CORPUS-INFORMED EVALUATION OF EFL TEXTBOOK–EXAMINATION ALIGNMENT IN VIETNAMESE UPPER-SECONDARY READING: EVIDENCE FROM TEXT COMPLEXITY INDICES

ĐÁNH GIÁ DỰA TRÊN NGỮ LIỆU VỀ SỰ TƯƠNG THÍCH GIỮA SÁCH GIÁO KHOA VÀ KỲ THI TRONG ĐỌC HIỂU TIẾNG ANH TRUNG HỌC PHỔ THÔNG Ở VIỆT NAM: MINH CHỨNG TỪ CÁC CHỈ SỐ ĐỘ PHỨC TẠP VĂN BẢN

Nhat Quang Vo[1], Quoc Khanh Nguyen[2]*, Thi Thuy Loan To[3]

[1]*Faculty of Foreign Languages, Ho Chi Minh University of Banking, Vietnam*
[2]*The University of Danang - Center for Continuing Education, Vietnam*
[3]*University of Social Sciences and Humanities, Vietnam National University Ho Chi Minh City, Vietnam*

*Corresponding author: av32.1b2cva1d047@oncce.udn.vn

(Received: August 01, 2026; Revised: September 07, 2026; Accepted: September 10, 2026)

DOI: 10.31130/ud-jst.2026.24(10).527E

Abstract - Vietnamese upper-secondary students use the Global Success English textbook sequence, while their end-of-school reading performance is assessed through the National High School Graduation Examination. This study examined the alignment between 72 textbook passages from Grades 10–12 and 14 examination passages from 2020–2026. A quantitative corpus-based descriptive-comparative design compared lexical diversity, lexical sophistication, syntactic complexity, cohesion, and readability. At the overall corpus level, lexical and readability measures showed no differences after Holm adjustment, although examination passages varied across periods. More consistent differences concerned cohesion and clause organization: examination passages contained fewer connectives and a higher density of dependent clauses. The findings indicate selective rather than uniform alignment and show how multidimensional corpus evidence can support materials evaluation, language assessment literacy, and EFL reading instruction in a high-stakes context.

Key words - Corpus linguistics; EFL reading; textbook–examination alignment; language assessment literacy; text complexity

Tóm tắt - Học sinh trung học phổ thông Việt Nam sử dụng bộ sách tiếng Anh Global Success, còn năng lực đọc cuối cấp được đánh giá qua Kỳ thi tốt nghiệp trung học phổ thông. Nghiên cứu xem xét mức độ tương thích giữa 72 bài đọc lớp 10–12 và 14 bài đọc đề thi giai đoạn 2020–2026. Thiết kế định lượng dựa trên ngữ liệu so sánh năm phương diện: đa dạng từ vựng, độ tinh vi từ vựng, độ phức tạp cú pháp, tính liên kết và độ dễ đọc. Ở cấp độ toàn ngữ liệu, các chỉ số từ vựng và độ dễ đọc không khác biệt sau hiệu chỉnh Holm, dù bài đọc đề thi biến động giữa các giai đoạn. Khác biệt rõ hơn nằm ở tính liên kết và tổ chức mệnh đề: bài đọc đề thi có ít từ nối hơn và mật độ mệnh đề phụ cao hơn. Kết quả cho thấy sự tương thích có chọn lọc; bằng chứng ngữ liệu đa chiều có thể hỗ trợ đánh giá học liệu, hiểu biết về đánh giá ngôn ngữ và dạy đọc tiếng Anh trong bối cảnh thi cử hệ quả cao.

Từ khóa - Ngôn ngữ học ngữ liệu; đọc hiểu tiếng Anh; sự tương thích sách giáo khoa–kỳ thi; hiểu biết về đánh giá ngôn ngữ; độ phức tạp văn bản

1. Introduction

In a reading curriculum, alignment is partly a matter of the texts students repeatedly encounter. Curriculum objectives and examination specifications may appear compatible, but that formal agreement says little about the vocabulary, clause structure, and discourse support built into the passages themselves. Constructive alignment concerns coherence among intended learning outcomes, teaching and learning activities, and assessment tasks [1]. The present study does not evaluate that full system. Instead, it isolates one narrower component, instructional–assessment textual alignment, by comparing the linguistic demands of sustained textbook and examination passages. When classroom texts offer one linguistic profile and a high-stakes examination presents another, the gap is experienced in everyday reading rather than only at the policy level.

This issue is particularly salient in Vietnam. The 2018 General Education English Curriculum adopts a competency-based orientation [2], [3]. Global Success is the textbook sequence examined in this study for Grades 10–12. The National High School Graduation Examination nevertheless continues to shape classroom priorities because it affects graduation and university admission. Whether textbook passages prepare students for examination texts is therefore a question for curriculum evaluation and teacher education.

A single readability score cannot answer that question. Two passages with similar sentence length may impose different demands if one contains less familiar vocabulary, denser clause relations, or fewer explicit links between ideas. Reading demand is therefore treated here as multidimensional, involving lexical diversity, lexical sophistication, syntactic complexity, cohesion, and surface readability [4], [5].

1.1. Conceptual and Empirical Background

Here, instructional–assessment textual alignment refers to the degree of convergence or divergence between the linguistic profiles of sustained textbook and examination reading passages. It is treated as a text-level component of

broader curriculum–assessment coherence rather than as evidence of full constructive alignment. Alignment is multidimensional rather than binary: convergence on one linguistic dimension does not offset divergence on another. Differences in vocabulary, clause organization, cohesion, or readability may indicate that some textual demands of the examination are less represented in regular instructional input and may contribute to washback [6], [7]. Cognitive Load Theory and recent L2 text-difficulty research further suggest that layered syntax and weak cohesive signaling can increase the processing resources needed to construct an integrated representation of a text [8], [9].

Readability formulas remain useful as broad indicators, but second-language readers also rely on lexical familiarity, syntactic cues, and connections among propositions [4], [5]. MTLD and HD-D represent lexical diversity [10]; AWL coverage estimates academic lexical load [11]; clausal and nominal measures distinguish forms of syntactic elaboration [12], [13]; and TAACO indices represent lexical overlap and connective support [14]. Recent L2 text-difficulty research further supports interpreting these dimensions jointly rather than relying on a single readability score [9]. No single index determines whether a corpus is harder. The relevant object is the profile formed by the indices and the processing demands that profile plausibly creates.

Corpus comparison makes such alignment observable. Gedik and Kolsal [15] compared Turkish high-school textbooks with university entrance examinations and demonstrated how a shared analytical framework can identify differences in lexical and syntactic complexity. Cheng and Chang [16] similarly reported lower readability, greater syntactic complexity, and higher vocabulary demands in Taiwanese college-entrance examination texts than in senior-high textbooks. A recent Chinese comparison also found lower readability in entrance-examination passages than in compulsory English textbooks [17]. In Vietnam**,** Lien et al. [18] documented the lexical demands of upper-secondary textbooks, while Phuong et al. [19] identified substantial lexical and readability demands in graduation-examination passages from 2015–2025. Nguyen and Tran [20] further reported substantially elevated lexical and readability demands in the 2025 examination relative to the curriculum's B1 benchmark. Within the Vietnamese literature identified here, however, these textbook and examination strands have not been directly compared across lexical, syntactic, cohesive, and readability dimensions.

Language assessment literacy provides a complementary interpretive rationale: teachers need evidence about which features of an assessment text create difficulty and whether regular materials provide sufficient preparation [21], [22]. The empirical framework therefore combines multidimensional text complexity with language assessment literacy, while constructive alignment is retained only as a system-level rationale.

1.2. Research Gap and Questions

The unresolved issue is the linguistic relationship between sustained reading passages in Global Success Grades 10–12 and those used in the National High School Graduation Examination. The present study operationalizes textual alignment as the pattern of convergence and divergence between instructional and assessment passages across five linguistic dimensions. It does not classify the corpora as wholly aligned or misaligned, but identifies dimensions in which their profiles converge and those in which examination demands diverge from regular textbook input.

RQ1. To what extent do the Global Success Grade 10–12 reading passages correspond with the National High School Graduation Examination reading passages in lexical diversity, lexical sophistication, syntactic complexity, cohesion, and readability?

RQ2. Which linguistic dimensions show the strongest pedagogically meaningful convergence or divergence, and what do these patterns imply for English teacher preparation, language assessment literacy, and reading pedagogy?

2. Methodology

2.1. Research Design and Corpus Materials

A quantitative corpus-based descriptive-comparative design was used. Passage boundaries were retained throughout the analysis, yielding 86 passage-level observations: 72 textbook passages and 14 examination passages. All descriptive and inferential analyses were conducted at the passage level rather than on grade- or year-aggregated files.

The Textbook Corpus contains 72 sustained passages from Global Success English for Grades 10–12 [23] – [25], with 24 passages from each grade. For each grade, the corpus included all ten core Reading passages and all ten Culture/CLIL passages, together with four Review passages. Coverage is therefore comprehensive for the core Reading and Culture/CLIL sections and supplemented by Review input, but it does not represent every form of language exposure in the textbooks. The Examination Corpus contains the two intact reading-comprehension passages from one operational English examination paper for each year from 2020 to 2026, yielding 14 passages. Alternative test-code versions were not treated as additional observations when they reproduced the same underlying reading passage. Because the examination format changed from 2025 [26], the 2020–2024 and 2025–2026 passages were additionally retained as predefined period subsets for secondary comparison. Passage length also differed between the corpora. Textbook passages had a median length of 214 words (IQR = 83.25; range = 125–343), whereas examination passages had a median length of 298.5 words (IQR = 77.25; range = 155–389). Although several indices are normalized or designed to reduce length effects, passage length can still influence some passage-level estimates. Length-related variation is therefore treated as a potential source of influence rather than as evidence of alignment or misalignment.

Functional comparability guided inclusion. The source unit was the main continuous reading text; unit titles, pre-reading prompts, task instructions, guided cloze passages,

item stems, answer options, and post-reading activities were excluded. The comparison therefore concerns sustained discourse rather than the surrounding teaching or testing apparatus. However, functional comparability does not imply equivalence in genre, topic, or communicative purpose. Textbook passages are designed for supported instruction, whereas examination passages are designed for assessment under time constraints; differences in discourse type and topic may therefore contribute to observed linguistic contrasts independently of textbook–examination alignment. Thus, "relevant texts" refers to passages meeting the predefined sustained-reading criterion within these bounded sources rather than every textual element in the textbooks or examination papers. Table 1 summarizes the composition and sampling scope of the two corpora.

Table 1. *Corpus Profile and Sampling Scope*

Corpus	Coverage	Included passage set	N
Textbook	10–12	Per grade: 10 Reading + 10 Culture/CLIL + 4 Review passages	72
Exam	2020 – 2026	2 intact sustained reading-comprehension passages per annual examination	14

Note. Each passage was retained as one statistical observation. The corpus targets sustained written reading discourse rather than all textbook or examination content. Test-code duplicates of the same underlying reading passage were not counted as additional observations.

2.2. Data Preparation and Quality Control

The passages were extracted manually and saved as UTF-8 plain-text files. A corpus register recorded provenance, word count, extraction notes, preprocessing decisions, and reasons for exclusion. The same cleaning sequence was applied to both corpora: task metadiscourse, exercises, item stems, answer options, headers, footers, examination codes, isolated numbers, and formatting artifacts were removed, and sentence boundaries were checked manually after PDF extraction.

Guided cloze texts were excluded because inserted blanks and item numbers disrupt sentence structure and primarily assess local lexico-grammatical choice. Proper nouns were retained for parsing and cohesion analysis, while sensitivity checks with and without named entities assessed their influence on lexical indices. A second trained coder checked at least 20% of the files for passage boundaries, punctuation restoration, and inclusion or exclusion decisions; disagreements were documented and resolved.

2.3. Analytical Instruments and Data Analysis

Five dimensions were analyzed. They were selected because L2 text difficulty is shaped jointly by lexical, syntactic, and discourse-level features rather than by surface readability alone [9]. MTLD and HD-D measured lexical diversity [10]; AWL coverage, computed with TAALES, represented academic lexical load [11], [27]. Mean length of clause, dependent clauses per clause and per T-unit, and complex nominals per clause and per T-unit represented clausal and phrasal complexity [12], [13]. Together, these measures distinguish clause length, subordination, and nominal elaboration. Cohesion was estimated from content-word overlap, argument overlap, and total and logical connective incidence using TAACO [14]. These measures represent explicit cohesive support between propositions. Flesch Reading Ease and Flesch–Kincaid Grade Level were treated as broad surface indicators rather than complete accounts of second-language reading difficulty. Both were calculated at passage level from the cleaned master texts using their standard formulas.

For every index, passage-level distributions were summarized using the mean, standard deviation, median, interquartile range, minimum, and maximum. Primary textbook–examination contrasts were tested using two-sided Mann–Whitney U tests with 72 textbook and 14 examination passages. Because 14 core linguistic indices were compared, Holm-adjusted p values were used to control the family-wise error rate. Cliff's δ quantified effect direction, with positive values indicating higher examination values and negative values indicating higher textbook values. Secondary Mann–Whitney comparisons contrasted Grade 12 passages (n = 24) with the 2020–2024 examination subset (n = 10) and separately with the 2025–2026 subset (n = 4), using the same effect-size and Holm-adjustment procedure. For interpretation, descriptive convergence was noted when the corpora were numerically similar and no robust Holm-adjusted distributional difference was observed. A nonsignificant result alone was not treated as evidence of statistical equivalence. Robust divergence required a directional contrast supported by its effect magnitude and a Holm-adjusted difference. Mixed patterns across indices were therefore reported as selective or partial textual alignment rather than reduced to a single binary classification.

No personal data were collected. Source passages were used only for scholarly linguistic analysis, and the article reports corpus-level statistics rather than reproducing complete copyrighted texts.

3. Results

All corpus contrasts reported below use the 86 passage-level observations. Inferential results are based on two-sided Mann–Whitney U tests with Holm-adjusted p values and Cliff's δ effect sizes.

Table 2 summarizes the corpus profiles across lexical diversity, lexical sophistication, readability, cohesion, and syntactic complexity.

The lexical measures did not show differences after Holm adjustment. Examination passages had somewhat higher medians for MTLD (76.38 vs. 73.32), HD-D (.828 vs..814), and AWL coverage (6.31% vs. 5.34%), with adjusted p values of.376,.138, and 1.000, respectively. MTLD was more dispersed in the examination corpus (IQR = 54.46) than in the textbook corpus (IQR = 26.90).

Readability showed directional differences but no overall difference after Holm adjustment. Median Flesch Reading Ease was lower in the examination corpus (53.68

vs. 55.64), whereas median Flesch–Kincaid Grade Level was higher (10.41 vs. 9.36); the adjusted p values were 1.000 and.376, respectively. Examination passages also showed greater dispersion on both readability measures. Table 3 presents the secondary comparisons between Grade 12 passages and the 2020–2024 and 2025–2026 examination subsets.

Table 2. *Multidimensional Corpus Profiles*

Measure	Textbook Md (IQR)	Exam Md (IQR)	Cliff's δ	U	Holm p
MTLD	73.32 (26.90)	76.38 (54.46)	.34	331	.376
HD-D	.814 (.041)	.828 (.058)	.42	293	.138
AWL coverage (%)	5.34 (3.67)	6.31 (5.62)	.10	455	1.000
Flesch Reading Ease	55.64 (11.34)	53.68 (16.57)	−.11	559	1.000
Flesch–Kincaid Grade Level	9.36 (2.25)	10.41 (3.36)	.34	335	.376
Content-word overlap	.088 (.050)	.074 (.069)	−.19	598	.822
Argument overlap	.139 (.095)	.114 (.077)	−.26	634	.649
All connectives (%)	8.15 (3.00)	5.94 (1.52)	−.53	769	.022
Logical connectives (%)	4.13 (2.06)	3.23 (1.18)	−.35	678.5	.376
MLC	11.65 (3.28)	9.45 (1.67)	−.64	826	.002
DC/C	.262 (.152)	.416 (.125)	.66	170	.001
DC/T	.351 (.274)	.750 (.382)	.70	150	<.001
CN/C	1.381 (.600)	1.255 (.192)	−.22	614.5	.793
CN/T	1.821 (.775)	2.202 (.603)	.32	342.5	.376

Note. Md = median; IQR = interquartile range. Positive Cliff's δ indicates higher examination values; negative δ indicates higher textbook values. Holm-adjusted p values control the 14-index comparison family. Individual nonsignificant contrasts are not treated as tests of statistical equivalence.

Table 3. *Grade 12 and Examination Period Comparisons*

Measure	Grade 12 Md (IQR)	Exam 2020–24 Md (IQR)	δ	Holm p	Exam 2025–26 Md (IQR)	δ	Holm p
MTLD	78.36 (23.86)	72.47 (32.16)	.00	1.000	158.61 (91.71)	.56	.630
HD-D	.820 (.029)	.820 (.042)	.10	1.000	.881 (.048)	.58	.630
AWL coverage (%)	5.71 (4.25)	4.08 (3.67)	−.32	1.000	9.58 (2.06)	.69	.346
FRE	53.78 (9.72)	56.42 (9.27)	.29	1.000	38.98 (8.08)	-.77	.154
FKGL	9.48 (2.65)	9.73 (1.28)	.05	1.000	12.60 (.80)	.94	.010

Note. δ represents each examination subset relative to Grade 12, with positive values indicating higher examination values. Holm p values are adjusted within the corresponding 14-index comparison family. The 2025–2026 subset contains four passages; this comparison is exploratory and is interpreted cautiously.

For the 2020–2024 subset, none of the five reported lexical or readability contrasts with Grade 12 remained significant after Holm adjustment. The 2025–2026 subset had higher medians for MTLD (158.61), HD-D (.881), AWL coverage (9.58%), and FKGL (12.60), and lower FRE (38.98) than Grade 12. Effect sizes ranged from moderate to very large for these contrasts, but only FKGL remained significant after Holm adjustment (δ =.94, Holm p =.010).

For cohesion, total connective incidence was lower in the examination corpus (Md = 5.94%) than in the textbook corpus (Md = 8.15%), U = 769, δ = −.53, Holm p =.022. Argument overlap, content-word overlap, and logical connective incidence were also lower in the examination corpus, but these contrasts did not remain significant after Holm adjustment.

Three syntactic contrasts remained significant after Holm adjustment. Median MLC was lower in the examination corpus (9.45 vs. 11.65, U = 826, δ = −.64, Holm p =.002). In contrast, DC/C was higher in the examination corpus (.416 vs..262, U = 170, δ =.66, Holm p =.001), as was DC/T (.750 vs..351, U = 150, δ =.70, Holm p <.001). Neither CN/C nor CN/T differed after adjustment.

4. Discussion

The comparison points to selective alignment. At the overall corpus level, the lexical and readability measures did not show Holm-adjusted differences. These nonsignificant contrasts indicate that robust distributional differences were not detected, but they do not establish statistical equivalence or, by themselves, demonstrate alignment. The overall profile is therefore interpreted as descriptive proximity on some measures alongside clear divergence on others. This pattern is less uniformly divergent than that reported by Gedik and Kolsal [15], who found higher lexical diversity, lexical sophistication, and syntactic complexity in Turkish university-entrance examinations, and by Cheng and Chang [16], who reported greater vocabulary load, lower readability, and higher syntactic complexity in Taiwanese examination texts. The present findings therefore extend this literature by showing that textbook–examination divergence may be concentrated in particular linguistic dimensions rather than appearing consistently across the full profile.

The four 2025–2026 passages showed a more demanding lexical and readability profile, but this comparison is exploratory because the subset contains only four passages. This descriptive pattern is consistent with Vietnamese examination research. Phuong et al. [19] reported low reading-ease scores and high lexical demands across the 2015–2025 examinations, while Nguyen and Tran [20] identified substantially elevated lexical and readability demands in the 2025 examination relative to the curriculum's B1 benchmark.

Recent work has examined the Global Success materials themselves. Hoa and Phuong [28] analyzed evaluative language in Global Success 12 reading texts, while Thi and Hang [29] evaluated accuracy- and fluency-oriented language production activities in English 10 – Global Success. Together with Phuong et al. [19], these studies provide relevant evidence on both textbook characteristics and examination reading demands in the Vietnamese upper-secondary context. The present study extends this line of work by directly comparing textbook and examination reading passages across lexical, syntactic, cohesive, and readability dimensions.

Descriptively, the examination corpus showed greater dispersion in MTLD (SD = 63.12 vs. 19.34), Flesch Reading Ease (12.76 vs. 9.85), and Flesch–Kincaid Grade Level (2.20 vs. 1.70). Given the small 2025–2026 subset,

however, this dispersion should not be interpreted as evidence of a temporal trend.

This distinction matters for curriculum–assessment coherence and washback. An examination need not reproduce textbook passages, and some increase in challenge is defensible. The problem arises when the type or level of challenge changes sharply without being represented in regular instruction. However, the observed differences need not reflect textbook inadequacy alone. Textbook passages are designed for supported instruction across a course, whereas examination passages serve assessment under time constraints, and variation in passage topic or discourse type may also contribute to the contrasts observed here. The 2025–2026 observations also coincide chronologically with the redesigned National High School Graduation Examination introduced from 2025 under the 2018 curriculum [26]. However, with only four passages in this subset, the present data cannot determine whether the observed differences reflect the examination redesign, year-specific passage selection, topic or discourse-type variation, or other factors. Greater transparency about the linguistic range of examination texts would help teachers preserve curriculum goals while preparing students realistically [1], [6].

The strongest divergence emerged from the combination of fewer connective cues and denser subordination. This partly supports earlier textbook–examination studies reporting greater syntactic complexity in assessment texts [15], [16], but the present pattern is more specific: examination passages had shorter clauses alongside substantially denser subordination. Greater subordination requires readers to identify hierarchical clause relations and retain dependent information while integrating it with the main proposition. When explicit connectives are also less frequent, some relations between propositions must be inferred rather than directly signaled, so partially integrated information may need to remain active during subsequent processing [8], [9]. These concurrent demands may increase processing load, particularly for L2 readers whose linguistic processing is less automatized [9]. Reading instruction should therefore include implicit cause and effect, contrast, exemplification, reference chains, and clause hierarchy alongside vocabulary and comprehension questions.

The syntactic pattern also cautions against using clause length alone as a proxy for complexity. Learners may benefit from identifying main clauses, bracketing subordinate material, tracing relative-clause antecedents, and restating the main proposition before integrating embedded information.

The lexical findings call for a different response. Grade 12 AWL coverage did not show a Holm-adjusted difference from the 2020–2024 examination subset, whereas the four 2025–2026 passages showed higher academic-word coverage descriptively. Because the recent subset is small, this contrast should not be treated as evidence of a stable period effect. More resilient instruction develops word-family knowledge, collocation, morphological analysis, and contextual inferencing in addition to deliberate vocabulary study [11].

For materials evaluation and language assessment literacy, each index answers a limited question. MTLD and HD-D describe lexical diversity; AWL coverage estimates academic lexical load; connective incidence represents overt discourse signaling; dependent-clause measures capture subordination; and Flesch–Kincaid Grade Level summarizes surface readability. Their value comes from reading the profile as a whole and linking each pattern to a plausible instructional response.

National membership is not evidence of linguistic alignment. The textbook and examination corpora belong to the same educational system, yet several overall lexical and readability measures did not differ after Holm adjustment, while connective incidence and dependent-clause density clearly differed. Materials evaluation should therefore supplement topic coverage and overall readability with a small set of interpretable lexical, syntactic, and discourse measures. The aim is not to replace teacher judgment, but to show more precisely what that judgment needs to consider.

5. Conclusion

This study compared sustained reading passages from Global Success Grades 10–12 with reading-comprehension passages from the 2020–2026 National High School Graduation Examination. The findings indicate selective textual alignment rather than uniform convergence or divergence across linguistic dimensions. At the overall corpus level, lexical and readability measures did not show Holm-adjusted differences, but this absence of significant differences should not be interpreted as statistical equivalence or as evidence of alignment by itself. The clearest divergences were lower connective incidence and greater dependent-clause density in the examination passages.

The main pedagogical issue lies in the interaction between discourse support and clause organization. A readability formula cannot distinguish a vocabulary problem from a clause-structure problem or a lack of connective cues. Combining lexical, syntactic, cohesive, and readability measures makes those sources of difficulty more visible and provides a firmer basis for materials evaluation and bridging instruction.

The findings remain bounded by one textbook sequence and 14 examination passages from seven examination years, with a substantially smaller examination sample and generally longer passages in the examination corpus. Because passage length can influence some passage-level linguistic estimates, this difference should also be considered when interpreting the corpus contrasts. The analysis excluded item stems, answer options, guided cloze passages, and test specifications, and automated tagging, parsing, and manual preprocessing introduce uncertainty. Future work should add textbook series and examination years and compare linguistic indices with learner performance, teacher judgments, and classroom evidence.

Within those limits, the corpus profile identifies both continuity and instructional gaps. Some lexical and readability measures did not show Holm-adjusted differences, but examination passages showed less explicit connective support and denser subordination. The four passages from the 2025–2026 examinations showed a more demanding lexical and readability profile descriptively, but this small subset is exploratory and cannot establish a temporal trend or an effect of the 2025 examination redesign. These distinctions can guide supplementary text selection, clause- and discourse-focused reading tasks, and teachers' interpretation of high-stakes reading demand.

AI Disclosure: Generative AI was used to support language editing and grammatical correction. The authors reviewed every change, retained responsibility for the study design and data analysis, and approved the final wording, interpretations, and conclusions.

REFERENCES

[1] J. B. Biggs, "Enhancing teaching through constructive alignment," *Higher Education*, vol. 32, no. 3, pp. 347–364, 1996. https://doi.org/10.1007/BF00138871

[2] Ministry of Education and Training, *General Education English Language Curriculum*, Circular No. 32/2018/TT-BGDĐT, 2018.

[3] H. V. Van, "Interpreting MoET's 2018 General Education English Curriculum," *VNU Journal of Foreign Studies*, vol. 38, no. 5, pp. 1–22, 2022. https://doi.org/10.25073/2525-2445/vnufs.4866

[4] A. C. Graesser, D. S. McNamara, M. M. Louwerse, and Z. Cai, "Coh-Metrix: Analysis of text on cohesion and language," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 2, pp. 193–202, 2004. https://doi.org/10.3758/BF03195564

[5] D. S. McNamara, A. C. Graesser, P. M. McCarthy, and Z. Cai, *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge, U.K.: Cambridge University Press, 2014. https://doi.org/10.1017/CBO9780511894664

[6] J. C. Alderson and D. Wall, "Does washback exist?," *Applied Linguistics*, vol. 14, no. 2, pp. 115–129, 1993. https://doi.org/10.1093/applin/14.2.115

[7] C. L. T. Nguyen and Q. T. N. Dao, "Assessing the washback effects of the 2025 National High School Graduation Exam: A cross-sectional study in Vietnam," *Vietnam Journal of Education*, vol. 9, no. 3, pp. 424–436, 2025. https://doi.org/10.52296/vje.2025.565

[8] J. Sweller, "Element interactivity and intrinsic, extraneous, and germane cognitive load," *Educational Psychology Review*, vol. 22, no. 2, pp. 123–138, 2010. https://doi.org/10.1007/s10648-010-9128-5

[9] X. Zhang and X. Lu, "Testing the relationship of linguistic complexity to second language learners' comparative judgment on text difficulty," *Language Learning*, vol. 74, no. 3, pp. 672–706, 2024. https://doi.org/10.1111/lang.12633

[10] P. M. McCarthy and S. Jarvis, "MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment," *Behavior Research Methods*, vol. 42, no. 2, pp. 381–392, 2010. https://doi.org/10.3758/BRM.42.2.381

[11] A. Coxhead, "A new academic word list," *TESOL Quarterly*, vol. 34, no. 2, pp. 213–238, 2000. https://doi.org/10.2307/3587951

[12] K. Kyle and S. A. Crossley, "Assessing syntactic sophistication in L2 writing: A usage-based approach," *Language Testing*, vol. 34, no. 4, pp. 513–535, 2017. https://doi.org/10.1177/0265532217712554

[13] X. Lu, "Automatic analysis of syntactic complexity in second language writing," *International Journal of Corpus Linguistics*, vol. 15, no. 4, pp. 474–496, 2010. https://doi.org/10.1075/ijcl.15.4.02lu

[14] S. A. Crossley, K. Kyle, and D. S. McNamara, "The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion," *Behavior Research Methods*, vol. 48, no. 4, pp. 1227–1237, 2016. https://doi.org/10.3758/s13428-015-0651-7

[15] T. A. Gedik and Y. S. Kolsal, "A corpus-based analysis of high school English textbooks and English university entrance exams in Turkey," *Theory and Practice of Second Language Acquisition*, vol. 8, no. 1, pp. 157–176, 2022. https://doi.org/10.31261/TAPSLA.9152

[16] Y.-S. Cheng and S.-C. Chang, "Do senior high school English textbooks sufficiently prepare students for the high-stakes college entrance examinations? A corpus-based analysis of text difficulty," *Journal of Textbook Research*, vol. 15, no. 1, pp. 43–80, 2022. https://doi.org/10.6481/JTR.202204_15(1).02

[17] Z. Zhou and M. C. de Almeida Prado, "A corpus-based comparative study of readability of passages in compulsory Chinese English textbooks and exams for middle school students," in *2024 13th International Conference on Educational and Information Technology (ICEIT 2024)*, Chengdu, China, Mar. 22–24, 2024. Piscataway, NJ, USA: IEEE, 2024, pp. 279–283. https://doi.org/10.1109/ICEIT61397.2024.10540975

[18] N. N. Lien, N. H. Mai, and N. H. Trang, "Vocabulary in English textbooks for Vietnamese upper-secondary students: A comparative analysis of reading passages," *TESL-EJ*, vol. 28, no. 2, pp. 1–25, 2024. https://doi.org/10.55593/ej.28110a10

[19] C. T. H. Phuong, T. L. Chi, and N. T. M. Hang, "Analysis of the reading passage difficulty and lexical demands in the Vietnamese High School Graduation English Tests (2015–2025)," *The University of Danang - Journal of Science and Technology*, vol. 24, no. 1, pp. 63–68, 2026. https://doi.org/10.31130/ud-jst.2026.24(1).681E

[20] N. T. K. An and T. T. Thanh, "Examining CEFR reading comprehension text complexity in Vietnam's 2025 National English Exam," *HNUE Journal of Science: Journal of Educational Science*, vol. 71, no. 4, pp. 57–66, 2026. https://doi.org/10.18173/2354-1075.2026-0057

[21] G. Fulcher, "Assessment literacy for the language classroom," *Language Assessment Quarterly*, vol. 9, no. 2, pp. 113–132, 2012. https://doi.org/10.1080/15434303.2011.642041

[22] Y. Xu and G. T. L. Brown, "Teacher assessment literacy in practice: A reconceptualization," *Teaching and Teacher Education*, vol. 58, pp. 149–162, 2016. https://doi.org/10.1016/j.tate.2016.05.010

[23] H. V. Van, H. T. X. Hoa, C. Q. Binh, H. T. H. Hai, K. T. T. Huong, and N. T. K. Phuong, *English 10: Global Success*. Hanoi, Vietnam: Vietnam Education Publishing House, 2022.

[24] H. V. Van, V. H. Ha, C. Q. Binh, H. T. H. Hai, K. T. T. Huong, and N. T. K. Phuong, *English 11: Global Success*. Hanoi, Vietnam: Vietnam Education Publishing House, 2023.

[25] H. V. Van, V. H. Ha, C. Q. Binh, H. T. H. Hai, and N. T. K. Phuong, *English 12: Global Success*. Hanoi, Vietnam: Vietnam Education Publishing House, 2024.

[26] Ministry of Education and Training, Quality Management Department, "Structure and format of the High School Graduation Examination from 2025," *Quality Management Department, Ministry of Education and Training*, Dec. 29, 2023. [Online]. Available: https://vqa.moet.gov.vn/vi/news/thong-bao/cau-truc-dinh-dang-de-thi-tot-nghiep-thpt-tu-nam-2025-74.html [Accessed July 03, 2026].

[27] K. Kyle, S. A. Crossley, and C. M. Berger, "The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0," *Behavior Research Methods*, vol. 50, no. 3, pp. 1030–1046, 2018. https://doi.org/10.3758/s13428-017-0924-4.

[28] N. T. Q. Hoa and H. L. Phuong, "A study of 'Judgement' in English reading texts related to Global Citizenship in the 'Global Success' Grade 12 textbook and workbook," *The University of Danang - Journal of Science and Technology*, vol. 24, no. 6B, pp. 1–7, 2026. https://doi.org/10.31130/ud-jst.2026.24(6B).209E

[29] L. N. B. Thi and N. T. M. Hang, "Evaluating accuracy and fluency in language production activities in 'Tiếng Anh 10 – Global Success'," *The University of Danang - Journal of Science and Technology*, vol. 24, no. 5C, pp. 24–29, 2026. https://doi.org/10.31130/ud-jst.2026.24(5C).293.