & Horvat-Choblya, 2024; Wu, 2025). Filling these gaps through a systems-oriented analysis addressing how generative AI varies linguistic resources to specifically stipulated band-level prompts is needed.

This study adopts a within-system analytical perspective to examine how IELTS band labels function as control signals that shape the linguistic patterning of AI-generated writing. This study investigates whether band labels alone are sufficient to trigger systematic differences in lexical diversity, syntactic structure, and discourse organisation in IELTS Writing Task 2 responses. The following research questions are addressed in order to accomplish these objectives: How does lexical diversity differ between AI-generated IELTS Writing Task 2 essays produced under Band 6.0 and Band 7.5 band-label prompts? How does sentence-level syntactic patterning vary between AI-generated essays conditioned by different IELTS band labels? How does the use of argumentative discourse markers differ between AI-generated responses produced under Band 6.0 and Band 7.5 band-label prompts?

3. METHODS

This study adopts a within-system, corpus-based research design to investigate how a generative AI model, in this case is Gemini 1.5 Flash model, varies its linguistic output when prompted with different IELTS band targets. Rather than comparing AI-generated texts with human writing, the study

conceptualises the AI as a text-producing system and examines how its internal linguistic patterning shifts under controlled prompt conditions.

Data and Corpus Construction

The dataset consists of twenty-four AI-generated IELTS Writing Task 2 essays. These texts were evenly divided into two conditions: twelve essays produced under a Band 6.0 prompt and twelve under a Band 7.5 prompt. All essays responded to the same IELTS task and topic and were generated within the same time period in order to minimise system-level variation. A strict length constraint of approximately 250–300 words was applied to all outputs so that differences in linguistic patterning could not be attributed to differences in text length.

To facilitate systematic analysis, all texts were compiled into a structured corpus using LancsBox X (version 5.5), with separate sub-corpora created for the Band 6 and Band 7 conditions.

Prompt Design and Control

The prompts were designed to be minimal to avoid linguistic priming. Each prompt included only the IELTS task, the target band label, and the word-length requirement. No specific instructions regarding vocabulary, grammar, or discourse structure were provided. This design ensured that any linguistic differences observed between Band 6 and Band 7 outputs could be attributed to the band label itself rather than overt stylistic guidance.

Analytical Framework

The analysis focused on three complementary dimensions of linguistic patterning: lexical diversity, text length control, and discourse organisation. For lexical diversity, we calculated MTLD and MATTR, two robust corpus-linguistic metrics that are relatively insensitive to text length variation and that are widely used in writing research. These measures were computed for each individual essay and then compared across the two band conditions.

The number of tokens and sentences in each essay was recorded to control for length effects. Thus, the study was able to check that any difference in lexical diversity or in discourse structure observed for the discourses under consideration did not result merely from longer or shorter texts.

Analytical Procedure

All lexical and discourse measures were extracted using LancsBox X. MTLD and MATTR values were calculated for each essay, and summary statistics were compared between Band 6 and Band 7. Discourse marker frequencies were extracted for each sub-corpus, and concordance lines were examined qualitatively to identify differences in rhetorical function rather than frequency alone. This combined quantitative–qualitative approach makes it possible to identify systematic linguistic shifts while also capturing how those shifts are realised in actual argumentative discourse.

Ethical Considerations

Due to the fact that the study is limited to analyzing AI-generated text and human subjects are removed from the study, informed consent and institutional ethics approval was not sought. The research involves system-generated outputs and does not analyze or refer to individual test takers.

4. RESULTS AND DISCUSSION

Lexical diversity was measured using MATTR (window= 50) and MTLD which indicates vocabularies variation in texts with similar length. The corpus consisted of 24 AI-generated IELTS task 2 essays which are divided between band 6 and band 7.5.

Vol 9, No 1 (2026): ESTEEM

Table 1 summaries the MTLD ranges for both groups

Band condition	MTLD range	Approximate mean
Band 6	117.33 – 158.02	≈ 135
Band 7.5	148.10 – 220.96	≈ 180

The results show a clear separation between the two band conditions. It shows that Band 6 essays have lower MMTLD values compared to band 7.5 which displayed substantially higher lexical diversity in which several text exceeded MTLD= 190 and one reaching 220.96.

This indicates that when instructed to generate band 7 responses, genAI produces text with significantly grater vocabularies variation even though the average text length (≈ 250–310 tokens) remains comparable across conditions. The difference therefore reflects lexical choice and distribution, not simply longer outputs.

Distribution of discourse markers

To examine the rhetorical organization, four high frequency discourse we chosen to be analysed using KWIC concordances in LancsBox which are however, therefore, for example, and on the other hand.

Table 2 presents their frequency across the two band conditions.

Marker	Band 6	Band 7
however	9 (8/12 texts)	10 (10/12 texts)
therefore	6 (6/12 texts)	3 (3/12 texts)
for example	3 (3/12 texts)	1 (1/12 texts)
on the other hand	6 (6/12 texts)	0 (0/12 texts)

Although however occurs at similar rates in both conditions, its discursive function differs sharply as found in the following section.

Qualitative patterns in discourse marker use

"However" as template vs argumentative concession

In Band 6 texts, however introduces formulaic counter-statements, such as:

"However, it is true that However, it is true that… However, experience shows that…"

These constructions have a mainly superficial contrastive use and often signal a predictable argumentative shift.

In Band 7 texts, however occurs within more complex concession structures, such as:

"However, it is important to acknowledge that public transport alone cannot solve…

However, this should not come at the expense of economic growth…"

In this particular use, however, "however" does not merely signal a contrast, but actively works as a means of balancing conflicting policy, economic, or environmental interests, which implies a more sophisticated level of rhetorical control.

Decline of "on the other hand" in Band 7

The expression on the other hand appears in half of the Band 6 texts but does not occur at all in Band 7 texts. This is significant because on the other hand is a marker of binary contrast framing ("one side vs another side"), which is typical of formulaic IELTS writing. Its disappearance in Band 7 suggests that the AI shifts from dualistic argument structure to multi-dimensional concession and synthesis, primarily realised through however and embedded clause structures.

"Therefore" and the level of abstraction

Band 6 texts use therefore twice as often as Band 7 texts. However, KWIC analysis shows that in Band 6, therefore

typically introduces local conclusions, for example:

"Therefore, expanding roads…"

In Band 7, therefore is used more sparingly but tends to introduce policy-level or abstract conclusions, such as:

"Therefore, road building should be a…"

This suggests that Band 7 texts prefer fewer but more semantically weighty conclusions, rather than frequent mechanical logical chaining.

Explicit vs implicit exemplification

The phrase for example appears three times in Band 6 but only once in Band 7. Band 6 relies on explicit signaling of examples, whereas Band 7 often embeds examples within the arguments so it runs smoothly. This reflects a shift toward more academic-style exemplification which is less dependent on explicit discourse labels.

Taken together, the results show that Band 7 prompting induces a systematic reconfiguration of the AI's linguistic behaviour which include significantly higher lexical diversity, reduced reliance on formulaic contrast markers, more nuanced concession structures, more abstract and integrated argumentation.

These differences cannot be explained by length or topic, as both conditions were generated from identical prompts except for the band specification.

Discussion

This study set out to examine whether prompting a generative AI system with different IELTS band targets leads to measurable and systematic differences in linguistic output. The results demonstrate that this is indeed the case: changing the band prompt from Band 6 to Band 7 produces not only longer or more fluent texts, but structurally different patterns of lexical and discourse organisation.

Prompt-specified proficiency and lexical control

The sharp contrast in MTLD values between Band 6 and Band 7 outputs shows that the model treats "band level" not merely as a stylistic label but as a cue to reconfigure its lexical selection mechanisms. Band 7 texts display substantially higher lexical diversity despite similar token counts. This suggests that the model actively increases lexical dispersion and avoidance of repetition when higher proficiency is requested.

From a language assessment perspective, this mirrors how human raters evaluate vocabulary: higher IELTS bands require not just more words, but wider and more flexible lexical use. The AI appears to have internalised this norm and operationalises it through probabilistic vocabulary variation rather than semantic depth alone.

However, the wide MTLD range within Band 7 (up to 220.96) also indicates instability. It shows that the model can generate highly diverse texts but does not do so consistently across outputs. This aligns with the growing literature on generative AI as stochastic rather than deterministic, even under identical prompts (sheng Cheng et al., 2025).

From formulaic to rhetorical discourse

One of the most noticeable finding is the use of discourse markers. From the result, it is known that band 6 texts rely heavily on template-like rhetorical framing especially through transitional phrases such as on the other hand, therefore, and explicit exemplifiers (for example). These words are widely used in any IELTS Task 2 essays where learners are taught to signal structure overtly (Pham, 2022).

In contrast, Band 7 texts largely abandon on the other hand and reduce explicit exemplification, instead favouring embedded concession structures built around however. This shift is linguistically meaningful: on the other hand constructs a binary argumentative frame, while however allows for gradual negotiation, qualification, and synthesis (Mubarok et al., 2024). This suggests that the AI models higher-band writing not as simply "more connectors" but as more complex interclausal relationships. In other words, higher-band prompts push the model from list-like argumentation toward hierarchically organised reasoning.

Logical relations and abstraction

The reduced use of therefore in Band 7 texts further supports this interpretation. Band 6 texts frequently use therefore to mark local causal links, often in a repetitive and mechanical way. Band 7 texts, by contrast, reserve therefore for global or policy-level conclusions. This reflects a shift from micro-level coherence (sentence-to-sentence linkage) to macro-level coherence (argument-to-conclusion integration). Such a distinction is central to advanced academic writing and aligns closely with IELTS descriptors for coherence and cohesion at Bands 7 and above (Wu, 2025)

What this reveals about AI "writing ability"

Taken together, these patterns indicate that the AI system does not simply imitate surface features of IELTS bands. Instead, it appears to have learned probabilistic mappings between band labels and abstract discourse strategies, such as: how arguments are structured, how concessions are framed, how evidence is integrated, how conclusions are drawn. However, the instability observed in MTLD and marker distribution suggests that this knowledge is not rule-based but statistical. The AI approximates higher-level writing through probability shifts, not through stable rhetorical planning. This has important implications for both AI evaluation and language education: what looks like "Band 7 writing" in AI is not a fixed competence but a stochastic simulation of proficiency.

Implications for IELTS, teaching, and AI use

For IELTS teaching, this means AI-generated Band 7 models should not be treated as reliable exemplars of human writing. While they exhibit advanced surface features, their rhetorical structure may fluctuate across outputs in ways that human learners do not. For AI research, the findings support the need to evaluate generative systems using repeated-output analysis, rather than single examples. Only by examining distributions across multiple generations can we see how prompt conditions truly shape linguistic behaviour.

5. CONCLUSION

This study demonstrates that a generative AI system systematically modulates its linguistic behaviour in response to IELTS band-level prompts, with evidence indicating that altering the band label alone is sufficient to produce meaningful and patterned variation in AI-generated writing. Using a within-system, corpus-based design, the findings show that essays prompted at Band 7 consistently exhibit greater lexical diversity, as measured by MTLD and MATTR, alongside more advanced rhetorical organisation characterised by reduced dependence on formulaic discourse markers and increased use of implicit cohesion and elaboration, whereas Band 6 outputs rely more heavily on explicit logical connectors and binary contrast structures. Collectively, these patterns suggest that the AI does not merely react to superficial prompt cues, but instead activates latent representations of writing proficiency that are probabilistically realised through band-level instructions. By shifting analytical focus from human–AI comparison to intra-model variation, this study offers a novel contribution to research

on generative AI in language assessment, highlighting band-specified prompts as powerful control signals that shape the deployment of lexical, syntactic, and discourse resources, with important implications for AI-supported test preparation, feedback, and writing instruction. At the same time, the study is limited by its reliance on a single AI model, a single IELTS task prompt, and a restricted set of linguistic measures; future research should therefore pursue cross-model comparisons, incorporate a broader range of prompts, genres, and proficiency levels, and integrate additional analytic dimensions such as syntactic complexity, semantic coherence, pragmatic appropriateness, and rater-based evaluation to develop a more comprehensive understanding of AI-mediated writing proficiency.

6. ACKNOWLEDGEMENT

Author should present list of acknowledgement at the end. Any financial or nonfinancial support for the study should be acknowledged.

7. REFERENCES

Al Hosni, J. (2025). Preserving Authorial voice in academic texts in the age of generative ai: A thematic literature review. *Arab World English Journal*, *16*(3), 244–258. https://doi.org/10.24093/awej/vol16no3.14

Amin, M. Y. M. (2023). AI and Chat GPT in Language Teaching: Enhancing EFL Classroom Support and Transforming Assessment Techniques. *International Journal of Higher Education Pedagogies*, *4*(4), 1–15. https://doi.org/10.33422/ijhep.v4i4.554

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*(March), 1–11. https://doi.org/10.3389/feduc.2024.1347421

Andriani, E. L. (2025). *ELTR Journal, e-ISSN 2579-8235 ,. 9*(2), 86–112.

Cappello, N., Rodriguez, D., Hernandez, E., & Canabate, D. (2025). Writing in movement : movement-based performing arts for developing physical education teachers ' identity Escribir moviéndose : artes escénicas en movimiento para desarrollar la identidad de docentes de educación física Authors How to cite in APA Keyword. *Retos: Nuevas Tendencias En Educación Física, Deporte y Recreación*, *63*, 829–845. https://doi.org/10.47197/retos.v63.111684

Devitska, A., & Horvat-Choblya, A. (2024). Linguistic domains: comparison of texts written by human and artificial intelligence. *Věda a Perspektivy*, *11(42)*. https://doi.org/10.52058/2695-1592-2024-11(42)-358-365

El Maamri, O. (2025). Towards the modelling of a writing process integrating generative ai to develop the argumentative skills of secondary school writers. *XLinguae*, *18*(1), 133–143. https://doi.org/10.18355/XL.2025.18.01.10

Erviti, A. I. (2017). *Discourse constructions in English: meaning, form and hierarchies: A study from the point of view of the Lexical Constructional Model*. https://dialnet.unirioja.es/servlet/tesis?codigo=121205

Fattah, H. A., Vadivel, B., Shaban, A. A., & Shanmugam, K. (2023). Enhancing English language education: the impact of ai integration in the classroom. *Journal of Humanities and Education Development*, *5*(6), 116–123. https://doi.org/10.22161/jhed.5.6.15

Fleckenstein, J., Meyer, J., Jansen, T., Keller, S. D., Köller, O., & Möller, J. (2024). Do teachers spot AI? Evaluating the detectability of AI-generated texts

among student essays. *Computers and Education: Artificial Intelligence*, *6*(January). https://doi.org/10.1016/j.caeai.2024.100209

Huang, H.-Y. C., Barrett, N. E., Tseng, Lo, M.-F., & Chiung-Jung. (2024). The Effectiveness of Shadowing Practice with Web-Based Apps: Towards Promoting the Comprehensibility and Lexical-Level Intelligibility of EFL Students' Presentations. *English Teaching and Learning*, *48*, 505–524. https://link.springer.com/article/10.1007/s42321-023-00145-w

Jaya, A., Hartono, R., Wahyuni, S., & Yulianto, H. J. (2025). Los efectos de la estrategia de aprendizaje basado en proyectos con actividad física en función del género sobre el rendimiento escolar y la confianza en sí mismos de los estudiantes. *Retos: Nuevas Tendencias En Educación Física, Deporte y Recreación*, *66*, 349–360. https://doi.org/10.47197/retos.v66.110067

Jiang, F. (Kevin), & Hyland, K. (2025). Does ChatGPT Write Like a Student? Engagement Markers in Argumentative Essays. *Written Communication*, *42*(3), 463–492. https://doi.org/10.1177/07410883251328311

Joshi, A., Dabre, R., Kanojia, D., Li, Z., Zhan, H., Haffari, G., & Dippold, D. (2024). Natural Language Processing for Dialects of a Language: A Survey. *Computer Science*, *1*(1). http://arxiv.org/abs/2401.05632

Lin, Z. (2025). Techniques for supercharging academic writing with generative AI. *Nature Biomedical Engineering*, *1*(9), 426–431. https://doi.org/10.1038/s41551-024-01185-8

Liu, L. (2025). Impact of AI gamification on EFL learning outcomes and nonlinear dynamic motivation: Comparing adaptive learning paths, conversational agents, and storytelling. *Education and Information Technologies*, *30*, 11299–11338. https://doi.org/10.1007/s10639-024-13296-5

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Buckingham Shum, S., Gašević, D., & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? *Computers and Education: Artificial Intelligence*, *3*(February). https://doi.org/10.1016/j.caeai.2022.100056

Mazari, N. (2025). Building Metacognitive Skills Using AI Tools to Help Higher Education Students Reflect on Their Learning Process Resumen Introduction. *RHS-Revista Humanismo y Sociedad*, *13*(1), 1–20. https://doi.org/https://doi.org/10.22209/rhs.v13n1a04

Mei, P., Brewis, D. N., Nwaiwu, F., Sumanathilaka, D., Alva-Manchego, F., & Demaree-Cotton, J. (2025). If ChatGPT can do it, where is my creativity? generative AI boosts performance but diminishes experience in creative writing. *Computers in Human Behavior: Artificial Humans*, *4*(December 2024), 100140. https://doi.org/10.1016/j.chbah.2025.100140

Mubarok, Y., Sudana, D., Yanti, D., Sugiyo, Aisyah, A. D., & Af'idah, A. N. (2024). Abusive Comments (Hate Speech) on Indonesian Social Media: A Forensic Linguistics Approach. *Academic Journal*, *14*(5), 1440. https://doi.org/10.17507/tpls.1405.16

Pham, T. H. Van. (2022). Efl learners' deployment of discourse markers in ielts essays: from beliefs to practices. *Tạp Chí Nghiên Cứu Nước Ngoài*. https://doi.org/10.25073/2525-2445/vnufs.4873

Putri, A. D., Jaya, A., & Marleni, M. (2023). Exploring the Students' Speaking

Ability Based on Their Different Personalities. *Esteem Journal of English Education Study Programme*, *6*(1), 10–16. https://doi.org/10.31851/esteem.v6i1.10203

Révauger, G., & Rident, F. (2025). "I, too", shall have to prompt? A study of EFL students and their unmonitored use of GenAI in the completion of an imitation task in poetry. *Alsic*, *1*(28). https://doi.org/10.4000/137dy

Rodriguez-Donaire, S. (2024). *Influence of Prompts Structure on the Perception and Enhancement of Learning through LLMs in Online Educational Contexts*. https://doi.org/10.5772/intechopen.1006481

Rosdiana, S., Noercolies, M. A., & Fauzan, M. H. (2024). The Use of Artificial Intelligence in Teaching Writing Skills. *EDUCASIA: Jurnal Pendidikan, Pengajaran, Dan Pembelajaran*, *9*(1), 45–56. https://doi.org/10.21462/educasia.v9i1.251

Saed, H. A., Haider, A. S., Al-Salman, S., & Hussein, R. F. (2021). The use of YouTube in developing the speaking skills of Jordanian EFL university students. *Heliyon*, *7*(7), e07543. https://doi.org/10.1016/j.heliyon.2021.e07543

Sharma, A., Kumar, P., & Dhamija, S. (2025). *From Prompts to Performance*. 223–252. https://doi.org/10.1002/9781394287499.ch10

Sherine, A., Seshagiri, A. V. S., & Sastry, M. M. (2020). Impact of whatsapp interaction on improving L2 speaking skills. *International Journal of Emerging Technologies in Learning*, *15*(3), 250–259. https://doi.org/10.3991/ijet.v15i03.11534

Sheng Cheng, X., Wu, P., & Liang, L. (2025). Detecting the knowledge structure of generative artificial intelligence research. *Journal of Information Science*. https://doi.org/10.1177/01655515251353188

Soughati, N., El Hadi, K., & Flyura, A. (2025). Technopedagogical Making Workshop and Writing/ Programming. *XLinguae*, *18*(1), 153–169. https://doi.org/10.18355/XL.2025.18.01.12

Torky, S. A. E. F. I., & Ahmed, N. H. A. E. G. S. (2025). Promoting EFL students' reading comprehension, grammatical competence, collocational competence and critical thinking disposition via data-driven pedagogical translation. *Egyptian Journal of Educational Sciences*, *5*(1), 191–266. https://doi.org/10.21608/ejes.2025.446747

Wu, J. (2025). Comparing Linguistic Features between Human-written High-Scoring IELTS Essays and AI-Generated ones. *Journal of Humanities and Social Sciences Studies*, *7*(8), 68–77. https://doi.org/10.32996/jhsss.2025.7.8.8

Xiaofan, W., & Annamalai, N. (2025). Investigating the use of AI tools in English language learning: A phenomenological approach. *Contemporary Educational Technology*, *17*(2). https://doi.org/10.30935/cedtech/16188

Zhang, M., & Crosthwaite, P. (2025). More human than human? Differences in lexis and collocation within academic essays produced by ChatGPT-3.5 and human L2 writers. *IRAL - International Review of Applied Linguistics in Language Teaching*, 1–28. https://doi.org/10.1515/iral-2024-0196