

A Corpus-Based Contrastive Study of EFL and ESL Argumentative Writing: Syntactic, Lexical, Cohesive, and Readability Profiles

Tan Liangshuang^{1,2}, Sabariah Abd Rahim¹

¹Universiti Malaysia Sabah, Malaysia, ²Heihe University, China

Corresponding Author Email: sab@ums.edu.my

DOI Link: <http://dx.doi.org/10.6007/IJARPED/v15-i3/28927>

Published Online: 14 September 2026

Abstract

This study investigates the multidimensional linguistic profiles of English as a Foreign Language (EFL) and English as a Second Language (ESL) undergraduate writers. Analysing a corpus of 264 argumentative essays written by Chinese EFL and Malaysian ESL undergraduates under identical timed conditions, the study utilizes Coh-Metrix 3.0 to evaluate syntactic complexity, lexical diversity, discourse cohesion, and readability. The results reveal that ESL writers produce syntactically more complex, longer sentences with stronger referential cohesion and deep semantic coherence, reflecting greater textual maturity and stronger written-register control. Conversely, EFL writers show higher lexical diversity scores, but this pattern appears to be closely related to shorter text length and a tendency to introduce loosely connected concepts rather than to sustain and develop core ideas across the essay. EFL writers also over-use explicit formulaic connectives to compensate for less implicit semantic continuity. The results offer theoretical connections between socio-educational ecologies and L2 cognitive processing and suggest the need for multidimensional text assessment and pedagogies that encourage sustained conceptual and referential continuity rather than explicit transitional templates.

Keywords: Coh-Metrix, Argumentative Writing, Syntactic Complexity, Lexical Diversity, Semantic Coherence

Introduction

Writing is a cognitively demanding activity and a key component of academic, professional, and educational success in second language (L2) learning (Hyland, 2019; Manchón, 2011). Among the major genres required in higher education, argumentative writing is particularly important because it requires students to formulate claims, justify positions, organise evidence, and construct logical relations between ideas (Cumming, 2001; Qin & Uccelli, 2016). These demands require writers to coordinate linguistic resources across multiple levels, including syntactic complexity, lexical choice, cohesion, semantic coherence, and readability. Such written profiles are shaped not only by individual proficiency but also by the socio-educational contexts in which English is learned and used (Berei, 2025; Halim & Wu, 2025). EFL learners often rely more heavily on classroom-based and assessment-oriented exposure

to English, whereas ESL learners may encounter English across wider educational, institutional, and social domains (Mao et al., 2025; Thirusanku & Yunus, 2014). These contextual differences may not imply a simple ESL advantage, but they suggest that EFL and ESL undergraduates may develop different linguistic profiles and stylistic preferences when completing similar argumentative writing tasks.

Corpus-based L2 writing research has contributed substantially to identifying linguistic features associated with writing development and text quality (e.g., Crossley & McNamara, 2014; Kargar, 2025; Lu, 2011; Staples & Reppen, 2016; Yoon & Polio, 2017). However, many learner-corpus studies have compared non-native-speaker writing with native-speaker baselines, which may obscure important variation among L2 learner populations by treating L2 writers as a relatively homogeneous group (Lu & Ai, 2015). Contrastive Interlanguage Analysis (CIA) provides a useful framework for comparing different learner populations and examining interlanguage variation in relation to learners' linguistic, educational, and contextual backgrounds (Gilquin & Granger, 2015; Granger, 2015). At the same time, because L2 writing development is multidimensional, single-index comparisons may provide only a partial account of learner writing. Coh-Metrix is therefore useful because it generates indices related to syntactic complexity, lexical characteristics and referential cohesion within one computational framework (Crossley & McNamara, 2011; Graesser et al., 2011; McNamara et al., 2014). Although Coh-Metrix has been widely used in L2 writing research, direct empirical comparisons between EFL and ESL undergraduates under identical task conditions remain relatively limited, especially when multiple textual dimensions are examined together (Fan & Gao, 2026; Gilquin & Granger, 2015; Mahadini et al., 2021). Accordingly, this study compares Chinese EFL and Malaysian ESL undergraduate argumentative writing under identical timed conditions, guided by the following research question:

To what extent do EFL and ESL undergraduate writers differ in the ten Coh-Metrix indices in this study?

Literature Review

Coh-Metrix Applications in L2 Writing Assessment

In the field of second language (L2) acquisition, the assessment of L2 writing has increasingly moved from traditional, labour-intensive manual evaluations toward multidimensional, computational analyses. Among the tools that have contributed to this methodological development, Coh-Metrix has been widely used as a natural language processing (NLP) tool for analysing text across multiple levels of language, discourse, and cognitive representation (Crossley & McNamara, 2011). Unlike traditional surface-level metrics that rely mainly on word and sentence length, Coh-Metrix operationalizes discourse processing theories by generating a wide range of indices related to lexical sophistication, syntactic complexity, referential cohesion, and semantic coherence (Graesser et al., 2011; Lu, 2017). This computational approach allows L2 researchers to investigate patterns of L2 writing development and to describe the linguistic profiles of diverse learner populations with greater consistency and replicability (Aryadoust, 2016).

One major line of empirical work with Coh-Metrix has been to investigate the extent to which the computationally derived indices of linguistic features are associated with human judgments of L2 writing quality. Automated metrics of lexical diversity, word frequency, and syntactic complexity have been shown to correlate with and sometimes predict holistic and

analytic ratings provided by human raters (Janebi Enayat, 2025; Matthews & Wijeyewardene, 2018). For example, lexical sophistication and phrasal complexity appear to have a stronger relationship with writing quality ratings than the simple frequency of surface-level cohesive devices, which L2 writers can overuse without necessarily improving the coherence of the text (Crossley & McNamara, 2011). Research based on Coh-Metrix has helped to develop a more detailed understanding of L2 writing ability by identifying linguistic features that are associated with rated writing quality, and has provided useful indicators for tracking learners' development of lexical, syntactic and discourse-level features over time (Aryadoust, 2016; Crossley & McNamara, 2014). Together, these results suggest that Coh-Metrix offers a systematic and replicable way to track L2 writing development across multiple dimensions of language.

Coh-Metrix-Based Indicators of Linguistic Complexity, Cohesion, and Readability

Syntactic complexity and lexical richness are two important dimensions of L2 writing quality that reflect the structural and lexical resources that writers use to construct written discourse (Crossley & McNamara, 2011; Jarvis & McNamara, 2012). Previous studies have used global metrics like subordination ratios. However, Coh-Metrix provides automated indices that assess different facets of syntactic complexity such as sentence length, syntactic simplicity, and structural elaboration (Lu, 2017; Polio & Yoon, 2018). At the global level, the average sentence length (DESSL) and syntactic simplicity index (PCSYNz) represent sentence-level structural patterns to account for the difference between the simpler repetitive constructions and the more sophisticated clausal constructions (Kim & Lim, 2019; McNamara et al., 2010; Xue et al., 2021). Both LDMTLD, a length-independent word-level measure, and PCCNCz, a word concreteness measure, reflect the cognitive efforts of the vocabulary choice that writers must make, moving from basic concrete words to more abstract registers of academic language (Crossley & Kyle, 2018; Nasserri & Thompson, 2021).

Quality of text depends not only on lexical and sentence-level features, but also on discourse organization, in particular the distinction between explicit cohesion and semantic coherence. Logical connectives (e.g. PCCONNP) may be used by L2 writers to show the relationship between ideas. However, computational analysis indicates that the higher rated writing might be more associated with CRFNO1 (referential overlap) and LSASS1 (Latent Semantic Analysis) cosines (Aljohani, 2026). Such a pattern can be explained by the lack of semantic continuity in the use of explicit cohesive devices, especially in the setting of foreign language learning, where learners tend to use formulaic transitional expressions (Pu et al., 2023; Xie, 2026). Comparing explicit and implicit cohesion measures can therefore help examine how different learning environments shape discourse-level organization.

Readability provides a supplementary perspective on the surface complexity of L2 texts. The present study uses a surface-based measure of Flesch-Kincaid Grade Level (RDFKGL). It measures the readability by the length of sentences and words to estimate the difficulty of the text (Kiselnikov et al., 2020). RDFKGL is not an overall measure of discourse quality or academic register, but it can be a useful supplementary measure of overall textual complexity (Chen & Luo, 2025; Crossley et al., 2008). Researchers can, by analysing this readability index, explore variation in the surface difficulty of learner texts linked to syntactic, lexical and cohesive features.

Linguistic Variations Across EFL and ESL Contexts

A common comparative approach in learner-corpus studies of L2 writing has been to compare non-native speaker (NNS) texts to a native-speaker English (ENL) baseline (e.g., Ai & Lu, 2013; Shen et al., 2023; Silva, 1993). ENL-baseline studies have demonstrated general linguistic and rhetorical differences between L1 and L2 writers. However, this approach has been criticized for treating L2 writers as a fairly uniform group. Lu and Ai (2015) claim that such comparisons may mask variation related to learners' first language (L1) background, proficiency level, and learning context. To tackle this challenge, Contrastive Interlanguage Analysis (CIA) has been used to compare different populations of L2 learners and investigate variation among learner varieties (Granger, 2015; Ha, 2023).

Furthermore, research with the use of multidimensional linguistic measures has suggested that L2 writing development might not be linear and unidirectional in different learning environments at all times (Becker & Nekrasova-Beker, 2026; Zhang & Li, 2025). Some EFL samples have been reported to show a relatively greater reliance on clausal elaboration and overt cohesive marking with more constrained patterns of lexical diversity (Fan & Gao, 2026). By contrast, studies of ESL writing have often examined learners' broader exposure to English in relation to lexical variety, phraseological range, grammatical accuracy, and register control, although findings across these dimensions are not always consistent (Abdi Tabari & Lee, 2026; Nasser & Thompson, 2021). Rather than assuming a general ESL advantage, this line of research suggests that EFL and ESL contexts may be associated with different linguistic profiles across syntactic, lexical, and discourse-level dimensions (Abdi Tabari & Huang, 2026; Lu & Ai, 2015). These studies highlight the value of multidimensional approaches for comparing interlanguage profiles across learning contexts.

Methodology

Corpus Description

To analyse multi-level linguistic variations in L2 academic writing, this study utilized learner corpora from two distinct environments: an EFL context in China and an ESL context in Malaysia. The EFL sub-corpus comprises argumentative essays by Chinese undergraduates in Northeastern China. The English these native Chinese speakers acquired was mainly through classroom teaching and they had limited opportunities to use English in their daily life other than schoolwork (Chen et al., 2020; Zou et al., 2024). The ESL sub-corpus, in contrast, consists of essays written by undergraduates at a Malaysian public university. The students use English as a de facto second language in administrative, educational and social domains that provide them with more regular opportunities for daily and academic language use (Gill, 2002; Thirusanku & Yunus, 2014). The writing tasks were carried out under controlled classroom conditions in both groups. Participants were asked to write argumentative essays in response to specific prompts in a timed session of 30 minutes. The prompts used in both settings were the same, asking students to express their opinions on familiar topics and social issues (e.g. technology, tourism), and justify them. No external reference materials, dictionaries, or translation tools were used in the writing process.

Sample Balancing and Data Cleaning

Prior to the computational analysis by Coh-Metrix 3.0, data cleaning and filtering procedures were performed so as to guaranty the corpus's consistency. Firstly, all raw essay files in .txt format were normalized by correcting obvious typos without altering grammatical structure

or lexical choices, correcting formatting inconsistencies such as irregular spacing, line breaks and encoding errors, and removing non-standard punctuation marks or characters. We removed non-textual elements (e.g., student IDs, titles, and copied prompts) to avoid affecting Coh-Metrix readability and lexical indices. The final corpus consisted of 264 valid essays, comprising 135 essays from the Chinese EFL group and 129 essays from the Malaysian ESL group. This dataset was then used for the subsequent comparative analyses.

Selected Measures

To compare linguistic variation in essays written by EFL and ESL learners, this study selected 10 representative indices from Coh-Metrix 3.0 (Graesser et al., 2004; McNamara et al., 2014; see Table 1). Following text and discourse processing frameworks, these indices were categorized into four dimensions: (1) descriptive attributes and syntactic complexity, capturing sentence length (DESSL) and syntactic simplicity (PCSYNz); (2) lexical diversity and word concreteness, measured through the relatively length-resistant Measure of Textual Lexical Diversity (LDMTLD) and word concreteness (PCCNCz); (3) cohesion and coherence, represented by logical connectives (PCCONNp), local referential overlap (CRFNO1), and semantic similarity at the sentence and paragraph levels based on Latent Semantic Analysis (LSASS1, LSASSp); and (4) readability, represented by the Flesch-Kincaid Grade Level (RDFKGL) as an indicator of textual difficulty.

Table 1

Selected Indices and Descriptions

No.	Index	Description
Descriptive Attributes & Syntactic Complexity		
1	DESWC	Word count: Total number of words in essay.
2	DESSL	Sentence length: Mean number of words per sentence.
3	PCSYNz	Syntactic simplicity: Coh-Metrix index of syntactic simplicity; higher values indicate less syntactically complex structures.
Lexical Diversity & Word Concreteness		
4	LDMTLD	Lexical diversity: Measure of Textual Lexical Diversity; estimates lexical variety with reduced sensitivity to text length.
5	PCCNCz	Word concreteness: Mean concreteness rating of content words, indicating how concrete or abstract the vocabulary is.
Cohesion & Coherence		
6	CRFNO1	Noun overlap: Proportion of adjacent sentence pairs sharing one or more nouns.
7	PCCONNp	Connective density: Incidence of logical connectives used to mark relations between clauses, sentences, or ideas.
8	LSASS1	Adjacent-sentence semantic similarity: Mean LSA cosine between adjacent sentences.

9	LSASSp	Intra-paragraph semantic similarity: Mean LSA cosine among sentence pairs within paragraphs.
Readability		
10	RDFKGL	Flesch-Kincaid Grade Level: Surface-based readability estimate derived from sentence length and syllables per word.

Data Analysis Process

Normality of each index was tested with the Kolmogorov–Smirnov and Shapiro–Wilk tests. The results showed that some indices such as PCSYNz in both groups, and DESWC in the EFL group did not significantly differ from normal distribution ($p > .05$), but several other indices differed significantly from normality ($p < .05$). However, parametric tests such as the independent-samples *t*-test are usually robust to moderate violations of normality, provided the sample sizes are large and sufficiently similar (Field, 2018). Given the relatively large and comparable sample sizes of the two groups (EFL: $N = 135$; ESL: $N = 129$), the independent-samples *t*-test was considered appropriate for the present analysis. Therefore, two-tailed independent-samples *t*-tests were conducted to examine differences between the EFL and ESL groups. To reduce overreliance on statistical significance, Cohen’s *d* was interpreted alongside *p*-values for all comparisons.

Results

This section presents the statistical results of the independent-samples *t*-tests and effect size estimations (Cohen’s *d*) comparing the writing performance of Chinese EFL learners ($N = 135$) and Malaysian ESL learners ($N = 129$) across 10 selected Coh-Metrix indices. The results are organized according to the four dimensions established in Chapter 3: (1) descriptive attributes and syntactic complexity, (2) lexical diversity and word concreteness, (3) cohesion and coherence, and (4) readability.

To ensure statistical rigor, Levene’s test for equality of variances was performed for each comparison. Wherever the assumption of equal variances was violated ($p < .05$), the corrected *t*-values, degrees of freedom (*df*), and two-tailed significance levels (*p*) under the “Equal variances not assumed” condition were reported and interpreted. Table 2 summarizes the descriptive statistics, *t*-test results, and Cohen’s *d* effect sizes for all 10 investigated indices.

Table 2

Descriptive and Inferential Statistics for Coh-Metrix Indices between EFL and ESL Groups

Index	EFL (N = 135) M (SD)	ESL (N = 129) M (SD)	t	df	p	Cohen's d
Descriptive Attributes & Syntactic Complexity						
DESWC	128.84 (30.37)	159.74 (50.20)	-6.019	208.77	< .001	-0.749
DESSL	12.95 (4.09)	16.65 (3.72)	-7.682	262	< .001	-0.946
PCSYNz	1.07 (0.88)	0.41 (0.80)	6.333	262	< .001	0.780
Lexical Diversity & Word Concreteness						
LDMTLD	94.96 (40.53)	72.48 (31.94)	5.017	252.95	< .001	0.614
PCCNCz	-0.20 (0.88)	-0.30 (0.84)	0.893	262	.373	0.110
Cohesion & Coherence						
CRFNO1	0.23 (0.19)	0.51 (0.28)	-9.529	225.92	< .001	-1.183
PCCONNp	3.30 (9.93)	6.17 (15.51)	-1.784	216.22	.076	-0.222
LSASS1	0.16 (0.07)	0.26 (0.09)	-9.536	241.67	< .001	-1.181
LSASSp	0.14 (0.08)	0.25 (0.09)	-10.193	252.57	< .001	-1.259
Readability						
RDFKGL	8.17 (2.84)	9.68 (1.75)	-5.237	224.48	< .001	-0.638

Descriptive Attributes and Syntactic Complexity

As illustrated in Table 2, the independent samples *t*-test revealed highly significant cross-contextual differences in basic text descriptive attributes and syntactic complexity between the two learner groups. Regarding text length (DESWC), Malaysian ESL learners produced significantly longer essays ($M = 159.74$, $SD = 50.20$) than Chinese EFL learners ($M = 128.84$, $SD = 30.37$), $t(208.77) = -6.019$, $p < .001$, with a medium-to-large effect size (Cohen's $d = -0.749$). In terms of sentence-level complexity, the average sentence length (DESSL) of the ESL group ($M = 16.65$, $SD = 3.72$) was significantly higher than that of the EFL group ($M = 12.95$, $SD = 4.09$), $t(262) = -7.682$, $p < .001$, representing a very large effect size (Cohen's $d = -0.946$).

In line with this finding, the syntactic simplicity index (PCSYNz) showed a statistically significant difference, $t(262) = 6.333$, $p < .001$, with a large effect size (Cohen's $d = 0.780$). The mean score of the EFL group was significantly higher ($M = 1.07$, $SD = 0.88$) than that of

the ESL group ($M = 0.41$, $SD = 0.80$). The results suggest that Chinese EFL learners tended to rely more on simpler syntactic constructions, whereas Malaysian ESL learners make use of more complex and varied syntactic structures, as a higher PCSYNz score indicates simpler and more frequent syntactic structures.

Lexical Diversity and Word Concreteness

Two measures of lexical complexity were captured by two indicators: lexical diversity (LDMTLD) and word concreteness (PCCNCz). There was a highly significant difference for lexical diversity as measured by the robust MTLD index (LDMTLD), $t(252.95) = 5.017$, $p < .001$, with a medium to large effect size (Cohen's $d = 0.614$). Interestingly, Chinese EFL learners displayed significantly higher lexical diversity ($M = 94.96$, $SD = 40.53$) than Malaysian ESL learners ($M = 72.48$, $SD = 31.94$).

Conversely, no statistically significant difference was found in word concreteness (PCCNCz) between the EFL group ($M = -0.20$, $SD = 0.88$) and the ESL group ($M = -0.30$, $SD = 0.84$), $t(262) = 0.893$, $p = .372$, with a negligible effect size (Cohen's $d = 0.110$). As a result, despite the EFL group's tendency to deploy a wider range of unique lexical items, learners from both contexts demonstrate a similar baseline in their use of abstract academic vocabulary.

Textual Cohesion and Semantic Coherence

The statistical analysis yielded the most striking differences in the domain of discourse-level cohesion and coherence, where the largest effect sizes in this study were recorded. For local referential cohesion, noun overlap (CRFNO1) was significantly higher in the ESL group ($M = 0.51$, $SD = 0.28$) than in the EFL group ($M = 0.23$, $SD = 0.19$), $t(225.92) = -9.529$, $p < .001$. The effect size was exceptionally large (Cohen's $d = -1.183$), indicating a substantial difference in how the two groups establish lexical ties between adjacent sentences.

In contrast, the normalized frequency of logical connectives (PCCONNp) showed only a marginal, non-significant difference at the conventional .05 level, $t(216.22) = -1.784$, $p = .076$, with a small effect size (Cohen's $d = -0.222$). Descriptively, the ESL group ($M = 6.17$, $SD = 15.51$) used slightly more logical connectives than the EFL group ($M = 3.30$, $SD = 9.93$). For LSA-based semantic similarity, both indices showed statistically significant differences. Adjacent-sentence LSA similarity (LSASS1) was significantly higher in the ESL group ($M = 0.26$, $SD = 0.09$) than in the EFL group ($M = 0.16$, $SD = 0.07$), $t(241.67) = -9.536$, $p < .001$, with a large effect size (Cohen's $d = -1.181$). Similarly, intra-paragraph sentence-level LSA coherence (LSASSp) was also significantly higher for the ESL group ($M = 0.25$, $SD = 0.09$) than for the EFL group ($M = 0.14$, $SD = 0.08$), $t(252.57) = -10.193$, $p < .001$, yielding the largest effect size in the entire study (Cohen's $d = -1.259$). These findings reveal that ESL learners construct texts with vastly stronger global semantic continuity and conceptual flow compared to their EFL counterparts.

Overall Readability

The Flesch–Kincaid Grade Level (RDFKGL) was used as a surface-based indicator of readability, primarily reflecting sentence length and word-level complexity. The independent-samples t -test revealed a statistically significant difference between the two groups, $t(224.48) = -5.237$, $p < .001$, with a medium-to-large effect size (Cohen's $d = -0.638$). The Malaysian ESL group produced texts with a higher estimated grade level ($M = 9.68$, $SD = 1.75$) than the Chinese EFL

group ($M = 8.17$, $SD = 2.84$), suggesting that the ESL essays were, on average, more complex according to conventional readability measures.

Discussion

This section discusses the findings for RQ: To what extent do EFL and ESL undergraduate writers differ in the ten Coh-Metrix indices in this study?

Syntactic Complexity and Information Packaging

Corpus analysis revealed a marked contrast in syntactic complexity and sentence length between the two cohorts. In particular, the Malaysian ESL group produced significantly longer and more syntactically complex sentences than their Chinese EFL counterparts, which may suggest that they have higher syntactic maturity and hierarchical clausal elaboration capability (Chung & Jeon, 2026; McNamara et al., 2014). Qualitative inspection of the Chinese EFL sub-corpus suggests a relatively conservative, risk-averse writing style, typically characterized by repetitive, single-clause “subject-verb-object” structures. More importantly, whenever EFL learners attempt to break away from this structural flatness to express complex logical relationships, their limited syntactic control often leads to grammatical breakdown.

The following examples in their writings show the learners’ grammatical breakdown due to the limitations of syntactic control.

(1) *“Secondly, though appear technology, we always need study, Technology just in some respects help us, but we must study rich our mind.”* (EFL_06)

(2) *“...like walk through thousands of kilometers, people create horse-drawn carriage, so people can across thousands of kilometers easily, but it doesn't mean technology making people lazy.”* (EFL_30)

These patterns indicate problems in syntactic encoding control when more than one proposition is being formulated in a single sentence. The attempted concessive clause in Example 1 disintegrates into a verbless fragment (“though appear technology”). In Example 2, the combination of several propositions (exemplification, causality, and contrast) produces a 51-word run-on sentence that lacks a core verb (“can across”). Kormos (2023) cited that cognitive models of L2 writing have demonstrated that the process of formulating complex ideas during writing is very demanding on working memory. When EFL learners try to enlarge basic SVO templates, they may suffer from overloading their attention resources, which leads to a trade-off where syntactic expansion is favoured at the expense of grammatical accuracy (Francis et al., 2002; Ruiz-Funes, 2015). Conversely, the Malaysian ESL cohort appeared better able to use clausal elaboration and phrasal compression to package multiple propositions within single sentence boundaries. Respond in one piece of writing:

“For me I believe that technology making people lazy it is because technology like media social such as Instagram, WhatsApp, Facebook can less face to face interaction... So, this can affect people to become lazy person.” (ESL_13)

Despite local grammatical infelicities, ESL_13 demonstrates syntactic elongation to accommodate complex logical relationships. The writer attempts to coordinate several ideas so as to create argumentative depth by embedding a prepositional list (“like media social such as...”) into a subordinate causal clause. This ability to strategically combine different kinds of informational units and to embed complex propositions within a single clause is one of the

hallmarks of advanced multilingual writing and correlates with higher argumentative text quality (Rioux & Thordardottir, 2026). The deep immersion quality of the ESL context may support such structural agility by fostering the lexical and syntactic flexibility required for online clausal elongation, whereas classroom-bound EFL settings tend to reinforce a more conservative, form-focused orientation (Choi et al., 2025). To conclude, these patterns suggest that different learning ecologies may shape attentional allocation during writing, with EFL learners oriented more toward local grammatical control and ESL learners better able to engage in broader proposition packaging.

Inflated Lexical Diversity and Conceptual Retrieval

The lexical complexity results present a notable interpretive tension. Contrary to the common belief that language proficiency is positively correlated to lexical variation (Peng et al., 2023), the statistical analysis results showed that the Chinese EFL group produced significantly more lexically diverse texts than the Malaysian ESL group. Micro-level qualitative inspection reveals that this higher diversity score should not be interpreted straightforwardly as evidence of superior lexical proficiency, but rather a structural byproduct of EFL learners' shorter text length and distinct conceptual retrieval mechanisms under timed constraints. As noted in L2 writing research, writers operating under higher cognitive constraints often struggle with real-time lexical retrieval, which can lead to a more fragmented pattern of idea development (Murphy & Roca de Larios, 2010; Yi & Ni, 2015). In response to this retrieval difficulty, these EFL writers frequently resort to "conceptual hopping"—introducing a wide array of non-repeated content words within a very short span. This prevents the traditional type-token ratio (TTR) from decaying and mathematically elevates lexical diversity scores in shorter texts (Ong & Zhang, 2010).

For instance, in essay EFL_19, the writer rapidly transitions from "books" and "words" to "cultures" and "mind" without elaboration: *"First, traveling help us get new knowledge. Books have many words but traveling make those words become real... second, traveling show us different cultures... and this open our mind"*. Conversely, Malaysian ESL learners produce longer, more cohesive texts characterized by "conceptual looping"—the systematic reintroduction and expansion of core conceptual entities (Crossley et al., 2012). This is exemplified in ESL_14, where the writer repeatedly develops and varies key concepts: *"...traveling exposes people to different cultures... By meeting new people... we gain a deeper understanding of the world and learn to appreciate diversity"*. While this conceptual looping mathematically reduces lexical diversity, it yields a more cohesive and rhetorically mature text structure valued by human raters; however, some computational assessments may not fully capture this maturity, as they typically associate higher essay quality with greater lexical diversity and lower text cohesion (Crossley & McNamara, 2011).

At the level of word concreteness, the two groups exhibited no statistically significant difference in PCCNCz, with both groups scoring in the negative (abstract) range. The alignment in style suggests that both groups have developed similar awareness of register, moving away from more conversational language to abstract academic discourse (Granados et al., 2022). However, the underlying cognitive processing remains distinct. In our study, the Chinese EFL learners' deployment of abstract words (e.g., "convenience," "communication") appears largely template-based. This is consistent with recent research in phraseology, which has found that L2 writers often use global lexical bundles to organize their argumentative

structure, but these formulaic sequences do not necessarily correspond to greater lexical specificity or precision (Aljohani, 2026). Rather, these abstract terms may be associated with memorized templates or with L1-mediated lexical choices. By contrast, for ESL learners, the use of abstract language (e.g., “diversity,” “barriers”) is more organic and integrated into a denser, highly accessible lexical network (Crossley et al., 2012), allowing them to weave these terms into cohesive, logically sustained arguments.

Surface Cohesion and Semantic Coherence

The different ways in which the two cohorts integrate abstract concepts are also reflected in the structural organization of their essays. While the density of explicit logical connectives (PCCONNp) shows no significant difference between the two cohorts, the Malaysian ESL group demonstrates a clear advantage in establishing implicit textual continuity. Specifically, the ESL writers have significantly higher scores in local referential cohesion (noun overlap, CRFNO1), adjacent-sentence semantic coherence (LSASS1), and intra-paragraph semantic coherence (LSASSp), all comparisons yielding very large effect sizes. This empirical pattern corresponds to the traditional distinction between surface-level cohesion and deep-level coherence (Crossley & McNamara, 2011). The Chinese EFL and Malaysian ESL learners are similar in their use of explicit logical signposts but differ greatly in their ability to build the implicit semantic overlap for true textbase continuity.

The Chinese EFL texts thus show a tendency to mark logical relations explicitly without consistently maintaining the semantic links needed for conceptual continuity. As Pu et al. (2023) and Wang and Xie (2022) have pointed out, EFL writers can compensate for poor topic development and global coherence through the use of local cohesive devices, particularly connectives. This superficial scaffolding is in line with Aljohani’s (2026) finding that global organizational markers can signal argumentative structure without contributing to local referential alignment. The transitions of the EFL writers are conceptually abrupt for there are few nouns shared between consecutive sentences (CRFNO1) and the readers have to exert additional cognitive effort to infer the missing semantic links (Crossley & McNamara, 2011). Conversely, ESL writers seem to build cohesive “conceptual chains” through systematic referential looping and semantic recurrence, perhaps minimizing the cognitive load on the reader.

This difference in semantic coherence may be partially associated with L1 rhetorical transfer and different cognitive demands in production. In the hypothetical context of Chinese as a topic-prominent language, EFL learners might be more likely to change topics at the clause boundary, resulting in more fragmented referential chains (Appel & Szeib, 2018). Moreover, the cognitive load of word retrieval at the level of the word may be greater than the working memory capacity for global cohesion checking in classroom-based EFL, which may lead to the use of memorized templates as a compensatory strategy (Pu et al., 2023). Conversely, Malaysian ESL writers, who are generally exposed to more naturalistic L2 input, may have more automated lexical networks. Such processing efficiency may release attentional resources to preserve an organic, associative semantic flow, without rigid structural scaffolding.

Readability as Cumulative Textual Maturity

In the final analysis, it is found that the Malaysian ESL group produced texts with a significantly higher Flesch-Kincaid Grade Level (RDFKGL) than the Chinese EFL group, in terms of overall readability and linguistic maturity. The gap on this metric involves a fundamental difference in textual maturity, indicating that ESL writers are more successful in the transition from conversational, spoken-like structures to formal, written academic prose. The criticism against traditional readability formulas is that they are based on surface-level metrics (e.g., sentence length, word length) rather than more meaningful measures of cognitive processing (Crossley et al., 2008; McNamara et al., 2002). But they can also rely on parts of structural elaboration because their computational components are direct representations of basic syntactic and lexical packaging (Kiselnikov et al., 2020).

The textual profile (relatively short, syntactically simple, short sentences, low informational density) is consistent with the lower grade level scores of the EFL writers in our corpus. Malaysian ESL writers, in contrast, demonstrated more mature structural characteristics with dense information, longer sentences, and more complex structures of clauses and phrases (Crossley et al., 2008; Nkhobo & Chaka, 2023). This interpretation is consistent with the findings of studies that showed measures at higher grade levels are highly correlated with longer sentence length and more elaborate syntactic structure, which are indicators of more advanced syntactic embedding (Kim & Lim, 2019). It appears that ESL students find it easier to adopt these conventions of the written register, whereas EFL writers demonstrate a lesser degree of syntactic and lexical elaboration, which could be related to the limits of classroom-based instruction and L1-based formulation practices. The metric patterns all together suggest the importance of structural sophistication for the development of a mature academic voice.

Conclusion

The results show that Chinese EFL and Malaysian ESL undergraduate writers differed significantly in most, but not all, of the ten Coh-Metrix indices examined in this study. Malaysian ESL writers produced longer texts, longer sentences, lower syntactic simplicity scores, stronger noun overlap, higher adjacent-sentence and intra-paragraph semantic similarity, and higher readability levels. In contrast, Chinese EFL writers showed significantly higher lexical diversity, while no significant differences were found in word concreteness or logical connective density. These findings indicate that EFL–ESL differences are not reducible to a general ESL advantage; rather, they reflect distinct profiles of linguistic resource allocation shaped by different socio-educational ecologies.

The novelty of this study lies in its direct, task-controlled comparison of EFL and ESL undergraduate argumentative writing through a multidimensional Coh-Metrix framework. Unlike studies that compare L2 writers mainly with native-speaker baselines or examine learners within a single context, this study foregrounds variation within L2 writing itself. Its core contribution is to show that learning context is reflected not only in surface-level linguistic complexity but also in deeper patterns of information packaging, referential continuity, and semantic coherence. In particular, the findings suggest that ESL writers' stronger discourse continuity may be associated with broader exposure to English across academic and social domains, whereas EFL writers' higher lexical diversity may partly reflect shorter texts and less sustained conceptual development.

The study therefore contributes to L2 writing research, language education, and assessment by arguing for profile-based rather than single-index interpretations of learner writing. Pedagogically, the findings suggest that EFL writing instruction should move beyond formulaic transitions and encourage sustained referential chains, semantic recurrence, and conceptual development across sentences and paragraphs. From an assessment perspective, the study highlights the importance of multidimensional evaluation that recognises both the strengths and constraints of writers from different learning ecologies. Such an approach can support fairer and more context-sensitive interpretations of undergraduate L2 argumentative writing.

References

- Abdi Tabari, M., & Lee, H. (2026). Examining the effects of task type, task order, and their interaction on L2 writing performance: Insights from independent and integrated tasks. *International Journal of Applied Linguistics*, 36(2), 1092–1108. <https://doi.org/10.1111/ijal.12837>
- Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. *Automatic Treatment and Analysis of Learner Corpus Data*, 59–72. <https://www.scopus.com/pages/publications/85105737611?origin=resultslist>
- Aljohani, S. (2026). Topic-specific vs Global lexical bundles and discourse-semantic text properties in L2 argumentative writing. *Australian Journal of Applied Linguistics*, 9, Article 103637. <https://doi.org/10.29140/ajal.2026.103637>
- Appel, R., & Szeib, A. (2018). Linking adverbials in L2 English academic writing: L1-related differences. *System*, 78, 115–129. <https://doi.org/10.1016/j.system.2018.08.008>
- Aryadoust, V. (2016). Understanding the growth of ESL paragraph writing skills and its relationships with linguistic features. *Educational Psychology*, 36(10), 1742–1770. <https://doi.org/10.1080/01443410.2014.950946>
- Becker, A., & Nekrasova-Beker, T. (2026). What's in a profile? Investigating linguistic features of L2 writing at advanced CEFR levels. *Journal of Second Language Writing*, 71, Article 101284. <https://doi.org/10.1016/j.jslw.2026.101284>
- Berei, E. B. (2025). Socio-educational resources for academic writing—Open-access, digital data for social work programs in Romanian universities. *Trends in Higher Education*, 4(3), Article 38. <https://doi.org/10.3390/higheredu4030038>
- Chen, L., & Luo, Q. (2025). A comparative study of text characteristics of CET-6, IELTS, and TOEFL reading passages based on computational tools. *Journal of English for Academic Purposes*, 77, Article 101556. <https://doi.org/10.1016/j.jeap.2025.101556>
- Chen, Q., Hao, C., & Xiao, Y. (2020). When testing stakes are no longer high: Impact on the Chinese College English learners and their learning. *Language Testing in Asia*, 10, Article 6. <https://doi.org/10.1186/s40468-020-00102-5>
- Choi, S., Park, S., & Tak, J.-Y. (2025). Exploring vocabulary patterns in ESL and EFL learner corpora: Implications for language teaching. *Korean Journal of English Language and Linguistics*, 25, 1378–1394. <https://doi.org/10.15738/kjell.25..202510.1378>
- Chung, H., & Jeon, M. (2026). A Coh-Metrix analysis of argumentative essays by Hong Kong ESL, Chinese EFL, and native writers. *English Language & Literature Teaching*, 32(1), 43–71. <https://doi.org/10.35828/etak.2026.32.1.43>
- Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42(3), 475–493. <https://doi.org/10.1002/j.1545-7249.2008.tb00142.x>

- Crossley, S. A., & Kyle, K. (2018). Assessing writing with the tool for the automatic analysis of lexical sophistication (TAALES). *Assessing Writing*, 38, 46–50. <https://doi.org/10.1016/j.asw.2018.06.004>
- Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life-Long Learning*, 21(2–3), 170–191. <https://doi.org/10.1504/IJCEELL.2011.040197>
- Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, 26, 66–79. <https://doi.org/10.1016/j.jslw.2014.09.006>
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2012). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, 29(2), 243–263. <https://doi.org/10.1177/0265532211419331>
- Cumming, A. (2001). Learning to write in a second language: Two decades of research. *IJES: International Journal of English Studies*, 1(2), 1–23. <https://revistas.um.es/ijes/article/view/48331>
- Fan, K., & Gao, T. (2026). The role of AI-mediated metacognitive interventions in EFL writing development: A cross-sectional study of textual accuracy, lexical sophistication, and cohesion. *Acta Psychologica*, 267, Article 107069. <https://doi.org/10.1016/j.actpsy.2026.107069>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- Francis, W. S., Romo, L. F., & Gelman, R. (2002). Syntactic structure, grammatical accuracy, and content in second-language writing: An analysis of skill learning and on-line processing. *Advances in Psychology*, 134, 317–337. [https://doi.org/10.1016/S0166-4115\(02\)80017-6](https://doi.org/10.1016/S0166-4115(02)80017-6)
- Gill, S. K. (2002). *International communication: English language challenges for Malaysia*. Universiti Putra Malaysia Press.
- Gilquin, G., & Granger, S. (2015). Learner language. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of corpus linguistics* (pp. 410–429). Cambridge University Press.
- Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3(2), 371–398. <https://doi.org/10.1111/j.1756-8765.2011.01121.x>
- Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal*, 115(2), 210–229. <https://doi.org/10.1086/678222>
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234. <https://doi.org/10.3102/0013189X11413260>
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193–202. <https://doi.org/10.3758/BF03195564>
- Granados, A., López-Jiménez, M. D., & Lorenzo, F. (2022). A longitudinal study of L2 historical writing. *Ibérica*, 43, 129–154. <https://doi.org/10.17398/2340-2784.43.129>
- Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, 1(1), 7–24. <https://doi.org/10.1075/ijlcr.1.1.01gra>
- Ha, M. J. (2023). Computer-aided analysis of syntactic elaboration of written argumentation across topics and L1s. *Computer Assisted Language Learning Electronic Journal*, 24(1),

242–268.

<https://www.scopus.com/pages/publications/85149128725?origin=resultslist>

Halim, H. B. A., & Wu, L. (2025). L2 motivational self system and learning emotions as predictors of syntactic complexity. *Acta Psychologica*, 261, Article 105699. <https://doi.org/10.1016/j.actpsy.2025.105699>

Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press.

Janebi Enayat, M. (2025). Computationally derived linguistic features of L2 narrative essays and their relations to human-judged writing quality. *Language Testing in Asia*, 15(1), Article 35. <https://doi.org/10.1186/s40468-025-00374-9>

Jarvis, S., & 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 *Approaching language transfer through text classification: Explorations in the detection-based approach* (pp. 106–126). Channel View Publications.

Kargar, A. (2025). Lexical diversity and sophistication in professional architectural discourse: A computational comparison of native and non-native English writing. *Hermes (Denmark)*, 65, 81–98. <https://doi.org/10.7146/hjlc.vi65.152533>

Kim, J., & Lim, K. (2019). A comparative analysis of Koreans' English writings and Google translations using Coh-Metrix 3.0. *Korean Journal of English Language and Linguistics*, 19(3), 452–474. <https://doi.org/10.15738/kjell.19.3.201909.452>

Kiselnikov, A., Vakhitova, D., & Kazymova, T. (2020). Coh-metrix readability formulas for an academic text analysis. *IOP Conference Series: Materials Science and Engineering*, 890(1), Article 012207. <https://doi.org/10.1088/1757-899X/890/1/012207>

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36–62. <https://doi.org/10.5054/tq.2011.240859>

Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, 34(4), 493–511. <https://doi.org/10.1177/0265532217710675>

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16–27. <https://doi.org/10.1016/j.jslw.2015.06.003>

Mahadini, M. K., Setyaningsih, E., & Sarosa, T. (2021). Using conventional rubric and Coh-Metrix to assess EFL students' essays. *International Journal of Language Education*, 5(4), 260–270. <https://doi.org/10.26858/ijole.v5i4.19105>

Manchón, R. M. (Ed.). (2011). *Learning-to-write and writing-to-learn in an additional language* (Vol. 31). John Benjamins Publishing.

Mao, F., Silver, R. E., & Luo, Q. (2025). The evolution of language education policies and international relations in China: A structural realist perspective. *Language Problems and Language Planning*, 49(2), 178–205. <https://doi.org/10.1075/lplp.24052.mao>

Matthews, J., & Wijeyewardene, I. (2018). Exploring relationships between automated and human evaluations of L2 texts. *Language Learning and Technology*, 22(3), 143–158. <https://doi.org/10.125/44661>

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57–86. <https://doi.org/10.1177/0741088309351547>

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

- McNamara, D. S., Louwerse, M. M., & Graesser, A. C. (2002). *Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension*. Technical Report, University of Memphis.
- Murphy, L., & Roca de Larios, J. (2010). Searching for words: One strategic use of the mother tongue by advanced Spanish EFL writers. *Journal of Second Language Writing*, 19(2), 61–81. <https://doi.org/10.1016/j.jslw.2010.02.001>
- Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, 47, Article 100511. <https://doi.org/10.1016/j.asw.2020.100511>
- Nkhobo, T., & Chaka, C. (2023). Syntactic pattern density, connectives, text easability, and text readability indices in students' written essays: A Coh-Metrix analysis. *Research Papers in Language Teaching and Learning*, 13(1), 121–136. <http://rpltl.eap.gr>
- Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, 19(4), 218–233. <https://doi.org/10.1016/j.jslw.2010.10.003>
- Peng, Y., Sun, J., Quan, J., Wang, Y., Lv, C., & Zhang, H. (2023). Predicting Chinese EFL learners' human-rated writing quality in argumentative writing through multidimensional computational indices of lexical complexity. *Assessing Writing*, 56, Article 100722. <https://doi.org/10.1016/j.asw.2023.100722>
- Polio, C., & Yoon, H.-J. (2018). The reliability and validity of automated tools for examining variation in syntactic complexity across genres. *International Journal of Applied Linguistics*, 28(1), 165–188. <https://doi.org/10.1111/ijal.12200>
- Pu, L., Heng, R., & Xu, B. (2023). Language development for English-medium instruction: A longitudinal perspective on the use of cohesive devices by Chinese English majors in argumentative writing. *Sustainability*, 15(1), Article 17. <https://doi.org/10.3390/su15010017>
- Qin, W., & Uccelli, P. (2016). Same language, different functions: A cross-genre analysis of Chinese EFL learners' writing performance. *Journal of Second Language Writing*, 33, 3–17. <https://doi.org/10.1016/j.jslw.2016.06.001>
- Rioux, E. J., & Thordardottir, E. (2026). How multilingual adolescents use connectives in argumentative texts: Insights into language proficiency and writing quality. *Reading and Writing*. Advance online publication. <https://doi.org/10.1007/s11145-026-10832-2>
- Ruiz-Funes, M. (2015). Exploring the potential of second/foreign language writing for language learning: The effects of task factors and learner variables. *Journal of Second Language Writing*, 28, 1–19. <https://doi.org/10.1016/j.jslw.2015.02.001>
- Shen, C., Guo, J., Shi, P., Qu, S., & Tian, J. (2023). A corpus-based comparison of syntactic complexity in academic writing of L1 and L2 English students across years and disciplines. *PLoS ONE*, 18(10), e0292688. <https://doi.org/10.1371/journal.pone.0292688>
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27(4), 657–677. <https://doi.org/10.2307/3587400>
- Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing*, 32, 17–35. <https://doi.org/10.1016/j.jslw.2016.02.002>
- Thirusanku, J., & Yunus, M. M. (2014). Status of English in Malaysia. *Asian Social Science*, 10(14), 254–260. <https://doi.org/10.5539/ass.v10n14p254>

- Wang, Y., & Xie, Q. (2022). Diagnosing EFL undergraduates' discourse competence in academic writing. *Assessing Writing*, 53, Article 100641. <https://doi.org/10.1016/j.asw.2022.100641>
- Xie, Y. (2026). Discourse cohesion and coherence and the writing quality of English argumentative essays: An analysis based on Coh-Metrix. *Science Insights Education Frontiers*, 32(2), 5249–5264. <https://eric.ed.gov/?id=EJ1508222>
- Xue, J., Zheng, L., Tang, X., Li, B., & Geva, E. (2021). Human ratings of writing quality capture features of syntactic variety and transformation in Chinese EFL learners' argumentative writing. *Frontiers in Psychology*, 12, Article 660796. <https://doi.org/10.3389/fpsyg.2021.660796>
- Yi, B., & Ni, C. (2015). Planning and working memory effects on L2 performance in Chinese EFL learners' argumentative writing. *Indonesian Journal of Applied Linguistics*, 5(1), 43–51.
- Yoon, H.-J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, 51(2), 275–301. <https://doi.org/10.1002/tesq.296>
- Zhang, C., & Li, R. (2025). A study on linguistic features of Chinese ESL and EFL learners' writing: A learning context perspective. *SAGE Open*, 15(3). <https://doi.org/10.1177/21582440251378082>
- Zou, S., Yan, X., & Fan, J. (2024). Establishing analytic score profiles for large-scale L2 writing assessment: The case of the CET-4 writing test. *Assessing Writing*, 60, Article 100826. <https://doi.org/10.1016/j.asw.2024.100826>