

Lexical and Syntactic Complexity across Adjacent IELTS Writing Task 2 Band Levels: A Corpus-Based Study

Truong Van Uyen^{1*}

¹Independent researcher, Vietnam

*Corresponding author's email: uyntrg3321@gmail.com

 <https://orcid.org/0009-0004-8693-939X>

 <https://doi.org/10.54855/ijte.26632>

® Copyright (c) 2026 Truong Van Uyen

Received: 16/01/2026

Revision: 18/07/2026

Accepted: 01/08/2026

Online: 12/08/2026

ABSTRACT

Keywords: IELTS Writing Task 2; lexical diversity; lexical sophistication; syntactic complexity; learner corpus

Lexical and syntactic complexity are important indicators of second language writing proficiency, particularly in assessment-oriented contexts such as IELTS Writing Task 2. However, limited attention has examined whether selected complexity measures can distinguish closely related IELTS band levels. A small-scale learner corpus of 30 essays, with 10 essays per band level, was analyzed quantitatively. Lexical diversity was measured using the Measure of Textual Lexical Diversity (MTLD), lexical sophistication was operationalized as the percentage of Academic Word List (AWL) items, and syntactic complexity was assessed through mean sentence length (MSL). One-way ANOVA showed a significant difference in lexical diversity, whereas AWL percentage and MSL showed non-significant upward trends. Post-hoc comparisons indicated that the significant MTLD difference occurred between Bands 6 and 8 rather than between immediately adjacent bands. Because the corpus consists of band-labeled instructional essays, the findings should be interpreted as exploratory rather than generalizable evidence of IELTS band distinctions.

Introduction

In recent years, learning and preparing for the International English Language Testing System (IELTS), a widely recognized test of English language proficiency, has become increasingly popular (Tran, 2015; Nguyen et al., 2020). The test assesses four language skills: listening, speaking, reading, and writing. Writing is one of the most vital skills in second language acquisition (Khudaverdiyeva, 2025). In the IELTS writing component, candidates are evaluated based on lexical resources, grammatical range and accuracy, task achievement, and coherence and cohesion, as linguistic characteristics play a major role in assigning band scores (IELTS, 2023). Therefore, it is essential to identify the linguistic features that meaningfully distinguish levels of writing achievement in both research settings and assessment contexts (Nguyen & Nguyen, 2022). Related evidence from the Vietnamese EFL context indicates that postgraduate learners also experience lexical and grammatical problems in academic writing, although these

difficulties were examined in a broader academic-writing context rather than specifically in IELTS assessment (Ho, 2024).

In second language (L2) writing research, lexical and syntactic complexity are widely regarded as key elements of writing development and proficiency (Ortega, 2003; Bulté & Housen, 2012). Based on Communicative Language Ability theory (Canale & Swain, 1980), and the Complexity–Accuracy–Fluency (CAF) framework (Housen et al., 2012), several studies consistently indicate that more proficient writers use a wide range of vocabulary and more elaborated grammatical structures (Bulté & Housen, 2012; Kyle & Crossley, 2015; O'Leary & Steinkrauss, 2022). Corpus-based research further confirms the importance of lexical diversity (LD), lexical sophistication (LS), and syntactic complexity (SC) in accounting for variation in L2 writing quality, particularly in academic and assessment contexts (Dinh, 2025).

Literature Review

Communicative Language Ability (CLA)

The theoretical basis of this research is Communicative Language Ability (CLA) (Canale & Swain, 1980), which was further established into a complete model in 1983 by Canale. According to Canale (1983), communicative competence is a complex concept, and it can be quantified by four principles, namely grammatical competence, sociolinguistic competence, discourse competence, and strategic competence. This theoretical premise fully explains language learners' motivation to use language effectively and appropriately in both oral and written forms (Bachman, 1990).

The use of CLA in L2 writing research has also gained much importance. Writing is no longer considered a mechanical process but communicative language use (Canale, 1983; Bachman, 1990; Hyland, 2003). Consequently, current studies define L2 writing ability as learners' potential for accessing linguistic means for communicative purposes in written discourse (Hyland, 2003; Ortega, 2012). Recent advances in theory have extended this approach further, incorporating dynamic system models that consider grammatical competence as part of an overall, constantly changing system that includes all the features that contribute to communicative competence (Larsen-Freeman, 2019). The abandonment of fixed, module-style trait models has emphasized the symbiotic acquisition of lexical, syntactic, and discourse levels, rather than treating grammatical competence as isolated from the others (Zhang et al., 2022).

While all four components of CLA possess theoretical relevance, empirical investigations do not invariably operationalize the entire model simultaneously. The selection of specific components is typically determined by research objectives and the nature of available data. In corpus-based investigations of written language, grammatical competence has received the most sustained empirical attention owing to its direct observability in textual output and the availability of established linguistic operationalizations (Ortega, 2003; Wolfe-Quintero et al., 1998). In contrast, sociolinguistic competence exhibits high context-dependency, discourse competence necessitates discourse-level analysis of coherence and cohesion mechanisms, and strategic competence manifests predominantly through interactional and real-time communication strategies. These latter components consequently present methodological challenges for corpus-based analysis of written texts and extend beyond the scope of the present investigation.

Accordingly, grammatical competence is a core component of CLA theory. Canale and Swain (1980) defined grammatical competence as "knowledge of lexical items, rules of morphology,

syntax, sentence grammar semantics, and phonology". These elements constitute the linguistic foundation of writing. Empirical evidence demonstrates that development in L2 writing has been captured consistently by lexical and syntactic complexity (Wolfe-Quintero et al., 1998; Ortega, 2003). Several recent corpus-based studies continue to underscore the significance of lexical and syntactic features in differentiating proficiency levels in L2 writing (Biber et al., 2016; Kyle & Crossley, 2018; Khushik & Huhta, 2020).

In the present study, CLA is therefore used in a restricted and focused way. The study does not attempt to measure the full construct of communicative language ability; rather, it draws specifically on grammatical competence to justify the analysis of lexical and syntactic features in written texts. This restricted focus is directly aligned with the purpose of the study, which is to compare IELTS Writing Task 2 essays across band-labelled levels. Accordingly, lexical diversity, lexical sophistication, and syntactic complexity are treated as observable textual indicators of selected aspects of grammatical competence that may contribute to band-level distinctions.

The Complexity–Accuracy–Fluency (CAF) framework

Grammatical competence is investigated using the linguistic complexity measures in the Complexity-Accuracy-Fluency (CAF) model. This offers an analytical systematic framework of assessing grammatical proficiency by presenting the expansion of lexical and syntactic resources in the writing production of the learner. Within this framework, Ellis (2003) explained that complexity was associated with the extent to which learners employed a broad scope of vocabulary and detailed grammatical patterns.

CAF framework has become extremely popular in the field of L2 writing research because it is a multidimensional, task-sensitive method of performance analysis (Housen et al., 2012). Recent empirical studies have shown that the framework is robust in a variety of settings and is sensitive to instructional manipulations. For example, Phuoc and Barrot (2022) used CAF measures (18) on L2 writing and recorded similarities as well as L1-based differences in CAF profiles, which are highly sensitive to learners' backgrounds and the genre of text. Imaz Agirre et al. (2025) reported that an academic-language intervention produced measurable improvements in SC and accuracy indices among secondary-school multilingual writers, as captured through automated extraction tools. Moreover, experimental research has depicted the expected CAF trade-offs in pedagogical settings and how various types of practice produced varying outcomes regarding their impact on the aspects of complexity, fluency, and accuracy (Suzuki & Kormos, 2020).

In the context of the present study, the CAF framework is particularly relevant because it provides a widely used basis for examining linguistic performance in L2 writing. However, the study does not seek to assess all three CAF dimensions. Its focus is placed on complexity, as the main concern is whether IELTS Writing Task 2 essays at different band levels display measurable differences in lexical and syntactic development. This focus is consistent with the purpose of the study, which is not to evaluate overall writing quality, but to examine whether selected textual features can help distinguish essays labeled as Bands 6, 7, and 8.

Accordingly, complexity is operationalized through three corpus-based indicators: lexical diversity, lexical sophistication, and syntactic complexity. Lexical diversity reflects the range of vocabulary used in a text, lexical sophistication captures the use of academic or advanced vocabulary, and syntactic complexity represents the structural elaboration of sentences. By narrowing the CAF framework to these observable complexity-related features, the study

maintains a clear connection between theory, measurement, and its specific aim of band-level text comparison.

Lexical Complexity: Diversity and sophistication

Within the broader construct of linguistic complexity, lexical and syntactic complexity are recognized as two focal and complementary dimensions in which grammatical competence is assessed in written production. Lexical knowledge has been extensively conceptualized as a multidimensional construct encompassing distinct but interrelated facets of vocabulary deployment (Laufer & Nation, 1995; Read, 2000). This multidimensional element has led to the distinction between LD and LS, which together reflect the breadth and depth of lexical resources used in language performance (Bulté & Housen, 2012; Kyle & Crossley, 2015).

McCarthy & Jarvis (2007) defined LD as "the range and variety of vocabulary deployed in a text". Although there are several indices to quantify LD, MTLTD was selected for this study due to its demonstrated independence from text length (McCarthy & Jarvis, 2010; Crossley & McNamara, 2012). MTLTD has been proven to be a reliable and stable estimate of LD across texts of varying lengths and demonstrates robust performance with learner language data, even with relatively brief and structurally uneven texts (Crossley & McNamara, 2012; Jarvis, 2013). Theoretically, MTLTD operationalizes LD through the rate at which new lexical items are used in a text, thereby capturing the sustained lexical variation characteristic of proficient writing (Jarvis, 2013). Many recent corpus-based studies continue to employ LD indices to track developmental patterns and instructional effects, with studies documenting developmental trajectories and competitive relationships between lexical and syntactic growth at specific proficiency stages (O'Leary & Steinkrauss, 2022).

LS, in contrast, captures the degree to which learners employ advanced, low-frequency, or academic vocabulary, thereby reflecting the depth and quality of lexical knowledge (Laufer & Nation, 1995; Read, 2000). Critically, LD and LS represent complementary yet distinct dimensions of lexical ability: a text may exhibit substantial LD through frequent variation of high-frequency vocabulary while remaining low in sophistication, or conversely, may employ a narrower lexical range with a higher proportion of advanced lexical items (Laufer & Nation, 1995; Read, 2000). In empirical L2 writing research, LS has been predominantly operationalized using frequency-based measures, particularly the proportion of vocabulary drawn from academic or low-frequency word lists. One widely adopted approach employs the Academic Word List (AWL), which captures learners' use of vocabulary beyond the most frequent word families and demonstrates strong associations with academic language proficiency (Coxhead, 2000). Previous investigations have established that AWL-based measures of LS effectively discriminate between proficiency levels in learner writing and distinguish higher- and lower-level academic texts (Laufer & Nation, 1995; Kyle & Crossley, 2015). Recent scholarship continues to document the contextual situatedness of LS, emphasizing that frequency-based and academic-vocabulary indices must be interpreted against genre and task constraints (O'Leary & Steinkrauss, 2022). Thus, the current research operationalizes LS as the content of AWLs in learners' written texts as a percentage, which is also compliant with practices in the field.

Syntactic Complexity

Together with lexical complexity, SC can be defined as the degree to which students use elaborated grammatical patterns in writing and is commonly considered to be one of the key pointers to L2 writing development (Ortega, 2003). It indicates learners' capacity to generate lengthy and complex structural sentences and clauses, thereby indicating grammatical

competence in written performance. Within CAF-oriented research, SC is commonly operationalized using observable structural features of texts, such as MSL and subordination indices (Bulté & Housen, 2012).

Although several indices have been proposed, this study adopts MSL as a main indicator of SC due to its robustness, widespread use in proficiency research, and suitability for small-scale learner corpora (Ortega, 2003; Bulté & Housen, 2012). Given the exploratory nature of the study, this selection is appropriate for reliable cross-band comparison while acknowledging that more fine-grained measures may capture additional dimensions of syntactic development.

Previous corpus-based research supports the discriminative value of SC measures across proficiency levels. For example, Khushik and Huhta (2020) found that length-based and clause-level indices distinguish lower CEFR levels, whereas phrasal and clause-level measures are more effective at higher proficiency bands. Similarly, proficiency-related variation in sentence and clause length, and subordination patterns, was also observed by Panagopoulos et al. (2024). Longitudinal studies also show that SC is dynamically developed over time and is associated with lexical complexity, which is indicative of trade-offs in how a learner assigns cognitive resources in the writing process (O’Leary & Steinkrauss, 2022; Zhang et al., 2022).

Theoretical integration

The present study adopts a theoretical position in which lexical knowledge is viewed as a multidimensional construct involving both diversity and sophistication. This position is consistent with research on lexical richness and with CAF-based approaches to L2 writing, as it provides a theoretically grounded basis for examining the lexical aspects of grammatical competence in written production (Bulté & Housen, 2012; Kyle & Crossley, 2015). The study also situates lexical and syntactic complexity within the broader CLA framework, in which grammatical competence is understood as a key component of communicative language ability. Accordingly, MTLD, AWL percentage, and MSL are used as corpus-based indicators of selected lexical and syntactic dimensions of IELTS Writing Task 2 performance.

Despite this increasing trend of research, several knowledge gaps remain to be addressed. Previous research has primarily explored broad proficiency contrasts, such as beginning versus advanced learners, or general proficiency differences across wider developmental stages (Ortega, 2003; Khushik & Huhta, 2020). However, in IELTS preparation contexts, teachers and learners are often concerned with movement from Band 6 to Band 7 or from Band 7 to Band 8. While there is a growing body of research on IELTS writing and related linguistic features, including phrasal and syntactic complexity (Wu, 2025), research that investigates adjacent higher IELTS Writing Task 2 band levels remain limited.

Second, while lexical complexity and syntactic complexity are frequently investigated together, their relative sensitivity to proficiency variation in IELTS-specific contexts is still underexplored. Emerging evidence suggests that lexical measures, particularly those capturing diversity and sophistication, may be more responsive to proficiency differences than broad length-based syntactic indices when band levels are close (Biber et al., 2016). Studies have shown that lexical diversity, academic vocabulary use, and phraseological competence can contribute to distinctions in IELTS and academic writing performance (Estaji & Hashemi, 2022). However, direct comparisons of LD, LS, and SC across Bands 6, 7, and 8 remain limited, especially in learner data from specific L2 backgrounds.

Third, IELTS writing research has often prioritized rating reliability, examiner consistency, and task effects, with comparatively less attention to corpus-based analyses of linguistic features underlying band distinctions (Hawkey, 2006; Du et al., 2025; Ilyas et al., 2025). Although

previous studies have examined aspects such as cohesion, rater agreement, and automated evaluation, there remains a need for empirically grounded studies linking measurable linguistic features to band-labeled writing performance. Methodological constraints, including small or unbalanced corpora and limited representation of higher bands, also restrict the generalizability of existing findings (Wu, 2025).

To address these gaps, the present study employs corpus-based quantitative methods to examine LD, LS, and SC in IELTS Writing Task 2 essays labeled as Bands 6, 7, and 8. It draws on Communicative Language Ability theory (Canale & Swain, 1980) and the CAF framework (Housen et al., 2012), operationalizing complexity using three measures: the Measure of Textual Lexical Diversity (MTLD) for LD (McCarthy & Jarvis, 2010), Academic Word List (AWL) percentage for LS (Coxhead, 2000), and Mean Sentence Length (MSL) for SC (Ortega, 2003). This multi-index approach responds to calls for more fine-grained corpus-based studies of writing proficiency (Khushik & Huhta, 2020; Wu, 2025).

Specifically, the study addresses the following research questions:

1. Does LD significantly differentiate IELTS Writing Task 2 essays across Bands 6, 7, and 8?
2. Is LS, measured by the proportion of AWL items, sensitive to differences between adjacent IELTS band levels?
3. Does MSL provide meaningful distinctions in SC across these band levels?

Through these research questions, the study aims to provide exploratory evidence on the linguistic features that may help distinguish IELTS Writing Task 2 essays across closely related band levels and to inform future corpus-based research on IELTS-oriented writing assessment.

Methods

Research design

The present study employed a quantitative small-scale learner corpus design to examine lexical diversity, lexical sophistication, and syntactic complexity across IELTS Writing Task 2 essays labeled as Bands 6, 7, and 8. A corpus-based design was selected because it allows systematic measurement of textual features and comparison of linguistic patterns across proficiency-related band labels. Given the modest size of the corpus, the study is intended as an exploratory analysis rather than a large-scale generalizable investigation of IELTS writing performance. Accordingly, the findings are interpreted as evidence of linguistic variation within a small set of band-labeled IELTS Task 2 essays, rather than as definitive evidence of certified IELTS proficiency development.

The use of three adjacent band levels was motivated by the pedagogical relevance of Band 6, Band 7, and Band 8 in IELTS preparation contexts. These levels represent commonly targeted stages for learners seeking to improve their writing performance. The design therefore enables a focused examination of whether selected corpus-based indices are sensitive to relatively fine-grained differences between higher-intermediate and advanced band-labeled essays.

Corpus and data selection

The dataset consisted of 30 IELTS Writing Task 2 sample essays, with 10 essays at each band level: Band 6, Band 7, and Band 8. The essays were collected from major Vietnamese IELTS preparation platforms, including DOL IELTS and KTDC IELTS, which provide sample

responses categorized by band level for instructional purposes. These sources were selected because they are widely used by IELTS candidates and teachers in Vietnamese IELTS preparation contexts.

The band scores used in this study should be understood as instructional band labels rather than independently verified official IELTS scores. The essays were not obtained from official IELTS examination scripts, and the study did not have access to examiner rating sheets, double-rating procedures, or inter-rater reliability statistics. Therefore, the band labels were treated as secondary source classifications assigned or reported by the preparation platforms. To reduce inconsistency, only essays explicitly labeled with a band score and presented as IELTS Writing Task 2 responses were included. Nevertheless, because the reliability of the original rating procedures could not be independently verified, band reliability is acknowledged as a methodological limitation.

The corpus is therefore best characterized as a collection of publicly available, band-labeled instructional IELTS Writing Task 2 sample essays rather than a corpus of authenticated test-taker scripts. As such, the findings are most directly relevant to linguistic patterns represented in IELTS preparation materials and should not be generalized uncritically to official IELTS test performance. The essays were selected from Task 2 argumentative or opinion-based writing prompts that were comparable in genre and task demands. Although the prompts were not identical across all essays, they were all consistent with IELTS Writing Task 2 requirements, requiring candidates to develop an extended written response to a social, educational, or general-interest issue. Prompt comparability was therefore addressed at the level of task type and genre rather than through the use of a single shared prompt.

Statistical analysis

Descriptive statistics (means, standard deviations) were performed first in order to set up an initial characterization regarding the different lexical and grammatical variables across bands. Subsequently, one-way ANOVA was employed in each case in order to determine whether there were statistically significant differences between three independent groups (bands 6, 7, 8). All statistical analyses were carried out in Stata, version 17. Using one-way ANOVA was appropriate in this case, since the independent variable (band level) was categorical and had more than two levels. Additionally, one-way ANOVA was preferred because, compared to multiple t-test analyses, it is safer in terms of Type I error rates, while still offering an adequate framework for group comparisons in learner corpora. Apart from significance, effect sizes were calculated in terms of partial eta squared in order to provide an adequate characterization regarding differences in lexical and grammatical variables between different bands. Tukey's HSD post-hoc tests were employed where appropriate to check differences among bands, while taking into account family-wise error rates.

Results

This section reports descriptive and inferential statistics for LD, LS, and SC across IELTS Writing Task 2 essays at Bands 6, 7, and 8. Means and standard deviations are presented first, followed by one-way ANOVA results and Tukey HSD post-hoc comparisons.

Table 1.

Descriptive statistics by IELTS band level

Measure	Band 6 (n = 10)	Band 7 (n = 10)	Band 8 (n = 10)
MTLD	89.888 (23.213)	106.703 (14.904)	127.19 (44.575)
AWL (%)	7.261 (3.225)	8.92 (2.745)	10.07 (3.441)
MSL	18.668 (2.40)	19.981 (3.062)	21.16 (2.417)

Table 1 shows the means and standard deviations of the three linguistic features across the three band-labeled groups. For LD, the mean MTLD score increased from Band 6 ($M = 89.888$, $SD = 23.213$) to Band 7 ($M = 106.703$, $SD = 14.904$) and Band 8 ($M = 127.190$, $SD = 44.575$). The relatively large standard deviation for MTLD in the Band 8 group suggests greater variability in lexical diversity among essays labeled as Band 8. A similar descriptive pattern was observed for LS. AWL percentage increased from 7.261% ($SD = 3.225$) at Band 6 to 8.920% ($SD = 2.745$) at Band 7 and 10.071% ($SD = 3.441$) at Band 8. For SC, MSL also increased slightly from 18.668 words ($SD = 2.400$) at Band 6 to 19.981 words ($SD = 3.062$) at Band 7 and 21.164 words ($SD = 2.417$) at Band 8. Overall, the descriptive statistics suggest upward trends across the three band-labeled groups, with the largest descriptive increase observed for MTLD.

Table 2.

ANOVA assumption checks

Measure	Shapiro-Wilk p Band 6	Shapiro-Wilk p Band 7	Shapiro-Wilk p Band 8	Brown-Forsythe p
MTLD	.883	.132	< .001	.544
AWL (%)	.830	.807	.339	.855
MSL	.297	.845	.065	.573

Before conducting the one-way ANOVA tests, the assumptions of normality and homogeneity of variance were examined. The results are reported in Table 2. Normality was assessed using Shapiro-Wilk tests within each band group. AWL percentage met the normality assumption across all three band-labeled groups, with p-values of .830 for Band 6, .807 for Band 7, and .339 for Band 8. MSL also met the normality assumption across the three groups, with p-values of .297, .845, and .065, respectively. Although the p-value for MSL in Band 8 was close to the conventional .05 threshold, it remained above this level. For MTLD, the normality assumption was met for Band 6 ($p = .883$) and Band 7 ($p = .132$), but not for Band 8 ($p < .001$). This suggests that MTLD scores in the Band 8 group were not normally distributed in this small sample. This result should be interpreted cautiously because each band group contained only 10 essays. Homogeneity of variance was assessed using the Brown-Forsythe test. The results did not indicate significant heterogeneity of variance for MTLD, $p = .544$, AWL percentage, $p = .855$, or MSL, $p = .573$. Therefore, the homogeneity assumption was considered acceptable for the purposes of the subsequent ANOVA tests. Given the small sample size and the non-normal distribution of MTLD in the Band 8 group, Kruskal-Wallis tests were also conducted as non-parametric robustness checks. These tests were used to examine whether the pattern of results remained broadly consistent when a distribution-free test was applied.

Table 3.

One-way ANOVA results and Kruskal-Wallis robustness checks

Measure	ANOVA F(df)	ANOVA p	Eta-squared	Kruskal-Wallis chi-square(df)	Kruskal-Wallis p
MTLD	F(2,27) = 3.81	.0349	.2201	chi-square(2) = 7.719	.0211
AWL (%)	F(2,27) = 2.01	.1535	.1296	chi-square(2) = 2.550	.2795
MSL	F(2,27) = 2.23	.1270	.1418	chi-square(2) = 3.631	.1628

Table 3 presents the one-way ANOVA results together with the Kruskal-Wallis robustness checks. For lexical diversity, the ANOVA result showed a statistically significant difference in MTLD across the three band-labeled groups, $F(2, 27) = 3.81$, $p = .0349$, $\eta^2 = .2201$. This indicates that MTLD varied across the groups in the present corpus. The effect size suggests that band label accounted for approximately 22% of the variance in MTLD scores, although this estimate should be interpreted cautiously given the small sample size. The Kruskal-Wallis test produced a similar result, $\chi^2(2) = 7.719$, $p = .0211$, suggesting that the MTLD finding was not solely dependent on the parametric ANOVA procedure. For lexical sophistication, measured by AWL percentage, the ANOVA result was not statistically significant, $F(2, 27) = 2.01$, $p = .1535$, $\eta^2 = .1296$. The Kruskal-Wallis result was also non-significant, $\chi^2(2) = 2.550$, $p = .2795$. Although AWL percentage increased descriptively from Band 6 to Band 8, these results do not provide sufficient evidence that AWL percentage differed reliably across the three groups in this dataset. For syntactic complexity, measured by MSL, the ANOVA result was also non-significant, $F(2, 27) = 2.23$, $p = .1270$, $\eta^2 = .1418$. The Kruskal-Wallis result was likewise non-significant, $\chi^2(2) = 3.631$, $p = .1628$. Therefore, although MSL showed a gradual descriptive increase across the band-labeled groups, the statistical tests did not provide sufficient evidence of a reliable group difference. Overall, the inferential results suggest that MTLD showed the clearest evidence of variation across the three band-labeled groups in this corpus. By contrast, AWL percentage and MSL showed only descriptive upward trends and did not reach statistical significance.

Table 4.

Tukey HSD post-hoc comparisons

Measure	Comparison	Mean difference	Tukey-adjusted p	95% CI
MTLD	Band 7 vs Band 6	16.815	.439	[-16.744, 50.374]
MTLD	Band 8 vs Band 6	37.302	.027	[3.743, 70.861]
MTLD	Band 8 vs Band 7	20.487	.301	[-13.072, 54.046]
AWL (%)	Band 7 vs Band 6	1.659	.476	[-1.834, 5.152]
AWL (%)	Band 8 vs Band 6	2.810	.133	[-0.683, 6.303]
AWL (%)	Band 8 vs Band 7	1.151	.696	[-2.342, 4.644]
MSL	Band 7 vs Band 6	1.313	.516	[-1.619, 4.245]
MSL	Band 8 vs Band 6	2.496	.106	[-0.436, 5.428]
MSL	Band 8 vs Band 7	1.183	.583	[-1.749, 4.115]

Note. Pairwise comparisons were conducted using Tukey's HSD test. *p*-values are Tukey-adjusted for multiple comparisons.

Table 4 reports the Tukey HSD post-hoc comparisons. These comparisons were conducted to identify which specific group differences contributed to the ANOVA results. Importantly, none of the immediately adjacent band comparisons, namely Band 6–Band 7 and Band 7–Band 8, reached statistical significance for any of the three measures.

For MTLT, the only statistically significant pairwise difference was found between Band 8 and Band 6. Essays labeled as Band 8 had higher MTLT scores than essays labeled as Band 6, mean difference = 37.302, $p = .027$, 95% CI [3.743, 70.861]. The difference between Band 7 and Band 6 was not statistically significant, mean difference = 16.815, $p = .439$, 95% CI [-16.744, 50.374]. The difference between Band 8 and Band 7 was also non-significant, mean difference = 20.487, $p = .301$, 95% CI [-13.072, 54.046]. This indicates that the MTLT result was mainly associated with the wider Band 6–Band 8 contrast, rather than with consistent differences between immediately adjacent band levels.

For AWL percentage, none of the pairwise comparisons reached statistical significance. The largest difference was between Band 8 and Band 6, mean difference = 2.810, $p = .133$, 95% CI [-0.683, 6.303]. The Band 7–Band 6 comparison, $p = .476$, and the Band 8–Band 7 comparison, $p = .696$, were also non-significant. These results suggest that AWL percentage did not provide clear evidence of pairwise differences among the three band-labeled groups.

For MSL, the post-hoc results also showed no statistically significant pairwise differences. The largest difference was observed between Band 8 and Band 6, mean difference = 2.496, $p = .106$, 95% CI [-0.436, 5.428]. The Band 7–Band 6 comparison, $p = .516$, and the Band 8–Band 7 comparison, $p = .583$, were likewise non-significant. Thus, MSL showed only a descriptive upward pattern and did not provide clear evidence of pairwise group differences in this corpus.

Taken together, the post-hoc results indicate that the only statistically significant pairwise difference was the MTLT difference between Band 6 and Band 8. The findings therefore should not be interpreted as showing that any of the three measures consistently separates adjacent IELTS band-labeled groups. A more cautious interpretation is that MTLT appeared relatively more sensitive to broader differences in this small set of band-labeled IELTS Writing Task 2 essays, whereas AWL percentage and MSL did not show statistically reliable group differences.

Discussion

This study examined the extent to which LD, LS, and SC vary across IELTS Writing Task 2 essays labeled as Bands 6, 7, and 8. The results indicate that LD is relatively more sensitive to band-level textual variation, whereas LS and SC show limited discriminative power across adjacent band levels. Given the small size of the corpus and the use of band-labeled instructional essays, these findings should be interpreted as exploratory evidence of textual variation rather than as conclusive evidence of IELTS band distinctions.

The disparity in LD, particularly in the Band 8–6 comparison, suggests that essays labeled as Band 8 tend to show greater vocabulary range and less lexical repetition than those labeled as Band 6. This finding is broadly in line with recent evidence that LD is associated with EFL writing quality and perceived writing quality (Yang et al., 2023; Wang, 2023). Yang et al. (2023) found that noun variation and other lexical richness indices contributed to the prediction of writing quality. However, the present finding should not be taken to mean that LD directly determines IELTS band scores. Rather, it suggests that MTLT may be more responsive than AWL percentage and MSL in capturing broader differences between lower- and higher-labeled essays in this dataset. The fact that the significant difference was found between Bands 6 and 8, but not between Bands 6 and 7 or Bands 7 and 8, also suggests that LD may reflect wider proficiency-related textual differences more clearly than fine-grained adjacent band differences.

LS, measured by AWL percentage, did not differ significantly across the band levels, although there was an upward trend. This could be because essays across IELTS preparation contexts may draw on similar academic or topic-related vocabulary, making frequency-based measures less discriminatory. As stated by Truong (2024), task-related variance in AWL coverage was also significant, which implies that task nature may contribute to academic vocabulary use alongside proficiency. A narrower interpretation is that AWL percentage may capture the presence of academic vocabulary, but it may not capture how appropriately, flexibly, or accurately such vocabulary is used in context. Therefore, the non-significant result does not imply that LS is unimportant in IELTS writing. It only suggests that AWL percentage alone may be too broad to distinguish closely related band-labeled essays in the present corpus. In addition, the small sample size may have limited statistical power. Future studies could therefore complement AWL-based measures with more fine-grained indicators, such as collocational accuracy and phraseological sophistication indices (Kim et al., 2025). This extension is particularly relevant to Vietnamese EFL writing because verb–noun and adjective–noun combinations have been identified as two common sources of lexical collocation errors among Vietnamese learners (Do & Le, 2023). This interpretation remains cautious, as the present study did not directly measure collocation, phraseological sophistication, or vocabulary accuracy.

The results showed no significant differences between bands in SC, which suggests that sentence length may not be a sufficiently sensitive indicator of structural sophistication in the present dataset. This cautious interpretation is also supported by research on Vietnamese EFL argumentative writing, which shows that syntactic difficulty often appears through specific error patterns, such as run-on sentences, preposition errors, article errors, sentence fragments, and pronoun errors, rather than through sentence length alone (Pham & Pham, 2024). Lahuerta Martínez (2024) suggested that longer sentences may sometimes result from coordination rather than subordination. This author also indicated that SC may plateau at higher levels while lexical complexity continues to develop. In comparison, Polat et al. (2024) found that sentence length was predictive of GPA among graduate students, indicating that sentence length may still have contextual relevance in some academic settings. In the present study, however, MSL increased gradually across Bands 6, 7, and 8, but this increase was not statistically significant. This suggests that sentence length alone may not be sufficiently sensitive to capture syntactic differences across closely related IELTS band levels. Longer sentences may result from coordination, repetition, or simple expansion, and therefore cannot automatically be interpreted as evidence of more advanced syntactic control. Future studies could use more specific measures, such as subordination ratios and phrasal elaboration, to examine syntactic complexity in greater detail (Polat et al., 2024). The present result should therefore be understood as a limitation of MSL as a single syntactic index, rather than as evidence that SC has no role in IELTS writing performance.

Taken together, the findings suggest a modest and measure-specific pattern. Among the three indices examined, MTLT showed the clearest statistical difference, whereas AWL percentage and MSL showed only non-significant upward tendencies. This pattern supports the view that complexity is multidimensional and that different complexity measures may not be equally sensitive to band-level text comparison. In particular, LD may be more visible in broader Band 6–Band 8 differences, while LS and SC may require more fine-grained measures to capture adjacent band-level variation.

Pedagogical implications

The pedagogical implications of the study should be interpreted cautiously because the study did not directly test instructional practices. The findings only suggest that IELTS writing instruction may usefully draw learners' attention to lexical variety and unnecessary repetition, since LD showed the clearest difference in this dataset. However, this should not be understood as encouraging students to insert difficult or academic vocabulary mechanically. The non-significant AWL result suggests that simply increasing the proportion of academic words may not be sufficient. Instead, vocabulary development may need to focus on appropriate word choice, paraphrasing, and context-sensitive use.

This cautious implication is broadly consistent with vocabulary-focused studies that emphasize contextual and phrase-based vocabulary development rather than isolated word-list learning. For example, Aljburi and Khaghaninejad (2024) discussed the value of associative vocabulary teaching strategies, while Boers et al. (2017) highlighted the relevance of phrase-focused vocabulary work. In relation to the present study, these studies should be understood only as pedagogical support for a cautious emphasis on meaningful vocabulary use. They are not direct evidence from the current dataset, which did not measure instructional effects, collocational accuracy, or phraseological development.

Besides, educators need to rethink suggestions that encourage longer sentences as a simple indicator of proficiency. According to Ayadi (2023), instructional attention may be placed on syntactic variety and functional constructions, such as controlled subordination and nominalization. This point is consistent with the non-significant MSL finding in the present study, which suggests that longer sentences should not be treated as a direct route to higher-band writing. A more reasonable pedagogical implication is that learners may benefit from clearer sentence control, more purposeful clause relationships, and more accurate expression of complex ideas, rather than from increasing sentence length alone.

Finally, corpus-informed feedback may have a limited role in helping learners notice lexical repetition and vocabulary range. Huong and Phuc (2023) suggested that lexical features are relevant to the quality of Vietnamese EFL students' online English writing. In an IELTS-oriented Vietnamese university context, lecturers perceived Google Docs-based writing e-portfolios as facilitating writing correction, peer review, collaborative learning, and learner engagement, while also reporting concerns about dependence on automatic correction and technological constraints (Le & Pham, 2026). In the present study, this point should be limited to lexical awareness only. Since MTLTD was the only measure showing a significant difference, the current data do not support broader claims about technology-mediated learning or overall IELTS score improvement.

Overall, the study provides exploratory evidence that LD may be more responsive than AWL-based LS and sentence-length-based SC in describing differences among the band-labeled essays examined here. The contribution of the study is not to establish fixed linguistic thresholds for IELTS bands, but to show that selected complexity measures vary in their usefulness for describing textual differences across Bands 6, 7, and 8. Future research should test these patterns with larger and more authentic IELTS datasets and employ finer-grained measures, including phraseological sophistication, subordination ratios, and clausal embedding, to provide a more comprehensive account of linguistic development across IELTS proficiency levels.

Conclusion

The study examined whether lexical diversity, lexical sophistication, and syntactic complexity help distinguish IELTS Writing Task 2 essays labeled as Bands 6, 7, and 8. The findings showed that lexical diversity, measured by MTLD, differed significantly across the three groups, with the only significant pairwise difference occurring between Bands 6 and 8. By contrast, AWL percentage and mean sentence length showed only non-significant upward trends. These findings suggest that MTLD may be more responsive than AWL percentage and sentence length in capturing broader proficiency-related differences in this small corpus, but they do not show that the selected measures consistently distinguish immediately adjacent band levels.

The findings should therefore be interpreted as exploratory evidence rather than as generalizable evidence of IELTS band distinctions. Pedagogically, the results suggest that IELTS writing instruction may benefit from encouraging lexical variety and reducing unnecessary repetition, while avoiding the mechanical insertion of academic vocabulary or the simple expansion of sentence length. However, because the study did not test instructional interventions, these implications should be treated as tentative. Future research should use larger samples, authenticated IELTS scripts, independently verified ratings, and finer-grained measures of lexical and syntactic development, such as phraseological sophistication, collocational accuracy, subordination ratios, and clausal embedding.

References

- Aljburi, M. A., & Khaghaninejad, M. S. (2024). Exploring the efficiency of associative vocabulary teaching strategies to foreign language learners. *International Journal of Instruction*, 17(3), 291–314. <https://doi.org/10.29333/iji.2024.17316a>
- Ayadi, M. (2023). Lexical richness and syntactic complexity as predictors of academic writing performance. *Journal of English Studies in Arabia Felix*, 2(1), 23–33. <https://doi.org/10.56540/jesaf.v2i1.43>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37(5), 639–668. <https://doi.org/10.1093/applin/amu059>
- Boers, F., Dang, T. C. T., & Strong, B. (2017). Comparing the effectiveness of phrase-focused exercises: A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014). *Language Teaching Research*, 21(3), 362–380. <https://doi.org/10.1177/1362168816651464>
- Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). John Benjamins. <https://doi.org/10.1075/llt.32.02bul>
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 2–27). Longman.

- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. <https://doi.org/10.1093/applin/I.1.1>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. <https://doi.org/10.2307/3587951>
- Crossley, S. A., & McNamara, D. S. (2012). Detecting the first language of second language writers using automated indices of cohesion, lexical sophistication, syntactic complexity and conceptual knowledge. In S. Jarvis & S. A. Crossley (Eds.), *Approaching language transfer through text classification* (pp. 106–126). *Multilingual Matters*. <https://doi.org/10.21832/9781847696991-005>
- Dinh, T. N. (2025). Complexity, fluency, and accuracy in written works of Vietnamese learners of English: A replication study. *International Journal of TESOL & Education*, 5(2), 29–56. <https://doi.org/10.54855/ijte.25522>
- Do, N. H. M., & Le, Q. T. (2023). Lexical collocation errors in essay writing: A study into Vietnamese EFL students and their perceptions. *International Journal of Language Instruction*, 2(2), 1–20. <https://doi.org/10.54855/ijli.23221>
- 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>
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- 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, Article 34. <https://doi.org/10.1186/s40468-022-00180-7>
- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge University Press.
- Ho, T. L. U. (2024). Problems with academic writing encountered by EFL postgraduate students at a university in the Mekong Delta. *International Journal of Language Instruction*, 3(1), 1–16. <https://doi.org/10.54855/ijli.24311>
- Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. John Benjamins. <https://doi.org/10.1075/llt.32>
- Huong, P. T. T., & Phúc, T. H. (2023). Lexical features and the quality of Vietnamese EFL students' online English writings. *Journal of University Teaching & Learning Practice*, 20(2). <https://doi.org/10.53761/1.20.02.06>
- Hyland, K. (2003). *Second language writing*. Cambridge University Press.
- IELTS. (2023). IELTS writing band descriptors. <https://ielts.org/cdn/Guides/ielts-writing-band-descriptors.pdf>
- Ilyas, D., Mimenbayeva, A., & Kadirbayeva, A. (2025). Neural network with fine-tuned BERT for IELTS writing evaluation. In *Proceedings of the 2025 International Conference on System Integration and Smart Technologies* (pp. 234–241). IEEE. <https://doi.org/10.1109/sist61657.2025.11139356>

- Imaz Agirre, A., Arias-Hermoso, R., & Ipiña, N. (2025). The effect of an intervention focused on academic language on CAF measures in the multilingual writing of secondary students. *International Review of Applied Linguistics in Language Teaching*, 63(2), 1373–1396. <https://doi.org/10.1515/iral-2023-0137>
- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63(s1), 87–106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Khudaverdiyeva, T. (2025). The importance of writing in language acquisition: A cognitive, communicative and pedagogical perspective. *Global Spectrum of Research and Humanities*, 2(3), 127–138. <https://doi.org/10.69760/gsrh.0203025014>
- Khushik, G. A., & Huhta, A. (2020). Investigating syntactic complexity in EFL learners' writing across Common European Framework of Reference levels A1, A2, and B1. *Applied Linguistics*, 41(4), 506–532. <https://doi.org/10.1093/applin/amy064>
- Kim, M., Kim, Y. J., & Kang, S. (2025). The dynamics of changes in linguistic complexity and writing scores in timed argumentative writing among beginning-level EFL learners. *International Review of Applied Linguistics in Language Teaching*, 63(4), 2891–2925. <https://doi.org/10.1515/iral-2023-0287>
- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757–786. <https://doi.org/10.1002/tesq.194>
- Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102(2), 333–349. <https://doi.org/10.1111/modl.12468>
- Lahuerta Martínez, A. C. (2024). The role of syntactic and lexical complexity in undergraduate writing quality. *Ibérica*, 47, 251–274. <https://doi.org/10.17398/2340-2784.47.251>
- Larsen-Freeman, D. (2019). On language learner agency: A complex dynamic systems theory perspective. *The Modern Language Journal*, 103(S1), 61–79. <https://doi.org/10.1111/modl.12536>
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322. <https://doi.org/10.1093/applin/16.3.307>
- Le, D. M. T., & Pham, T. S. (2026). Lecturers' perceptions of Google Docs for e-portfolios to improve students' IELTS writing. *International Journal of Language Instruction*, 5(1), 18–44. <https://doi.org/10.54855/ijli.26512>
- McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, 24(4), 459–488. <https://doi.org/10.1177/0265532207080767>
- McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392. <https://doi.org/10.3758/BRM.42.2.381>
- Nguyen, H. N., & Nguyen, D. K. (2022). Vietnamese learners' performance in the IELTS Writing Task 2: Problems, causes, and suggestions. *International Journal of TESOL & Education*, 2(1), 170–189. <https://doi.org/10.54855/ijte.222111>
- Nguyen, H. V., Nguyen, M. X. N. C., & Dao, P. (2020). The impact of IELTS on English language teachers in Central Vietnam (IELTS Research Reports Online Series No. 1).

- British Council, Cambridge Assessment English, and IDP: IELTS Australia.
<https://www.ielts.org/teaching-and-research/research-reports>
- O'Leary, M., & Steinkrauss, R. (2022). Syntactic and lexical complexity in L2 English academic writing: Development and competition. *Ampersand*, 9, Article 100096.
<https://doi.org/10.1016/j.amper.2022.100096>
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518.
<https://doi.org/10.1093/applin/24.4.492>
- Ortega, L. (2012). Epilogue: Exploring L2 writing–SLA interfaces. *Journal of Second Language Writing*, 21(4), 404–415. <https://doi.org/10.1016/j.jslw.2012.09.002>
- Panagopoulos, P., Andria, M., Mikros, G., & Varlokosta, S. (2024). Writing in L2 Greek: Exploring the effect of L2 proficiency and learning context on complexity, accuracy, and fluency. *Journal of Second Language Writing*, 64, Article 101111.
<https://doi.org/10.1016/j.jslw.2024.101111>
- Pham, Q. N., & Pham, V. P. H. (2024). Common syntactic errors in the writing skills of the 3rd year students at Van Lang University. *International Journal of TESOL & Education*, 4(2), 118–144. <https://doi.org/10.54855/ijte.24427>
- Phuoc, N. N., & Barrot, J. S. (2022). Complexity, accuracy, and fluency in L2 writing across proficiency levels: A matter of L1 background? *Assessing Writing*, 54, Article 100673.
<https://doi.org/10.1016/j.asw.2022.100673>
- Polat, N., Mahalingappa, L., & Mancilla, R. (2024). IELTS and written syntactic complexity as predictors of GPA of multilingual international graduate students. *AERA Open*, 10, 1–16. <https://doi.org/10.1177/23328584241258729>
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.
- Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, 42(1), 143–167.
<https://doi.org/10.1017/S0272263119000421>
- Tran, T. D. (2015). *An exploratory study of current assessment practices for improving the learning of English as a foreign language (EFL) in two Vietnamese universities* [Doctoral dissertation, Queensland University of Technology]. <https://eprints.qut.edu.au/84622/>
- Truong, T. V. (2024). Analyzing lexical characteristics in IELTS Writing Task 2 essays: A study of Vietnamese students from non-English majors. *International Journal of Research Publication and Reviews*, 5(12), 3233–3245.
<https://doi.org/10.55248/gengpi.5.1224.250102>
- Wang, Y. (2023). The interplay of linguistic complexity measures and their relation to the quality of L2 argumentative writing. *Communications in Humanities Research*, 11, 31–39. <https://doi.org/10.54254/2753-7064/11/20231365>
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. University of Hawai'i Press.
- 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>

Yang, Y., Yap, N. T., & Ali, A. M. (2023). Predicting EFL expository writing quality with measures of lexical richness. *Assessing Writing*, 57, Article 100762. <https://doi.org/10.1016/j.asw.2023.100762>

Zhang, S., Zhang, H., & Zhang, C. (2022). A dynamic systems study on complexity, accuracy, and fluency in English writing development by Chinese university students. *Frontiers in Psychology*, 13, Article 787710. <https://doi.org/10.3389/fpsyg.2022.787710>

Biodata

Truong Van Uyen is an independent researcher. She holds an MA in Education from Leeds Beckett University, UK, and a BA in Business English from Ho Chi Minh City University of Education, Vietnam. Her research interests include English language teaching methodology and course evaluation.