TEXTUAL COMPLEXITY IN VIETNAM'S POST-REFORM ENGLISH EXAMS: A 2025–2026 CORPUS COMPARISON

Nguyen Quoc Khanh[1]; Vo Nhat Quang[2]✉; To Thi Thuy Loan[3]

[1]The University of Da Nang, Da Nang, Vietnam; [2]Ho Chi Minh University of Banking, Ho Chi Minh City, Vietnam; [3]University of Social Sciences and Humanities, Vietnam National University HCMC, Ho Chi Minh City, Vietnam

✉ 020401250009@st.buh.edu.vn

https://doi.org/10.63506/jilc.1002.454

(Received: 25/06/2026; Revised: 28/07/2026; Accepted: 18/08/2026)

Abstract: This study examines textual complexity in the 2025 and 2026 Vietnamese National High School Graduation English exams. Using a small-scale corpus-based descriptive case comparison, the study analyzed one official paper from each year and section-level sub-corpora, including reading passages, cloze texts, item stems, answer options, and sentence-arrangement fragments. Four automated corpus-analysis tools were used to measure lexical diversity, lexical sophistication, syntactic complexity, and cohesion. Results show no uniform increase or decrease in complexity. The 2026 paper was shorter and slightly less lexically diverse overall, but section-level profiles varied. Reading passages became less lexically diverse yet more syntactically complex. Because the study is limited to two official textual cases and several short sub-corpora, findings are interpreted as descriptive evidence, not proof of test difficulty or year-level effects.

Key words: Linguistic complexity; corpus analysis; English examination; Vietnam; cohesion

SO SÁNH ĐỘ PHỨC TẠP VĂN BẢN TRONG ĐỀ THI TIẾNG ANH TỐT NGHIỆP THPT NĂM 2025–2026

Tóm tắt: Nghiên cứu này phân tích độ phức tạp văn bản trong đề thi tiếng Anh tốt nghiệp THPT của Việt Nam năm 2025 và 2026. Với thiết kế so sánh mô tả dựa trên ngữ liệu, nghiên cứu phân tích một đề chính thức mỗi năm và các tiểu ngữ liệu gồm bài đọc hiểu, văn bản điền khuyết, câu dẫn, phương án trả lời và các thành phần câu trong bài tập sắp xếp câu. Bốn công cụ phân tích ngữ liệu tự động được dùng để đo độ đa dạng từ vựng, độ tinh vi của từ vựng, độ phức tạp cú pháp và tính liên kết văn bản. Kết quả không cho thấy xu hướng tăng hoặc giảm thống nhất về độ phức tạp văn bản. Đề năm 2026 ngắn hơn và có độ đa dạng từ vựng thấp hơn một chút khi xét toàn đề, nhưng kết quả phân tích ở từng phần có sự khác biệt. Các bài đọc hiểu có độ đa dạng từ vựng thấp hơn nhưng độ phức tạp cú pháp cao hơn. Do chỉ dựa trên hai đề thi và một số tiểu ngữ liệu ngắn, các kết quả chỉ mang tính mô tả, không đủ cơ sở để kết luận về độ khó của đề thi hoặc xu hướng thay đổi giữa các năm.

Từ khóa: Độ phức tạp ngôn ngữ; phân tích ngữ liệu; đề thi tiếng Anh; Việt Nam; liên kết văn bản

1. Introduction

High-stakes English examinations play an important role in shaping language teaching, learning, and assessment practices in Vietnam (Alderson & Wall, 1993; Tong & Pham, 2024). The English component of the Vietnamese National High School Graduation Examination is particularly significant because it is connected to graduation recognition, the evaluation of teaching and learning quality, and university and college admission purposes (Ministry of Education and Training [MOET], 2024). Since 2025, the examination has entered a new phase linked to the 2018 General Education Curriculum and competence-oriented assessment (MOET, 2023, 2024), under which students completing general education are expected to reach Level 3 of Vietnam's six-level language proficiency framework (MOET, 2018). This reform context is important because English assessment in Vietnam has long been associated with debates about grammar-focused testing, reading comprehension, multiple-choice formats, and curriculum-assessment alignment (Nguyen & Dao, 2025). Following the 2025 English examination, public discussion focused on text length, advanced vocabulary, processing time, and suitability for the expected student level (Tuoi Tre News, 2025). However, such debates often rely on general impressions, highlighting the need for systematic linguistic evidence to examine how textual demands vary across different components of the examination.

Existing research has examined the Vietnamese high school graduation English exam in relation to washback, curriculum-assessment alignment, lexical demand, and readability (Cao et al., 2026; Nguyen & Dao, 2025; Tong & Pham, 2024). These studies provide useful evidence, but three gaps remain. First, existing analyses have tended to focus on reading passages, while the English paper also includes cloze texts, item stems, answer options, and sentence arrangement fragments. Second, previous work has often relied on readability formulas or lexical profiling, although linguistic complexity may involve distinct lexical, syntactic, and discourse-level components (Bulté & Housen, 2012). Third, there is still limited section-level evidence comparing the 2025 and 2026 papers, the first two years in which the post-reform format was implemented. To address these gaps, this study compares the two papers across lexical diversity, lexical sophistication, syntactic complexity, and cohesion using automated corpus-analysis tools.

A conceptual distinction is made between textual complexity and test difficulty. Textual complexity refers here to measurable linguistic properties of the exam materials, including lexical, syntactic, and cohesion-related features (Bulté & Housen, 2012; Rafatbakhsh & Ahmadi, 2023). Test difficulty, by contrast, may reflect interactions among textual features, task or item characteristics, and test-taker characteristics (Bachman & Palmer, 1996; Rafatbakhsh & Ahmadi, 2023). Because the present study does not analyze student responses or item statistics, it does not determine whether one paper was psychometrically easier or harder. Instead, it treats a national English examination paper as a set of task components and describes how textual complexity varies across reading passages, cloze texts, item stems, answer options, and sentence arrangement fragments. The study seeks to contribute to the understanding of textual complexity in high-stakes language assessment by distinguishing it from test difficulty, applying automated corpus-analysis tools to different examination components, and providing descriptive evidence from the first two years of the post-reform format.

The study addresses the following research questions:

RQ1. How do the two analyzed official 2025 and 2026 English exam papers differ in lexical diversity, lexical sophistication, syntactic complexity, and cohesion?

RQ2. How are lexical, syntactic, and cohesion-related features distributed across the main task components of the two analyzed papers?

2. Literature Review

2.1 High-stakes English testing and corpus-based exam analysis

Previous studies have shown that high-stakes English exams can affect textbook coverage, classroom time, materials, activities, and assessment practices (Tong & Pham, 2024). Nguyen and Dao (2025) examined the alignment of the 2025 English examination with the national curriculum, analyzed its test items using Bloom's taxonomy, and investigated its washback effects on students' learning. Corpus-based methods can complement this work because they allow researchers to examine measurable textual features in authentic exam materials rather than relying on impressionistic judgments (Biber et al., 1998; McEnery & Hardie, 2011).

In Vietnam, existing corpus-based exam research has examined lexical demand, vocabulary coverage, and readability. Vu (2019) analyzed high-stakes English exam papers from 2002 to 2018, Nguyen et al. (2023) examined lexical coverage and AWL occurrence from 2019 to 2022, and Cao et al. (2026) analyzed reading passages from 2015 to 2025. Beyond Vietnam, Rafatbakhsh and Ahmadi (2023) showed that English as a foreign language (EFL) reading test difficulty may be influenced by text features, item features, and their interaction, while Nguyen and Le (2024) compared International English Language Testing System (IELTS) Academic Reading tests and practice tests through lexical, syntactic, and discourse-level indices. These studies provide a foundation for examining linguistic characteristics of examination materials across multiple dimensions.

2.2 Multidimensional linguistic complexity

Textual linguistic complexity is increasingly understood as a multidimensional construct. It cannot be reduced to word count, sentence length, or readability score alone. For second-language learners, linguistic demands may arise from multiple linguistic features, including lexical diversity, lexical sophistication, syntactic complexity, and cohesion (Bulté & Housen, 2012; Rafatbakhsh & Ahmadi, 2023).

Lexical complexity generally includes lexical density, lexical diversity, and lexical sophistication (Read, 2000). Lexical diversity concerns the range of words used, while lexical sophistication concerns the extent to which a text contains less frequent, later-acquired, more academic, or less familiar words (Kyle et al., 2018, 2021). Syntactic complexity refers to how information is organized through sentence structures, clauses, phrases, and nominal groups. Students may know individual words but still misunderstand grammatical relationships between ideas, so syntactic complexity should be analyzed separately from lexical complexity (Kyle & Crossley, 2018; Lu, 2010).

Cohesion refers to linguistic features that connect ideas across sentences and larger textual units, including lexical repetition, overlap, and connectives (Halliday & Hasan, 1976). Automated

cohesion research similarly distinguishes local, global, and text-level cohesion indices (Crossley et al., 2016). A multidimensional view is therefore useful for exam analysis because different task components may create different types of linguistic demand.

2.3 Theoretical framework

The present study is grounded in a multidimensional view of textual linguistic complexity in language assessment. Rather than treating textual complexity as a single property, the framework considers lexical, syntactic, and discourse-level features across different task components (Bulté & Housen, 2012).

Drawing on this perspective, this study operationalizes four dimensions of textual complexity using automated corpus-analysis tools. Lexical diversity is operationalized with the Tool for the Automatic Analysis of Lexical Diversity (TAALED; Kyle et al., 2021), lexical sophistication with the Tool for the Automatic Analysis of Lexical Sophistication (TAALES; Kyle et al., 2018), syntactic complexity with the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC; Kyle, 2016), and cohesion with the Tool for the Automatic Analysis of Text Cohesion (TAACO; Crossley et al., 2016). These tools operationalize selected observable dimensions of textual complexity rather than the full construct of test difficulty.

3. Methodology

3.1 Research design and data sources

This study adopted a small-scale corpus-based comparative design to examine the textual linguistic complexity of the 2025 and 2026 Vietnamese National High School Graduation English exams. A corpus-based design was appropriate because it allowed the linguistic features of examination texts to be examined systematically rather than through intuition (Biber et al., 1998; McEnery & Hardie, 2011). The purpose was not to evaluate students' test performance or to make statistical generalizations across multiple exam forms, but to provide a descriptive and multidimensional linguistic profile of two official post-reform English exam papers.

The data consisted of one official English exam paper from each year, specifically exam code 1101 for both 2025 and 2026. Because the study did not independently analyze all parallel test codes within each year, the findings are not interpreted as representing all possible code-level variations. Cross-year comparisons are therefore limited to the official textual forms analyzed in this study, and comparisons involving answer options are interpreted with particular caution because answer-option content may vary across codes.

The two papers were selected because they represent the first two operational years of the examination format aligned with the 2018 General Education Curriculum. MOET also published official answer-key materials for the 2025 and 2026 examinations, which helped verify the official status of the materials used in this study (MOET, 2025, 2026). Only English textual content was included; Vietnamese instructions, administrative information, page numbers, question numbers, answer labels, and layout-related elements were removed during corpus cleaning.

3.2 Corpus construction and analytical tools

Each exam paper was cleaned and converted into plain-text format to ensure machine readability and consistent preprocessing (McEnery & Hardie, 2011). To support both whole-paper and section-level analysis, the cleaned data were divided into several analytic corpora: the full operational corpus, the stimuli-only corpus, the reading-passage corpus, the cloze-text corpus, the item-stem corpus, the answer-option corpus, and the sentence-arrangement corpus. Reading passages and cloze texts were treated as continuous stimuli; item stems comprised the English wording of questions or prompts; answer options comprised the English alternatives attached to each item; and sentence-arrangement fragments were treated as fragmentary material. Sentence-arrangement fragments (Questions 14–18 in 2025 and 36–40 in 2026) were retained within the lexical item-stem corpus as a nested subset and were not counted again in the full-operational totals. Administrative question markers, answer-option labels, arrangement-fragment labels, and processing headings were excluded consistently from the relevant analytic inputs. Reading passages were retained within the stimulus component rather than duplicated in the answer-option layer. For syntactic analysis, the item-stem inputs excluded the five sentence-arrangement tasks in each year, leaving 18 non-arrangement stems per paper (180 and 190 TAALED tokens in 2025 and 2026, respectively).

Four automated corpus tools were used. TAALED 1.4.1 analyzed lexical diversity through the moving-average type-token ratio with a 50-word window (MATTR-50), Hypergeometric Distribution D (HD-D), and the Measure of Textual Lexical Diversity (MTLD). These indices were selected because they are less dependent on text length than simple type-token ratio (Covington & McFall, 2010; Kyle et al., 2021; McCarthy & Jarvis, 2010; Zenker & Kyle, 2021). Following Zenker and Kyle's (2021) caution about minimum text length and index stability, MATTR-50 and HD-D were treated as the primary lexical diversity indicators, while large MTLD contrasts in short sub-corpora were treated as exploratory descriptive signals. TAALES 2.2 analyzed lexical sophistication through British National Corpus (BNC) written content-word frequency, age of acquisition (AoA), and Academic Word List (AWL) density (Coxhead, 2000; Kuperman et al., 2012; Leech et al., 2001). TAASSC 1.3.8 analyzed syntactic complexity through mean length of sentence (MLS), clauses per T-unit (C/T), dependent clauses per clause (DC/C), and complex nominals per T-unit (CN/T) (Kyle, 2016; Kyle & Crossley, 2018; Lu, 2010). TAACO 2.0 analyzed cohesion through adjacent content-word overlap, repeated content lemmas, and connective incidence (Crossley et al., 2016).

These indices were selected to represent complementary aspects of each construct: frequency, acquisition-related familiarity, and academic-word orientation for lexical sophistication; sentence length, clause density, subordination, and nominal complexity for syntactic complexity; and lexical overlap, repetition, and connective use for cohesion.

For reproducibility, TAALED 1.4.1 provided the outputs basic_ntokens, basic_ncontent_tokens, lexical_density_tokens, mattr50_aw, hdd42_aw, and mtld_original_aw; the reported lexical-diversity indices were based on all words. TAALES 2.2 provided BNC_Written_Freq_CW_Log, Kuperman_AoA_CW, and All_AWL_Normed using the corresponding frequency, psycholinguistic, and academic-word-list index categories. Individual Item Output was unselected in both tools. The last measure was used directly as the normalized proportion of all words belonging to the AWL. TAASSC was run once with Clause Complexity,

Phrase Complexity, and Syntactic Sophistication selected and once with L2SCA Variables selected; the reported outputs were MLS, C_T, DC_C, and CN_T. TAACO was run without a source text or semantic-overlap model; all lemma-token categories, Sentence, Adjacent, and Adjacent 2 lexical-overlap options, TTR, Connectives, and Givenness were enabled, and the reported outputs were adjacent_overlap_cw_sent, repeated_content_lemmas, and all_connective.

3.3 Data analysis and interpretation

After the analytic files described in Section 3.2 were constructed and manually audited, they were processed through TAALED, TAALES, TAASSC, and TAACO. The comma-separated values (CSV) outputs were merged into a master dataset using file name, year, and section as identifiers. Descriptive comparisons were conducted between matched 2025 and 2026 sections using the same index, first at the whole-corpus or stimuli level and then across corresponding task components.

To reduce segmentation error, the corpus files were checked through a manual audit before automated processing. The audit focused on whether section boundaries were correctly identified, whether non-English or administrative elements had been removed consistently, whether answer options and arrangement fragments had been separated from continuous stimuli, and whether cloze blanks had been handled consistently. Uncertain cases were resolved through author discussion before the final files were processed. Because the study did not use independent double coding of all segmentation decisions, no inter-rater reliability coefficient is reported. Files from 2025 and 2026 were never merged. For TAALED, TAALES, and TAASSC, multiple source texts belonging to the same analytic layer were retained in the same plain-text file and separated by blank lines. For the boundary-aware cohesion analysis, however, each continuous reading passage and cloze text was processed as a separate TAACO input file so that adjacent-sentence overlap was not calculated across unrelated text boundaries. Section-level adjacent content-word overlap was aggregated across within-text adjacent sentence pairs only, while repeated-content-lemma and connective measures were pooled using TAACO word counts.

Blanks in cloze texts were removed so that the remaining text represented the visible linguistic input available to test-takers. This decision may alter local syntactic structure and cohesive marking, especially when the missing item is a connective, preposition, or content word. The blank-removed indices therefore describe the post-deletion inputs rather than complete discourse. To assess the effect of blank removal on syntactic analysis, the correct responses were aligned with the question wording and answer-option sequences and reinserted into all 17 cloze positions in each paper. The blank-removed and answer-restored texts were analyzed using TAASSC 1.3.8 with identical L2SCA settings. The restored inputs represent reconstructed completed texts rather than the materials originally visible to test-takers.

No inferential statistical tests were conducted because the study examined two purposefully selected official exam papers rather than a large sample of independent test forms. The design does not support statistical generalization beyond the two papers analyzed (Brezina, 2018; Gries, 2015). The analysis therefore focused on descriptive contrasts.

To guide the interpretation of descriptive contrasts, the study applied an a priori interpretation rule. A section-level contrast was treated as a main descriptive finding only when it showed at least two converging signals within one dimension or across theoretically related

dimensions. Isolated single-index changes with no corroboration were not treated as main findings. MTLD shifts in short corpora, AWL changes in small sections, and cohesion values for non-continuous components were reported only as secondary observations. Cohesion indices were interpreted directly only for continuous texts, namely reading passages and cloze texts.

Several methodological limitations should be noted. The findings are descriptive, automated corpus tools do not replace expert judgment, the segmentation audit did not provide an inter-rater reliability coefficient, syntactic complexity indices were not reported for fragmentary sections, and length-sensitive indices were treated cautiously (Kyle et al., 2021; McCarthy & Jarvis, 2010; Zenker & Kyle, 2021).

4. Results

This section presents the descriptive results of the multidimensional corpus-based analysis of the 2025 and 2026 Vietnamese National High School Graduation English exams. The results are organized into five areas: corpus composition, lexical diversity, lexical sophistication, syntactic complexity, and cohesion. Values are reported to three decimal places, but interpretation does not rely on third-decimal differences.

4.1 Corpus composition

Table 1 presents the corpus composition of the 2025 and 2026 exam papers across seven analytic corpora. Overall, the 2026 full operational corpus was shorter than the 2025 corpus, but this reduction was not evenly distributed across components. The stimuli-only, reading-passage, and cloze-text corpora increased in length, whereas answer options and arrangement fragments decreased.

Table 1

Corpus Composition by Year and Section

Section	Tokens 2025	Tokens 2026	Content tokens 2025	Content tokens 2026	Lexical density 2025	Lexical density 2026
Full operational corpus	2488	2442	1436	1314	0.577	0.538
Stimuli only	1021	1120	624	572	0.611	0.511
Reading passages	619	652	404	323	0.653	0.495
Cloze texts	402	468	220	249	0.547	0.532
Item stems	601	541	319	288	0.531	0.532
Answer options	866	781	495	456	0.572	0.584
Arrangement fragments	421	349	228	195	0.542	0.559

Note. Tokens refer to TAALED basic_ntokens from the cleaned English files. Content tokens are lexical words, and lexical density is the ratio of content tokens to total tokens. Stimuli-only includes reading passages and blank-removed cloze texts. The full-operational corpus comprises stimuli, lexical item stems, and answer options (2025: 1,021 + 601 + 866 = 2,488; 2026: 1,120 + 541 + 781 = 2,442). Arrangement fragments are nested within lexical item stems and are not counted again. Separately filtered TAASSC item-stem inputs are excluded from these totals.

4.2 Lexical diversity

Table 2 reports three lexical diversity indices: MATTR-50, HD-D, and MTLD. MATTR-50 and HD-D were treated as the primary indicators, while MTLD was interpreted cautiously for shorter sub-corpora, as explained in Section 3.2.

Table 2

Lexical Diversity by Year and Section

Section	MATTR-50 2025	MATTR-50 2026	HD-D 2025	HD-D 2026	MTLD 2025	MTLD 2026
Full operational corpus	0.789	0.782	0.881	0.862	77.431	73.792
Stimuli only	0.858	0.816	0.896	0.861	141.813	108.508
Reading passages	0.900	0.804	0.911	0.842	245.134	79.991
Cloze texts	0.788	0.830	0.853	0.873	77.363	131.445
Item stems	0.785	0.783	0.842	0.826	67.810	61.327
Answer options	0.701	0.718	0.855	0.858	45.297	36.547
Arrangement fragments	0.817	0.850	0.867	0.873	112.211	186.689

Note. MATTR-50 = moving-average type-token ratio with a 50-word window; HD-D = Hypergeometric Distribution D; MTLD = Measure of Textual Lexical Diversity. Higher values indicate greater lexical diversity. MATTR-50 and HD-D are primary indicators; MTLD is interpreted cautiously in short sub-corpora.

As presented in Table 2, the 2026 full operational corpus showed slightly lower lexical diversity across all three reported indices. Both primary lexical-diversity indices declined in the stimuli-only and reading-passage corpora; item-stem MATTR-50 was nearly unchanged, whereas HD-D and MTLD decreased. By contrast, answer options showed slightly higher MATTR-50 and HD-D in 2026 but lower MTLD, so no consistent lexical-diversity direction was inferred for this section. The reading passages showed the clearest decrease across the primary indicators. In contrast, the cloze texts and arrangement fragments showed higher lexical diversity in 2026. However, the large MTLD contrasts in reading passages and arrangement fragments should be interpreted cautiously because section-level corpora are short and MTLD can fluctuate strongly in limited text samples.

The whole-paper contrast may partly reflect changes in corpus composition, with longer stimuli but shorter item stems and answer options in 2026 (1,120 versus 1,021; 541 versus 601; and 781 versus 866 tokens, respectively). However, lower lexical-diversity values were also observed in the stimuli-only corpus, which excludes answer options (MATTR-50: 0.858 vs. 0.816; HD-D: 0.896 vs. 0.861; MTLD: 141.813 vs. 108.508). The whole-paper pattern therefore cannot be attributed to the shorter answer-option section alone.

At the individual-passage level, the two 2025 reading passages contained 286 and 333 tokens, with MATTR-50/HD-D values of 0.894/0.896 and 0.906/0.905, respectively. The two 2026 passages contained 302 and 350 tokens, with corresponding values of 0.864/0.867 and 0.758/0.803. Thus, both 2026 passages were lower values on both primary lexical-diversity measures than both 2025 passages, so the aggregate pattern was not driven by a single passage. However, with only two passages per year, this consistency does not rule out random variation or establish a general year-level effect. A file-level audit also confirmed that the reading-passage

corpora were processed using the same TAALED settings; the large MTLD contrast was therefore retained but treated as secondary to the converging MATTR-50 and HD-D pattern.

4.3 Lexical sophistication

Table 3 presents selected lexical sophistication indices from TAALES, including BNC written frequency, age of acquisition (AoA), and Academic Word List (AWL) density. TAALES provides indices related to word frequency, range, academic vocabulary, psycholinguistic properties, and n-gram information (Kyle et al., 2018). Higher BNC written frequency values indicate more frequent words, whereas higher AoA values indicate later-acquired words. The selected indices are based on established lexical and psycholinguistic resources, including the British National Corpus, age-of-acquisition ratings, and the Academic Word List (Coxhead, 2000; Kuperman et al., 2012; Leech et al., 2001).

Table 3

Lexical Sophistication by Year and Section

Section	BNC freq log 2025	BNC freq log 2026	AoA 2025	AoA 2026	AWL normed 2025	AWL normed 2026
Full operational corpus	-1.097	-0.956	7.216	6.961	0.093	0.105
Stimuli only	-1.207	-0.879	7.370	6.602	0.108	0.080
Reading passages	-1.330	-0.905	7.561	6.903	0.122	0.087
Cloze texts	-1.002	-0.846	7.045	6.205	0.087	0.069
Item stems	-0.974	-1.051	6.671	7.010	0.066	0.116
Answer options	-1.045	-0.994	7.363	7.400	0.093	0.132
Arrangement fragments	-0.933	-1.108	6.603	7.268	0.056	0.128

Note. BNC freq log = log-transformed British National Corpus written frequency; AoA = age of acquisition; AWL normed = normalized proportion of words belonging to the Academic Word List. Higher BNC values indicate more frequent vocabulary; higher AoA values indicate later-acquired vocabulary.

Table 3 shows mixed patterns in lexical sophistication. The stimuli-only corpus, reading passages, and cloze texts contained more frequent and earlier-acquired words in 2026. By contrast, item stems and arrangement fragments showed higher age-of-acquisition values, suggesting a shift toward later-acquired vocabulary in these components. AWL density is reported as supplementary information only. Although AWL values increased in the full operational corpus, item stems, answer options, and arrangement fragments, these changes are not treated as main findings because AWL is an academic-English list rather than a direct benchmark for Grade 12 communicative competence, and small section-level corpora may turn small raw word differences into visible normalized changes.

4.4 Syntactic complexity

Table 4 reports selected syntactic-complexity indices from TAASSC. TAASSC provides indices related to clausal complexity, phrasal complexity, and syntactic sophistication. Measures such as mean length of sentence, clauses per T-unit, dependent clauses per clause, and complex nominals per T-unit have been widely used in automated syntactic complexity analysis (Kyle & Crossley, 2018; Lu, 2010). The analysis was conducted only on sections containing sufficiently complete sentence-level texts. Answer options and sentence-arrangement fragments were

excluded from syntactic analysis; the item-stem row was calculated using separately filtered inputs containing only non-arrangement question stems.

Table 4

Syntactic Complexity by Year and Section

Section	MLS 2025	MLS 2026	C/T 2025	C/T 2026	DC/C 2025	DC/C 2026	CN/T 2025	CN/T 2026
Stimuli only (blanks removed)	19.473	16.985	1.534	1.680	0.337	0.437	2.345	1.973
Reading passages	18.629	20.875	1.463	1.838	0.283	0.500	2.049	2.324
Cloze texts (blanks removed)	20.950	13.528	1.706	1.526	0.448	0.362	3.059	1.632
Cloze texts (answers restored)	25.150	15.222	1.842	1.619	0.400	0.382	3.105	1.667
Item stems (filtered)	10.526	10.947	1.000	1.105	0.053	0.143	1.421	1.263

Note. MLS = mean length of sentence; C/T = clauses per T-unit; DC/C = dependent clauses per clause; CN/T = complex nominals per T-unit. The full operational corpus, answer options, and sentence-arrangement fragments were excluded from syntactic analysis. Filtered item-stem inputs comprised 18 non-arrangement stems per paper (TAALED basic_ntokens: 180 in 2025 and 190 in 2026). The stimuli-only row includes cloze texts with blanks removed. Paired cloze inputs were analyzed before and after restoring the correct answers (TAASSC nwords: 419/503 for 2025 and 487/548 for 2026). The restored-answer results describe reconstructed completed texts rather than direct measures of test difficulty.

In Table 4, the reading passages showed higher values across all four syntactic-complexity indices in 2026. The stimuli-only corpus showed shorter mean sentence length but higher clause-related indices. Cloze texts showed lower values in 2026 across all four indices under both blank-removed and answer-restored conditions, although the restored inputs represent completed-text sensitivity checks. In the separately filtered non-arrangement item stems, MLS, C/T, and DC/C increased in 2026, whereas CN/T decreased.

4.5 Cohesion

Table 5 presents selected cohesion indices from TAACO, including adjacent content-word overlap, repeated content lemmas, and connective incidence.

Table 5

Cohesion Indices by Year and Section

Section	Adjacent CW overlap 2025	Adjacent CW overlap 2026	Repeated content lemmas 2025	Repeated content lemmas 2026	Connectives 2025	Connectives 2026
Stimuli only	0.041	0.063	0.174	0.189	0.075	0.046
Reading passages	0.041	0.092	0.188	0.207	0.070	0.050
Cloze texts	0.043	0.036	0.152	0.165	0.082	0.040

Note. Adjacent CW overlap = adjacent content-word overlap. Values are boundary-aware aggregates from separately processed continuous texts; no adjacent-sentence comparisons were

made across unrelated text boundaries. Repeated content lemmas and connectives were pooled using TAACO word counts.

After boundary-aware aggregation, adjacent content-word overlap increased in the stimuli-only corpus and reading passages in 2026 but decreased slightly in cloze texts. Repeated content lemmas increased across all three reported sections, whereas connective incidence decreased. Cohesion is therefore interpreted only for continuous stimulus texts.

4.6 Section-level synthesis

Across the five result tables, only patterns supported by converging evidence are treated as main descriptive findings. The strongest pattern appears in the reading passages. In 2026, this section showed lower lexical diversity across the primary lexical diversity indices and more frequent, earlier-acquired vocabulary, but higher syntactic complexity across all reported syntactic indices. This provides the clearest evidence of a changed section-level textual profile. The passage-level check showed the same descriptive direction for both individual reading passages, although the small number of passages does not support a general year-level inference.

The cloze texts showed a different but more cautious pattern. They were longer and more lexically diverse in 2026, while their syntactic indices remained lower across both blank-removed and answer-restored inputs. This sensitivity analysis supports the stability of the descriptive contrast, but the reconstructed completed texts should not be equated with the examination materials originally displayed to test-takers.

Item stems, answer options, and arrangement fragments are interpreted more cautiously. The lexical item-stem layer contained fewer tokens in 2026, whereas separately filtered syntactic indices showed mixed changes; AWL changes are treated as secondary observations. Overall, the results indicate uneven section-level textual profiles across the two official exam cases, not a general redistribution of difficulty and not a year-level effect.

5. Discussion

The present study examined the textual linguistic complexity of the 2025 and 2026 Vietnamese National High School Graduation English exams through a multidimensional corpus-based framework. As shown in Tables 1–5, the two papers differed across lexical diversity, lexical sophistication, syntactic complexity, and cohesion, but the differences did not move in one direction. Because the study is based on one official textual form per year and several short sub-corpora, the findings are interpreted as descriptive evidence rather than proof of test difficulty or year-level reform effects.

5.1 Uneven distribution of textual features

At the whole-paper level, the 2026 full-operational corpus contained slightly fewer tokens (2,442 versus 2,488) and showed modestly lower lexical diversity. However, the section-level results complicate this pattern. Some sections became shorter, while others became longer, and some showed lower lexical diversity while others showed higher lexical diversity or higher academic word density. This supports multidimensional accounts of linguistic complexity, which treat lexical, syntactic, and discourse features as complementary rather than interchangeable dimensions (Bulté & Housen, 2012). It also shows that an exam paper can usefully be analyzed as a collection of task components rather than as a homogeneous text. However, this study can

only describe textual properties and cannot show how strongly they affected test-takers' actual performance.

5.2 Lexical and syntactic shifts in reading and cloze sections

The reading passages showed a clear contrast between lexical and syntactic indices. In 2026, they showed lower lexical diversity, lower AWL density, more frequent words, and earlier-acquired vocabulary, but higher values across the reported syntactic complexity indices. This suggests a descriptive shift in the textual profile of the reading passages, from vocabulary-oriented complexity toward more structurally complex sentence organization. The passage-level check supports this descriptive interpretation: both 2026 reading passages showed lower MATTR-50 and HD-D values than both 2025 passages. Given the small number of passages, however, this remains a case-specific pattern rather than evidence of a general year-level shift.

This interpretation should remain cautious. The study does not prove that students experienced the 2026 reading passages as more syntactically difficult. It only shows that the analyzed text contained more syntactic features often associated with structural processing, such as clausal complexity, subordination, and complex nominals (Kyle & Crossley, 2018; Lu, 2010). The cloze texts showed higher lexical diversity but lower syntactic-index values in 2026, a contrast that persisted after the correct answers were restored. This sensitivity finding concerns reconstructed completed texts and does not establish differences in cloze-item difficulty.

5.3 Item stems, lexical sophistication, and cohesion as supplementary evidence

The lexical item-stem layer contained fewer tokens and somewhat later-acquired vocabulary in 2026, whereas the separately filtered non-arrangement stems showed higher MLS, C/T, and DC/C but lower CN/T. This matters because item wording is part of the textual input that test-takers process, and item-related features may interact with passage features (Rafatbakhsh & Ahmadi, 2023). However, these patterns should not be overstated. AWL values are retained only as a supplementary indicator of academic-word orientation because AWL was designed for academic English, not as a school-leaving or Level 3 benchmark.

The boundary-aware cohesion analysis showed higher adjacent content-word overlap in the 2026 reading passages but a slight decrease in the cloze texts. Repeated content lemmas increased in both sections, whereas connective incidence decreased. These indices are therefore treated as supplementary descriptions of cohesion within continuous stimulus texts rather than as whole-paper cohesion measures.

5.4 Cautious implications for teaching, exam preparation and test review

The findings offer cautious implications for classroom teaching, exam preparation, and test review. For classroom teaching, the results suggest that reading instruction should not focus only on difficult vocabulary or word lists. Teachers may also help learners unpack longer sentences, identify clause relationships, interpret complex noun phrases, and trace cohesive relations between ideas. This is especially relevant because the 2026 reading passages showed lower lexical diversity but higher syntactic complexity.

For exam preparation, practice could be organized by task component rather than treated as general reading practice. Reading passages, cloze texts, item stems, answer options, and

sentence arrangement fragments can display different textual profiles, so students may benefit from learning how each section is worded and structured. This implication remains provisional because the study did not analyze student responses or item-level performance.

For test review, future exam evaluation should monitor section-level textual profiles alongside curriculum alignment and item performance evidence. Corpus-based indices can help identify where textual complexity is located across an exam paper, but they cannot by themselves show whether the exam measures communicative competence, pragmatic competence, or real-life language use. They should therefore be used as one source of textual evidence within a broader assessment review process.

6. Conclusion

This study conducted a multidimensional corpus-based comparison of the 2025 and 2026 Vietnamese National High School Graduation English exams. Drawing on TAALED, TAALES, TAASSC, and TAACO, it examined lexical diversity, lexical sophistication, syntactic complexity, and cohesion across whole-paper and section-level corpora.

The findings show that the 2026 paper did not display a uniform increase or decrease in textual complexity. At the whole-paper level, the 2026 corpus was shorter and showed lower lexical diversity. At the section level, however, the pattern was uneven. Reading passages showed lower lexical diversity, but higher syntactic complexity. Cloze texts became longer and more lexically diverse, while lower syntactic-index values in 2026 persisted after correct answers were restored. The lexical item-stem layer contained fewer tokens, filtered item-stem syntax showed mixed changes, and boundary-aware cohesion patterns differed between the continuous stimulus sections.

The study contributes to corpus-based exam analysis by extending attention beyond reading passages to cloze texts, item stems, answer options, and sentence arrangement fragments. It also shows how multiple automated tools can construct a multidimensional profile of exam-text complexity. However, it analyzed only one official textual form from each of two exam years and focused on textual features rather than student performance. Future research should examine more years and combine corpus indices with item-level performance data, learner responses, expert judgment, or external reference corpora. Without an external benchmark corpus, the study cannot judge whether the values are appropriate for Level 3 learners. Overall, the study supports only a relative comparison of the two exam cases and keeps a clear boundary between textual complexity and test difficulty.

References

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115–129. https://doi.org/10.1093/applin/14.2.115

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press. https://doi.org/10.1017/CBO9780511804489

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press. https://doi.org/10.1017/9781316410899

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/lllt.32.02bul

Cao, T. H. P., Thai, L. C., & Nguyen, T. M. H. (2026). Analysis of the reading passage difficulty and lexical demands in the Vietnamese high school graduation English tests (2015–2025). *Tạp chí Khoa học và Công nghệ Đại học Đà Nẵng, 24*(1), 63–68. https://doi.org/10.31130/ud-jst.2026.24(1).681E

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics, 17*(2), 94–100. https://doi.org/10.1080/09296171003643098

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods, 48*(4), 1227–1237. https://doi.org/10.3758/s13428-015-0651-7

Gries, S. T. (2015). Statistics for learner corpus research. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 159–182). Cambridge University Press. https://doi.org/10.1017/CBO9781139649414.008

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* [Doctoral dissertation, Georgia State University]. ScholarWorks @ Georgia State University. https://scispace.com/pdf/measuring-syntactic-development-in-l2-writing-fine-grained-3fpsdv7759.pdf

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

Kyle, K., Crossley, S. A., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods, 50*, 1030–1046. https://doi.org/10.3758/s13428-017-0924-4

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly, 18*(2), 154–170. https://doi.org/10.1080/15434303.2020.1844205

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. Longman.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

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*, 381–392. https://doi.org/10.3758/BRM.42.2.381

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press. https://doi.org/10.1017/CBO9780511981395

Ministry of Education and Training (2018). *General education English language curriculum.* Ha Noi. https://www.britishcouncil.org/sites/default/files/english_language_grade_3_12_2018-eng_translation.pdf

Ministry of Education and Training (2023, December 29). *Cấu trúc định dạng đề thi tốt nghiệp THPT từ năm 2025* [Format structure of the high school graduation examination from 2025]. https://vqa.moet.gov.vn/vi/news/thong-bao/cau-truc-dinh-dang-de-thi-tot-nghiep-thpt-tu-nam-2025-74.html

Ministry of Education and Training (2024, December 24). *Ban hành Quy chế thi tốt nghiệp THPT từ năm 2025* [Issuing the regulations for the high school graduation examination from 2025]. https://vqa.moet.gov.vn/vi/news/tin-tuc-su-kien/ban-hanh-quy-che-thi-tot-nghiep-thpt-tu-nam-2025-167.html

Ministry of Education and Training (2025, July 7). *Công bố đáp án các môn thi của Kỳ thi tốt nghiệp THPT 2025* [Announcement of answer keys for the 2025 High School Graduation Examination]. https://vqa.moet.gov.vn/vi/news/tin-tuc-su-kien/cong-bo-dap-an-cac-mon-thi-cua-ky-thi-tot-nghiep-thpt-2025-208.html

Ministry of Education and Training (2026, June 19). *Công bố đáp án các môn thi của Kỳ thi tốt nghiệp THPT 2026* [Announcement of answer keys for the 2026 High School Graduation Examination]. https://moet.gov.vn/tin-tuc/tin-tong-hop2/cong-bo-dap-an-cac-mon-thi-cua-ky-thi-tot-nghiep-thpt-2026.html

Nguyen, C. L. T., & Dao, Q. T. N. (2025). Assessing the washback effects of the 2025 National High School Graduation Exam: A cross-sectional study in Vietnam. *Vietnam Journal of Education, 9*(3), 424–436. https://doi.org/10.52296/vje.2025.565

Nguyen, H. T. M., & Le, N. V. A. (2024). Text complexity of Cambridge-delivered IELTS Academic Reading tests: Comparability with IELTS Academic Reading practice tests from other publishers. *TESL-EJ, 28*(2). https://doi.org/10.55593/ej.28110a4

Nguyen, T. N., Do, H. V., Vu, D. H. P., & Do, T. P. A. (2023). A corpus-based investigation into lexical coverage and AWL occurrence in the English test of Vietnam's National High School Graduation Examination. *HNUE Journal of Science: Educational Sciences, 68*(4), 21–33. https://vjol.info.vn/DHSP-GD/en/article/view/113795/

Rafatbakhsh, E., & Ahmadi, A. (2023). Predicting the difficulty of EFL reading comprehension tests based on linguistic indices. *Asian-Pacific Journal of Second and Foreign Language Education, 8*, Article 41. https://doi.org/10.1186/s40862-023-00214-4

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press. https://doi.org/10.1017/CBO9780511732942

Tong, T. L. C., & Pham, T. H. N. (2024). EFL teaching practices under the washback of high-stakes tests: What aspects are affected? *Hue University Journal of Science: Social Sciences and Humanities, 133*(6B), 127–146. https://doi.org/10.26459/hueunijssh.v133i6D.7407

Tuoi Tre News (2025, June 29). *Fierce debate erupts over challenging English test in Vietnam's national high school graduation exam*. https://news.tuoitre.vn/fierce-debate-erupts-over-challenging-english-test-in-vietnams-national-high-school-graduation-exam-103250629110408848.htm

Vu, D. V. (2019, December). A corpus-based lexical analysis of Vietnam's high-stakes English exams [Conference presentation]. *20th English in Southeast Asia Conference, National Institute of Education, Nanyang Technological University, Singapore*. https://duyvv.wordpress.com/wp-content/uploads/2019/11/a-corpus-based-lexical-analysis-of-vietname28099s-high-stakes-english-exams.pdf

Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing, 47*, Article 100505. https://doi.org/10.1016/j.asw.2020.100505