Evaluating Academic Reading Input in Vietnamese EFL Textbooks: Lexical, Syntactic, and Cohesive Evidence for Materials Design and Classroom Pedagogy

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

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

[1]Faculty of Foreign Languages, Ho Chi Minh University of Banking, Ho Chi Minh City, Vietnam

[2]Center for Continuing Education, The University of Da Nang, Da Nang, Vietnam

[3]Faculty of Education, University of Social Sciences and Humanities, Vietnam National University Ho Chi Minh City, Ho Chi Minh City, Vietnam

Received: June 12, 2026 *Revised: July 19, 2026* *Accepted: August 17, 2026*

Abstract

The acquisition of academic vocabulary and complex linguistic structures is essential as learners move from general English proficiency toward academic literacy. In many Asian EFL contexts, textbooks remain a major source of lexical and structural input. This study presents a multi-dimensional corpus-based evaluation of 72 reading passages in the Global Success textbook series for Vietnamese upper-secondary students. Using TAALED, TAALES, TAASSC, and TAACO, the passages were analyzed as separate text-level units to examine lexical diversity, lexical sophistication, syntactic complexity, and cohesion across grades. Mean Academic Word List (AWL) coverage showed a non-linear descriptive pattern, declining from 5.90% (*SD* = 2.98) in Grade 10 to 5.41% (*SD* = 2.94) in Grade 11 before rising to 6.95% (*SD* = 3.52) in Grade 12. Mean Measure of Textual Lexical Diversity (MTLD) scores increased descriptively from 70.64 (*SD* = 14.55) to 74.04 (*SD* = 17.50) and 79.24 (*SD* = 21.15), respectively. However, Kruskal–Wallis tests found no statistically significant between-grade differences in AWL coverage or MTLD. Syntactic and cohesion measures also showed substantial passage-level variation, with only small descriptive differences in grade-level means. Overall, the findings do not support a uniform progression across Grades 10–12 and are better characterized by considerable heterogeneity among individual passages. The study demonstrates how text-level corpus analysis can identify areas that may warrant consideration for additional vocabulary and structural support in Vietnamese upper-secondary reading instruction.

Keywords: Lexical coverage, Lexical diversity, Syntactic complexity, Text cohesion, EFL textbooks, Vietnamese upper-secondary education

Introduction

The landscape of English Language Teaching (ELT) in Vietnam has shifted significantly as the country integrates more deeply into the global academic and economic community. Central to this transformation is the 2018 General Education Program, a major curricular reform introduced by the Vietnamese Ministry of Education and Training (MOET) (Hoang, 2015; Ministry of Education and Training, 2018). The reform promotes a communicative, competence-based orientation in English education. Recent studies have also begun to examine how this orientation is reflected in the *Global Success* series, including its cultural representation. As part of this reform, the *Global Success* textbook series has been widely adopted in Vietnamese schools as a major source of English instruction. Aligned with the Common European Framework of Reference for Languages (CEFR), the series aims to develop communicative competence while building the academic literacy needed for higher education.

Linguistic development, encompassing vocabulary knowledge and syntactic mastery, is widely recognized as a key predictor of reading comprehension and overall language proficiency. For upper-secondary students (Grades 10–12), learning academic vocabulary and navigating complex sentence structures become particularly important as they move from everyday English toward academic discourse. Coxhead's (2000) Academic Word List (AWL) provides a key benchmark for this transition, supporting students preparing for university study and English-medium academic contexts. Simultaneously, the capacity to process sophisticated syntactic patterns and varied cohesive devices is critical for establishing fluent text comprehension.

Despite the widespread use of the *Global Success* series, its multi-dimensional linguistic rigor remains underexplored. Existing studies focus mainly on cultural representation or selected sections of lower secondary and Grade 10 materials. Few corpus-based studies have examined how lexical, syntactic, and cohesive features vary across the entire upper-secondary reading curriculum. To address this literature gap, the present study adopts a multi-dimensional corpus linguistics approach using the Tool for the Automatic Analysis of Lexical Diversity (TAALED), the Tool for the Automatic Analysis of Lexical Sophistication (TAALES), the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC), and the Tool for the Automatic Analysis of Cohesion (TAACO) to analyze the reading passages in Grades 10, 11, and 12. Reading passages were deliberately selected as they serve as the primary source of academic input in EFL classrooms where outside exposure is limited. Identifying uneven descriptive patterns across grade levels may inform both materials evaluation and classroom planning. Although the corpus analysis does not examine teacher practices directly, its findings can indicate passages or linguistic features for which teachers may consider additional vocabulary support, structural explanation, or guided reading activities.

From a TESOL perspective, this study contributes to materials evaluation and academic literacy pedagogy in three ways. First, it provides empirical evidence on how lexical, syntactic, and cohesive features are distributed across the grade levels of a nationally adopted EFL textbook series. Second, it

examines whether passage-level distributions reveal potential discontinuities in lexical, syntactic, and cohesive input that may warrant pedagogical consideration. Third, it provides context-specific linguistic evidence that may inform academic vocabulary instruction, expository reading exposure, and textbook sequencing in Vietnamese upper-secondary education. Although the analytical approach may be replicated in other EFL textbook contexts, the present findings are limited to the Global Success series and should not be treated as directly transferable beyond this setting. These contributions show how corpus-based textbook evaluation can move beyond descriptive material review and generate practical evidence for classroom reading pedagogy, curriculum planning, and academic literacy preparation. The study is guided by the following research questions:

1. What are the lexical profiles of the reading passages in the *Global Success* series across Grades 10, 11, and 12, in terms of lexical diversity (TAALED) and academic sophistication (TAALES)?
2. How does syntactic complexity, measured through TAASSC indices of dependency structure and clausal subordination rather than sentence length alone, vary across the three grade levels?
3. How do content-word overlap and connective use, as measured by TAACO, vary across Grades 10, 11, and 12?

By addressing these questions, this study provides data-driven insights into the multi-dimensional linguistic demands of contemporary Vietnamese ELT materials. The findings offer practical implications for curriculum developers, textbook authors, and teachers seeking to strengthen structural and vocabulary instruction and better prepare students for the demands of academic English.

Literature Review

Lexical Coverage, Academic Vocabulary, and Reading Comprehension Thresholds

The relationship between vocabulary size and reading comprehension has been widely examined in applied linguistics. A key concept in this research is lexical coverage, defined as the percentage of running words in a text that a reader already knows (Nation, 2006). Identifying the lexical threshold, the minimum coverage required to understand a text without heavy external support, has important implications for textbook design, curriculum sequencing, and material selection.

Early work by Laufer (1989) proposed a minimal threshold of 95% lexical coverage, suggesting that readers must know at least 95 out of every 100 words to grasp the general meaning of an academic text. At this level, readers encounter about one unknown word every two lines, which often allows meaning to be inferred from context (Laufer & Ravenhorst-Kalovski, 2010). Later studies argue that 95% represents only a survival level. Nation (2006) shows that roughly 98% coverage is needed for fluent, independent reading and incidental vocabulary learning, reducing unknown words to about one in every 50 tokens. Schmitt et al. (2011) further demonstrate a linear relationship between lexical

knowledge and comprehension. Their findings suggest that learners typically need 98–99% vocabulary knowledge to reach about 70% comprehension. Building on these results, Laufer and Ravenhorst-Kalovski (2010) propose a dual-threshold model: 95% coverage for adequate comprehension (4,000–5,000 word families) and 98% for optimal comprehension (around 8,000). This framework provides practical benchmarks for evaluating EFL textbooks.

Within this broader concern with lexical coverage, academic vocabulary occupies a central position in preparing learners for advanced reading. While the General Service List (GSL) covers roughly 75–80% of general written texts, it is widely considered insufficient for academic discourse. To address this gap, Coxhead (2000) developed the Academic Word List (AWL), demonstrating that its 570 word families account for about 10% of the running words in academic texts. When combined with high-frequency words, mastery of these items raises learners' lexical coverage of academic texts closer to comprehension thresholds (Nation, 2006). AWL coverage therefore provides a descriptive basis for comparing the academic-vocabulary profiles of EFL textbook passages with those reported for academic corpora.

Multi-Dimensional Analytical Framework for Textbook Text Complexity

Textbook evaluation should distinguish the linguistic complexity of texts from the difficulty experienced by individual readers. Vocabulary use varies significantly across contexts (Hyland & Tse, 2007), and characterizing textbook reading input therefore requires attention to multiple linguistic properties rather than vocabulary alone. In the present study, lexical diversity, lexical sophistication, syntactic complexity, and text cohesion are treated as characteristics of the texts themselves. They are not interpreted as direct measures of learners' cognitive load, processing effort, experienced difficulty, or comprehension. Recent work on L2 text difficulty has related linguistic features to external proficiency-level classifications (Zhang & Lu, 2025); the present study does not include such learner- or rater-based measures. Accordingly, the four dimensions are used here to provide a multi-dimensional linguistic profile of the reading passages.

Beyond lexical properties, syntactic complexity constitutes another dimension of the linguistic structure of academic text. In second language acquisition (SLA) research, advanced academic prose is increasingly recognized not merely by sentence elongation or clausal embedding, but by phrasal-level complexity and compressed discourse structures (Biber et al., 2011; Kyle, 2016). As textbooks transition from narrative to expository registers, structural complexity manifests through dense nominal subject dependencies and prepositional phrases that pack high amounts of information into compact grammatical units. Evaluating these fine-grained structural variables through automated tools allows systematic computational measurement of selected linguistic features across the textbook passages.

Concurrently, text cohesion represents an important dimension of reading-text organization. Textual cohesion involves explicit linguistic cues, including content-word overlap and connectives, that signal relationships within and across sentences (Crossley et al., 2019). These features may support

textual continuity, but their relationship with reading difficulty is not unidirectional. Greater lexical overlap or more frequent connective use does not automatically make a passage easier, while lower values do not by themselves demonstrate greater inferencing demands. Their pedagogical significance depends on the interaction among reader proficiency, text type, vocabulary, syntax, and instructional context. Accordingly, the present study treats TAACO indices as descriptions of cohesion patterns rather than direct measures of comprehension or inferencing difficulty.

Based on these considerations, the present study adopts a multi-dimensional analytical framework for profiling textbook text complexity. In this framework, textual complexity is operationalized through four interrelated linguistic dimensions: lexical diversity, lexical sophistication, syntactic complexity, and text cohesion. Lexical diversity reflects the range of vocabulary to which learners are exposed and is operationalized through MTLD using TAALED. Lexical sophistication reflects the academic and frequency-based status of vocabulary and is operationalized through AWL coverage and COCA Academic Frequency using TAALES. Syntactic complexity reflects how information is organized through dependency structure and clausal subordination. It is operationalized through TAASSC indices of nominal subjects, prepositional modifiers, and direct objects per clause, together with dependent clauses per clause and complex T-units per T-unit.

Text cohesion reflects the degree to which semantic and logical relationships are signaled explicitly across sentences. It is operationalized through TAACO indices of adjacent content-word overlap and all, basic, and logical connectives. The analysis adopts a focused rather than exhaustive selection of indices because the four tools generate numerous measures, several of which capture closely related linguistic properties. The indices were selected according to three criteria: their direct correspondence with the dimensions specified in the analytical framework, their interpretability for textbook and reading-material evaluation, and their support in previous validation or application studies. This selection represents distinct aspects of lexical diversity, lexical sophistication, dependency structure, clausal subordination, and cohesion while limiting redundant measures and unnecessary statistical comparisons.

This analytical framework directly informs the three research questions of the study. RQ1 addresses the lexical dimension by examining lexical diversity and academic sophistication. RQ2 addresses the syntactic dimension by examining dependency-based structural patterns and clausal subordination. RQ3 addresses the cohesion dimension by examining content-word overlap and connective use through TAACO. Together, these dimensions allow the study to move beyond isolated word-frequency counts and provide a more integrated account of how academic reading demands are distributed across Grades 10, 11, and 12 in the *Global Success* series.

Automated Corpus-Based Textbook Evaluation

The emergence of corpus linguistics has transformed the evaluation of language teaching materials by shifting the field from intuition-based judgments to empirical analyses of lexical, syntactic,

and cohesive input. In recent years, automated linguistic profiling has been increasingly used to examine whether instructional texts are appropriately sequenced for L2 learners. For example, Jin et al. (2020) examined syntactic complexity in adapted teaching materials and showed that grade-level progression can be benchmarked through syntactic indices rather than through general readability formulas alone. Similarly, Shao et al. (2022) compared Chinese EFL learners' argumentative writing with expert essays from EFL textbooks and found that textbook essays contained more elaborate phrasal structures, including noun phrases, prepositional phrases, and embedded modifiers. These findings suggest that textbook input may differ from learner production not only in vocabulary difficulty, but also in the density of phrasal elaboration.

More recently, Zhang and Lu (2025) used a multi-feature linguistic complexity model to classify 1,181 English texts for L2 learners according to CEFR levels. Their study incorporated lexical diversity, lexical sophistication, syntactic complexity, and cohesion-related measures, including indices generated through TAALED, TAALES, and TAACO. The findings showed that text difficulty for L2 learners cannot be adequately explained by traditional readability formulas alone. Instead, lexical sophistication, nominal complexity, and cohesion features jointly contribute to how teachers classify reading materials across proficiency levels. This is particularly relevant to textbook evaluation because it demonstrates that pedagogical text difficulty is multi-dimensional and should be interpreted through several linguistic layers.

However, previous corpus-based studies of textbooks and teaching materials remain limited in three important respects. First, many studies focus on a single dimension, such as vocabulary coverage or syntactic complexity, rather than examining how lexical sophistication, syntactic density, and cohesion interact. Second, studies that use automated indices often treat numerical outputs as material-design evidence without sufficiently linking them to teacher mediation and reading pedagogy. Third, relatively few studies have examined whether a nationally adopted EFL textbook series provides a coherent progression of academic reading input across consecutive grade levels.

The present study addresses these limitations by integrating lexical diversity, lexical sophistication, syntactic complexity, and cohesion within one analytical framework. Rather than treating TAALED, TAALES, TAASSC, and TAACO as isolated measurement tools, the study examines how the selected linguistic features are distributed across Grades 10, 11, and 12 and considers their potential pedagogical implications. In this study, curriculum sequencing refers narrowly to the ordering and distribution of linguistic features across the three grade levels. It does not represent an evaluation against a formal model of optimal sequencing. Similarly, classroom scaffolding is not analyzed as an empirical variable because the study does not examine classroom observations, teacher practices, or teacher manuals. Instead, scaffolding refers to possible pedagogical responses suggested by the linguistic profiles, such as vocabulary support, noun-phrase unpacking, connective awareness, and guided inferencing activities.

Corpus-Based Textbook Evaluation in the Asian EFL Context

Studies across Asian EFL contexts have highlighted concerns about both the lexical focus of textbook materials and the adequacy of high-frequency lexical targets for learners (Dang & Webb, 2016; Yu & Renandya, 2021). These materials frequently lack sufficient AWL coverage and the repetition needed for incidental acquisition of academic vocabulary (Dang & Webb, 2016; Yu & Renandya, 2021). This limitation may have lasting consequences; for example, insufficient exposure at the secondary level has been linked to ongoing difficulties with academic vocabulary among university students (Evans & Morrison, 2011).

Together, these studies highlight a recurring challenge in Asian EFL curricula: textbook design often prioritizes grammatical sequencing and thematic variety over systematic lexical development, leaving materials concentrated at basic high-frequency levels (Yu & Renandya, 2021). However, much of this research has focused primarily on vocabulary coverage and frequency profiles. Less attention has been paid to how lexical sophistication, syntactic complexity, and text cohesion vary together across grade levels. This is an important limitation because the linguistic complexity of textbook reading input is multi-dimensional. A textbook may show adequate lexical progression but still provide insufficient syntactic density or overly explicit cohesion. Conversely, it may become syntactically denser without providing enough academic vocabulary support.

These multi-dimensional linguistic profiles may have implications for TESOL pedagogy and teacher mediation. Corpus-based textbook evaluation can identify textual features that teachers may wish to consider when planning reading support, although linguistic profiles alone do not establish which instructional responses will be effective. For passages characterized by relatively dense academic vocabulary or limited explicit cohesion, possible pedagogical responses include vocabulary support, attention to lexical relations, and guided inferencing activities. Similarly, passages containing substantial prepositional modification or compressed noun phrases may provide occasions for teachers to draw learners' attention to these structures. These examples are presented as pedagogical possibilities informed by the linguistic characteristics of the texts rather than as instructional effects demonstrated by the present study.

The Vietnamese Context and the Research Gap

In Vietnam, the use of corpus linguistics for textbook evaluation remains relatively limited. Existing studies on the *Global Success* series have mainly focused on qualitative aspects, such as cultural representation in the Grade 11 volume (Vuong et al., 2025). In terms of vocabulary, Pham et al. (2025) noted the presence of advanced lexical resources but pointed to a lack of communicative practice in Grade 10. Similarly, Mai et al. (2024) analyzed the lexical demands of listening sections and found noticeable differences in complexity across publishers.

Despite these contributions, comprehensive cross-grade analyses of academic-language features remain scarce. While recent studies on the *Global Success* series have explored cultural representation (Vuong et al., 2025) or uni-dimensional vocabulary profiles for individual grades using traditional frequency lists (Pham et al., 2025), a critical research gap remains. It is therefore unclear whether the series provides a consistent multi-dimensional progression or whether it reflects the linguistic plateaus frequently observed in other Asian EFL contexts (Yu & Renandya, 2021). To our knowledge, no published study has applied an integrated multi-dimensional analytical framework combining lexical diversity, lexical sophistication, syntactic complexity, and cohesion to the reading passages across Grades 10–12 of the Global Success series. The present study addresses this gap by using TAALED, TAALES, TAASSC, and TAACO to profile the reading passages of the *Global Success* series across Grades 10, 11, and 12.

Materials and Methods

Corpus Compilation

To systematically characterize the multi-dimensional linguistic features of the reading input in Vietnamese upper-secondary textbooks, a specialized pedagogical mini-corpus was compiled from the Global Success English textbooks for Grades 10, 11, and 12 (Hoang et al., 2022, 2023, 2024). Although modest in token size, the corpus was designed to represent the sustained written reading input provided in the designated reading-oriented sections of the three upper-secondary textbooks. In this bounded textbook context, representativeness was defined in relation to the specific instructional sections under investigation rather than the complete linguistic content of each book (Biber, 1993).

For each grade, the corpus included all ten core Reading passages, all ten Culture/CLIL passages, and four Review passages, yielding 24 texts per grade and 72 texts in total. It therefore represents the sustained written input in these reading-oriented sections rather than the textbooks as a whole. Dialogues, speaking scripts, task instructions, and classroom-management prompts were excluded because they do not constitute sustained written reading discourse and typically display more conversational vocabulary and shorter clauses (Biber, 1988).

The passages also differed descriptively in theme and register. Grade 10 combined personal, social, cultural, and environmental topics in narrative and informational texts. Grade 11 emphasized socio-cultural and everyday-life topics through narrative, explanatory, and discussion-based passages. Grade 12 contained more knowledge-oriented and expository topics, including artificial intelligence, environmental issues, and globalization. These descriptions provide context for interpreting the quantitative findings but do not constitute formal thematic or genre coding.

All 72 passages were retained as separate text files and treated as individual units of analysis. This design allowed within-grade variation to be estimated and between-grade patterns to be examined statistically. The findings should therefore be interpreted as evidence about the reading component of the Global Success upper-secondary series, not its complete linguistic content.

Data Cleaning and Preparation

All 72 selected reading texts were digitally extracted, manually verified, and saved as plain text (.txt) files. To ensure high data fidelity for computational processing, the preparation phase followed a standardized digitization and cleaning workflow. The targeted passages were first converted using Optical Character Recognition (OCR) software. The digitized files were then checked line by line against the printed textbook versions to remove OCR transcription errors and formatting artifacts.

Data cleaning was conducted manually using a predefined exclusion rubric (see Appendix A). Three authors independently applied the predefined exclusion rubric (Appendix A) to a random pilot sample of three texts (approximately 1,366 tokens, representing roughly 9% of the total corpus). The coders evaluated which textbook elements (e.g., proper nouns, task instructions) should be retained or removed for each specific tool. This pilot phase yielded an almost perfect Fleiss' kappa coefficient, $\kappa = 0.896$ (Fleiss, 1971). Minor discrepancies, primarily regarding the treatment of hyphenated words and multi-word proper nouns, were resolved through focused discussion until full consensus was reached before the team proceeded to clean the remaining corpus. This procedure ensured that the final corpus was consistently prepared for automated analysis.

The cleaning process removed task directions, question numbers, bullet symbols, isolated digits, Vietnamese glosses, and other non-discursive textbook elements. These items were excluded because computational tools treat such strings as analyzable tokens, which can distort indices of lexical diversity, lexical sophistication, syntactic complexity, and cohesion. For lexical analyses using TAALED and TAALES, proper nouns, geographic names, and organizational acronyms were excluded to prevent the artificial inflation of lexical diversity scores and to isolate generalizable vocabulary. Conversely, for syntactic and cohesion analyses using TAASSC and TAACO, the original sentence structures, including proper nouns, were preserved. Removing them would disrupt natural dependency parsing and intersentential cohesion patterns. Consequently, the total word counts generated by TAASSC were approximately 1.6% higher than the lexical token counts generated by TAALED. This difference resulted primarily from the task-specific cleaning procedures: proper nouns, geographic names, and organizational acronyms were excluded from the lexical files but retained in the syntactic files to preserve complete sentence structures for dependency parsing. Minor additional differences may also reflect tool-specific treatment of hyphenated forms, contractions, and punctuation boundaries. The discrepancy therefore represents a predictable consequence of using separately prepared inputs and different tokenization procedures rather than missing or inconsistent corpus data.

Analytical Instruments and Framework

Rather than relying on basic vocabulary counts, this study utilizes a state-of-the-art suite of natural language processing (NLP) tools developed for academic text analysis. The primary instruments and their core indices are summarized in Table 1.

Table 1.

Multi-Dimensional NLP Suite and Core Indices

Tool	Dimension Tested	Primary Core Indices
TAALED (v1.4.1)	Lexical Diversity	Measure of Textual Lexical Diversity (MTLD)
TAALES (v2.2)	Sophistication	Normed AWL, COCA Academic Frequency
TAASSC (v1.3.8)	Syntactic Complexity	Dependency ratios (nsubj, prep, and dobj per clause); dependent clauses per clause (DC/C); complex T-units per T-unit (CT/T)
TAACO (v2.0.4)	Cohesion & Logic	Adjacent Content-Word Overlap; All, Basic and Logical Connectives

Although the tools generate numerous indices, the study selected a restricted set corresponding directly to the four dimensions specified in the analytical framework and offering clear pedagogical interpretation. MTLD represented lexical diversity; AWL coverage and COCA Academic Frequency represented lexical sophistication; three dependency ratios and two subordination indices represented syntactic complexity; and content-word overlap and connective use represented cohesion. This focused selection limited overlapping variables and unnecessary statistical comparisons. The automated outputs are therefore treated as systematic measurements of the selected linguistic features rather than as inherently objective evaluations. Their interpretation is tied to the operational definitions, selected indices, and preprocessing procedures used in the present study.

TAALED (Tool for the Automatic Analysis of Lexical Diversity): Used to calculate the *Measure of Textual Lexical Diversity (MTLD)*. MTLD tracks vocabulary variation and is highly resistant to text length fluctuations, ensuring reliable comparisons across sub-corpora (Kyle et al., 2021).

TAALES (Tool for the Automatic Analysis of Lexical Sophistication): The principal indices were *All_AWL_Normed* and COCA Academic Frequency (Kyle & Crossley, 2015; Kyle et al., 2018). *All_AWL_Normed* represents AWL-matched running-word tokens divided by the total word count. It was multiplied by 100 and reported as AWL coverage. For example, a value of 0.059 represents

approximately 5.9% coverage. Lower COCA Academic Frequency values indicate less frequent and potentially more specialized academic vocabulary.

TAASSC (Tool for the Automatic Analysis of Syntactic Sophistication and Complexity): Dependency structure was measured through nominal subjects, prepositional modifiers, and direct objects per clause. Dependent clauses per clause (*DC_C*) and complex T-units per T-unit (*CT_T*) represented clausal subordination. Together, these measures distinguish dependency-based patterns from subordinate-clause use without relying on sentence length alone (Kyle, 2016).

TAACO (Tool for the Automatic Analysis of Cohesion): The selected indices measured adjacent content-word overlap and the normalized incidence of all, basic, and logical connectives (Crossley et al., 2019). Connective values were multiplied by 100 and reported as percentages. These indices describe cohesion patterns rather than directly measuring comprehension or inferencing difficulty.

Analytical Procedure

The 72 cleaned reading passages were processed as separate text files through each automated tool using the default parameters for written texts. The specific tool versions are reported in Table 1. Each passage was coded according to its grade level, resulting in three groups of 24 texts, and was retained as an individual observation. This decision was necessary because concatenating the passages would have produced only one aggregated observation per grade level. The resulting absence of within-grade variance, rather than concatenation in itself, would have prevented statistically defensible inferential comparisons across the three grades.

For each linguistic index, grade-level means and standard deviations were calculated to describe the central tendency and dispersion of the 24 passages. Because several indices showed substantial within-grade variation and did not consistently meet the assumptions required for parametric comparisons, between-grade differences were examined using the Kruskal–Wallis test. Statistical significance was evaluated at $\alpha = .05$, and epsilon-squared (ε^2) was calculated to estimate the effect size of each omnibus comparison. Dunn's pairwise tests with Holm adjustment were planned for any statistically significant omnibus result. However, because none of the selected Kruskal–Wallis tests reached the .05 significance level, no pairwise post-hoc comparisons were conducted.

The analysis remained descriptive and diagnostic in orientation. Statistical tests were used to determine whether the observed grade-level patterns were sufficiently consistent across individual passages, rather than to establish a predetermined developmental sequence. References to textbook themes and registers in the Discussion are used only to contextualize possible linguistic patterns. They are not presented as findings from a separate qualitative thematic analysis.

Results and Discussion

The text-level analysis provides a multi-dimensional profile of lexical, syntactic, and cohesive features across the three grade levels. To maintain a direct correspondence with the research questions, the findings are presented in the same sequence as the three questions. Section 4.1 addresses lexical diversity and lexical sophistication in response to RQ1. Section 4.2 examines dependency structure and clausal subordination in response to RQ2. Section 4.3 reports content-word overlap and connective use in response to RQ3. Pedagogical implications are presented after the three empirical sections.

Lexical Diversity and Sophistication (TAALED & TAALES)

Table 2 summarizes the text-level lexical diversity and sophistication results for the 24 individual passages in each grade.

Table 2.

Lexical Diversity and Sophistication Across Grades 10–12

Grade	Total tokens	MTLD *M* (*SD*)	AWL coverage (%) *M* (*SD*)	COCA Academic Frequency *M* (*SD*)
10	4,065	70.64 (14.55)	5.90 (2.98)	8,527.91 (1,795.32)
11	5,234	74.04 (17.50)	5.41 (2.94)	8,204.46 (1,756.52)
12	5,914	79.24 (21.15)	6.95 (3.52)	8,041.45 (1,890.15)

Note. Total tokens are corpus totals generated by TAALED. The remaining values are grade-level means, with standard deviations in parentheses, based on 24 individual passages per grade. MTLD = Measure of Textual Lexical Diversity; AWL = Academic Word List. AWL coverage represents the percentage of running-word tokens matched to the AWL and was calculated as *All_AWL_Normed* × 100. Lower COCA Academic Frequency values indicate the use of words occurring less frequently in the academic component of the Corpus of Contemporary American English.

At the text level, the grade-level means retained a descriptive contrast between lexical diversity and academic vocabulary coverage. Mean MTLD values were 70.64 (*SD* = 14.55), 74.04 (*SD* = 17.50), and 79.24 (*SD* = 21.15) for Grades 10, 11, and 12, respectively. Mean AWL coverage showed a non-linear pattern, declining from 5.90% (*SD* = 2.98) in Grade 10 to 5.41% (*SD* = 2.94) in Grade 11 before rising to 6.95% (*SD* = 3.52) in Grade 12. However, the Kruskal–Wallis tests showed no statistically significant between-grade differences in either MTLD, $H(2) = 2.06$, $p = .357$, $\varepsilon^2 = .001$, or AWL coverage, $H(2) = 2.11$, $p = .348$, $\varepsilon^2 = .002$. The substantial within-grade variation indicates that these patterns should be interpreted as tentative distributional tendencies rather than evidence that Grade 11

performs a specific fluency-building function or that Grade 12 represents a confirmed stage of academic lexical acceleration.

Mean COCA Academic Frequency values were 8,527.91 (*SD* = 1,795.32) in Grade 10, 8,204.46 (*SD* = 1,756.52) in Grade 11, and 8,041.45 (*SD* = 1,890.15) in Grade 12. Because lower values indicate lower-frequency and potentially more specialized academic vocabulary, the grade-level means show a weak descriptive directional pattern. However, the between-grade difference was not statistically significant, $H(2) = 1.14$, $p = .565$, $\varepsilon^2 < .001$, and the substantial within-grade variation does not support interpreting this pattern as a consistent developmental progression in lexical sophistication.

The findings partially converge with Yu and Renandya's (2021) concern about uneven academic vocabulary exposure in Asian EFL textbooks. However, the studies are not directly equivalent because Yu and Renandya examined both vocabulary coverage and repetition, whereas the present study did not track recurring AWL word families across units. The lower mean AWL coverage in Grade 11 should therefore be treated as a context-specific descriptive tendency rather than a confirmed plateau or evidence of insufficient repetition. Although previous research suggests that thematic and grammatical sequencing may not produce a consistent increase in academic vocabulary, differences in textbook systems, educational levels, corpus design, and lexical measures limit direct comparison. The thematic and register profiles described in Section 3.1 may help contextualize the pattern, but they were not formally coded as analytical variables. Future research could examine the recurrence and distribution of academic word families across lessons and grades. Figures 1 and 2 visualize the passage-level distributions underlying the grade-level means for MTLD and AWL coverage. The overlapping distributions illustrate the substantial within-grade variation and support the cautious interpretation of the descriptive patterns.

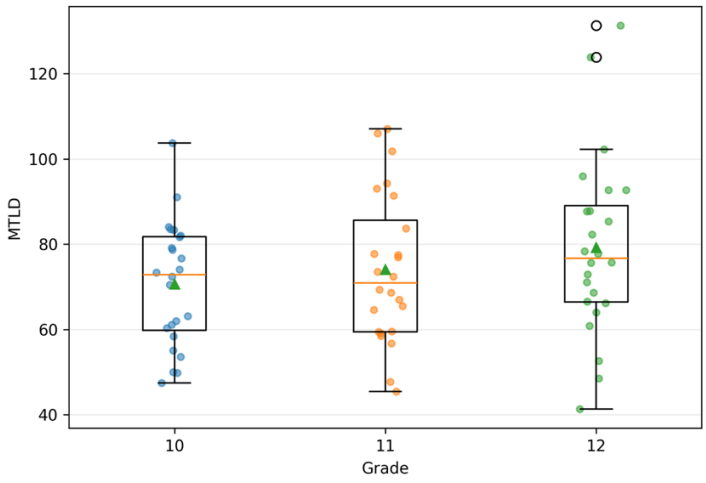


Figure 1. *Distribution of MTLD Scores Across Individual Passages in Grades 10–12*

Note. Each grade includes 24 individual reading passages. Boxes represent the interquartile range, horizontal lines indicate medians, triangles indicate means, and dots represent individual passages.

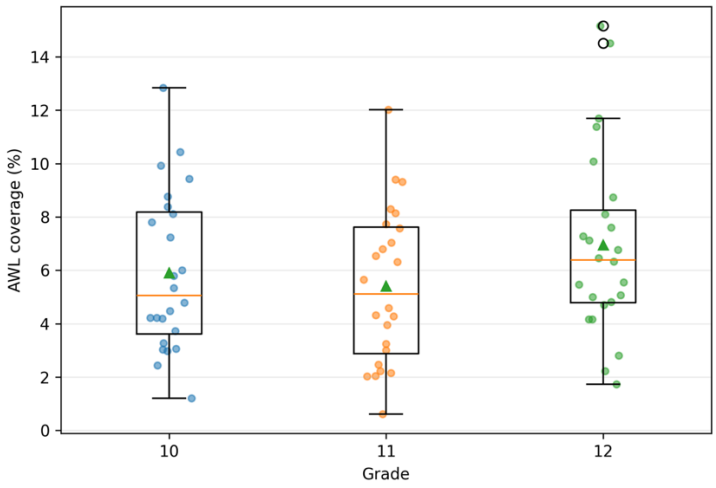


Figure 2. *Distribution of AWL Coverage Across Individual Passages in Grades 10–12*

Note. Each grade includes 24 individual reading passages. Boxes represent the interquartile range, horizontal lines indicate medians, triangles indicate means, and dots represent individual passages. AWL coverage was calculated as *All_AWL_Normed* × 100.

As a descriptive comparative reference point, Coxhead (2000) reported that Academic Word List families account for approximately 10% of the running words in university-level academic texts. This corpus-based figure is used here for comparison rather than as a threshold that upper-secondary textbook materials are expected to attain. In the present study, the Grade 12 passages recorded a mean text-level AWL coverage of 6.95% (*SD* = 3.52). When the Grade 12 passages were considered collectively and weighted by text length, AWL coverage was 7.17%. Both figures remain below the level commonly reported for university-level academic texts.

This comparison should be interpreted cautiously. It does not constitute a statistical test against the university benchmark, nor does it imply that every upper-secondary textbook passage should reproduce the lexical density of university prose. Moreover, AWL coverage describes the academic vocabulary contained in the reading materials rather than learners' vocabulary knowledge or their overall comprehension of the texts. This comparison provides a descriptive indication of how the academic-vocabulary profile of the Grade 12 passages differs from that reported for university-level academic texts. It therefore identifies a possible area for pedagogical consideration rather than a deficiency in the textbook series itself. As students approach higher education, teachers may consider supplementing textbook input with academically oriented reading materials and opportunities to recycle academic word families, depending on learners' proficiency and instructional aims. These suggestions are pedagogical implications of the observed lexical profiles; their effects on vocabulary development or reading comprehension were not tested in the present study.

Syntactic Complexity (TAASSC)

Table 3 summarizes the passage-level distributions of dependency structure and clausal subordination across the three grade levels. The selected indices represent nominal subjects, prepositional modifiers, and direct objects per clause, together with dependent clauses per clause and complex T-units per T-unit.

Table 3.

Syntactic Complexity Across Grades 10–12

Grade	Total words	Nominal subjects per clause *M* (*SD*)	Prepositional modifiers per clause *M* (*SD*)	Direct objects per clause *M* (*SD*)	DC/C *M* (*SD*)	CT/T *M* (*SD*)
10	4,144	0.559 (0.135)	0.311 (0.112)	0.534 (0.145)	0.279 (0.132)	0.320 (0.189)
11	5,279	0.547 (0.116)	0.339 (0.112)	0.494 (0.155)	0.275 (0.130)	0.326 (0.146)
12	6,038	0.555 (0.135)	0.380 (0.172)	0.442 (0.123)	0.281 (0.088)	0.344 (0.159)

Note. Total words are corpus totals generated by TAASSC. The remaining values are grade-level means, with standard deviations in parentheses, based on 24 individual passages per grade. DC/C = dependent clauses per clause; CT/T = complex T-units per T-unit.

Nominal-subject use remained relatively stable across the three grades, with means of 0.559 (*SD* = 0.135) in Grade 10, 0.547 (*SD* = 0.116) in Grade 11, and 0.555 (*SD* = 0.135) in Grade 12. The Kruskal–Wallis test showed no statistically significant between-grade difference, $H(2) = 0.12$, $p = .943$, $\varepsilon^2 < .001$.

Mean prepositional-modification values were 0.311 (*SD* = 0.112) in Grade 10, 0.339 (*SD* = 0.112) in Grade 11, and 0.380 (*SD* = 0.172) in Grade 12. Although the grade-level means were descriptively higher in the upper grades, the between-grade difference was not statistically significant, $H(2) = 2.32$, $p = .314$, $\varepsilon^2 = .005$, and substantial passage-level variation was evident.

Mean direct objects per clause were 0.534 (*SD* = 0.145) in Grade 10, 0.494 (*SD* = 0.155) in Grade 11, and 0.442 (*SD* = 0.123) in Grade 12. Although the grade-level means were descriptively lower in the upper grades, the difference did not reach statistical significance, $H(2) = 4.45$, $p = .108$, $\varepsilon^2 = .036$. This pattern should therefore not be interpreted as evidence of a systematic shift away from transitive clausal structures.

The two measures of clausal subordination varied only minimally across grades. Mean dependent clauses per clause were 0.279 (*SD* = 0.132), 0.275 (*SD* = 0.130), and 0.281 (*SD* = 0.088) for Grades 10, 11, and 12, respectively, with no statistically significant difference, $H(2) = 0.28$, $p = .868$, $\varepsilon^2 < .001$. Mean complex T-units per T-unit were 0.320 (*SD* = 0.189), 0.326 (*SD* = 0.146), and 0.344

(*SD* = 0.159) for Grades 10, 11, and 12, respectively, with no statistically significant between-grade difference, $H(2) = 0.23$, $p = .889$, $\varepsilon^2 < .001$.

Overall, the syntactic results are characterized more strongly by passage-level variation than by grade-level differentiation. Although some grade-level means differed descriptively, none of the selected syntactic indices showed a statistically significant between-grade difference, and the findings do not support a uniform Grade 12 syntactic surge.

Text Cohesion and Connectives (TAACO)

The cohesion results, presented in Table 4, show substantial variation among individual passages within each grade. Adjacent content-word overlap represents the degree of lexical continuity between neighboring sentences. Higher values indicate greater repetition of content vocabulary across sentences. However, overlap alone should not be treated as direct evidence of reading ease or inferencing demand.

Table 4.

Text Cohesion and Overlap Across Grades 10–12

Grade	Adjacent content-word overlap *M* (*SD*)	All connectives (%) *M* (*SD*)	Basic connectives (%) *M* (*SD*)	Logical connectives (%) *M* (*SD*)
10	0.088 (0.044)	8.22 (2.45)	4.56 (1.30)	4.71 (2.06)
11	0.094 (0.029)	9.04 (2.24)	5.39 (1.48)	4.43 (1.77)
12	0.098 (0.051)	8.34 (2.08)	4.83 (1.49)	4.19 (1.50)

Note. Values are grade-level means, with standard deviations in parentheses, based on 24 individual passages per grade. Connective values were converted into percentages by multiplying the corresponding normalized TAACO indices by 100.

Mean adjacent content-word overlap was 0.088 (*SD* = 0.044) in Grade 10, 0.094 (*SD* = 0.029) in Grade 11, and 0.098 (*SD* = 0.051) in Grade 12. However, the Kruskal–Wallis test showed no statistically significant difference across the three grade levels, $H(2) = 0.63$, $p = .729$, $\varepsilon^2 < .001$. The overlapping passage-level distributions do not support a reliable grade-level progression in content-word overlap. They also do not provide evidence that Grade 11 uses lexical repetition as a compensatory scaffolding mechanism or that Grade 12 passages are systematically less repetitive.

Connective use also showed descriptive rather than statistically reliable differences. Grade 11 recorded the highest mean for all connectives and basic connectives, while the grade-level mean for logical connectives was lower in Grade 12 than in Grade 10. Nevertheless, the between-grade differences were not statistically significant for all connectives, $H(2) = 1.51$, $p = .470$, $\varepsilon^2 < .001$; basic

connectives, $H(2) = 4.60$, $p = .100$, $\varepsilon^2 = .038$; or logical connectives, $H(2) = 0.93$, $p = .628$, $\varepsilon^2 < .001$. These results indicate considerable passage-level variation rather than a stable cohesion sequence across the three grades. The passage-level variation suggests that any consideration of connective awareness or guided inferencing should be informed by the characteristics of individual passages rather than by an assumed grade-level scaffolding mechanism.

Pedagogical Implications for Classroom Reading Instruction

The substantial passage-level variation suggests that pedagogical support may be better considered in relation to the linguistic characteristics of individual texts than to an assumed uniform grade-level progression. The following implications are therefore presented as pedagogical possibilities informed by the linguistic profiles and the literature reviewed above, rather than as instructional effects demonstrated by the present study.

Teachers may consider complementing some narrative and socio-cultural passages with short expository texts on accessible scientific, environmental, social, or technological topics where this is appropriate to learners' proficiency and instructional aims. Such supplementation could provide additional opportunities for exposure to informational registers and academic vocabulary. Vocabulary recycling activities, including attention to word families and collocations, may also be considered where particular passages contain relevant lexical demands.

For students approaching higher education, teachers may consider supplementary passages that provide additional exposure to academic vocabulary, with text selection guided by learners' proficiency and instructional aims. The approximately 10% AWL coverage reported for university-level academic texts is used here only as a descriptive comparative reference and should not be interpreted as an instructional target for upper-secondary materials. Teachers may also consider drawing attention to extended noun phrases, prepositional modification, subordinate clauses, and cohesive relations when these features are prominent in particular passages. Activities such as identifying head nouns and modifiers, unpacking dense structures, or tracing connections across sentences represent possible forms of instructional support; their effectiveness was not evaluated in the present study.

Conclusion

Through text-level corpus profiling, this study examined lexical diversity, lexical sophistication, syntactic complexity, and cohesion across 72 individual reading passages in the Global Success upper-secondary series. The principal pattern across the selected indices was substantial variation among individual passages within each grade. None of the selected Kruskal–Wallis comparisons reached statistical significance, and the corresponding effect sizes were negligible or small. Although some grade-level means differed descriptively, these differences do not support a

confirmed Grade 11 academic vocabulary plateau, a Grade 12 lexical-syntactic surge, or a uniform developmental progression across the three grades.

The pedagogical relevance of these findings lies primarily at the level of individual passages rather than fixed grade-level stages. Teachers may use the linguistic profiles of particular texts to identify areas in which additional attention to academic vocabulary, prepositional and clausal structures, or cohesive relations could be considered. Supplementary expository reading input may likewise be considered where appropriate, but Grade 11 should not be assigned a predetermined bridging function, and Grade 12 should not be treated as uniformly more demanding. These implications should be interpreted as possible instructional responses to the observed textual characteristics rather than as interventions whose effects were established in the present study.

For curriculum developers and textbook writers, the substantial variation among individual passages suggests that academic language demands should be monitored at both the grade and text levels. The present findings do not establish a confirmed Grade 11 plateau or demonstrate that a uniformly linear increase in complexity would necessarily provide better pedagogical sequencing. Instead, materials development may benefit from balancing thematic diversity with repeated exposure to academic word families, varied expository reading input, and manageable structural demands throughout the upper-secondary cycle. This interpretation is broadly consistent with previous research showing that lexical progression in Asian EFL textbooks may be uneven, although differences in textbook systems, educational levels, corpus design, and analytical measures limit direct comparison (Yu & Renandya, 2021). Further evaluation involving learner comprehension and classroom implementation is needed before specific changes to textbook sequencing can be recommended.

The study also shows the value of multi-dimensional computational profiling for TESOL materials evaluation. Vocabulary coverage alone cannot adequately characterize the linguistic complexity of textbook reading passages. Future evaluations should combine lexical, syntactic, and cohesion indices with learner comprehension data, classroom observation, and multimodal textbook analysis. Such triangulation would provide a more complete account of how national EFL textbooks prepare learners for academic English literacy.

This study is subject to several methodological limitations. First, although the 72 passages were retained as separate units of analysis, the corpus was drawn from a single textbook series and included only 24 passages per grade. The statistical comparisons therefore describe variation within this bounded corpus and should not be generalized to other textbook series or to Vietnamese upper-secondary EFL materials as a whole. In addition, the modest group sizes and substantial passage-level heterogeneity may have limited the ability of the tests to detect small between-grade effects. The absence of statistically significant differences should therefore not be interpreted as evidence that the three grade levels are linguistically equivalent. Future research should replicate the analysis using larger corpora drawn from multiple textbook series and publishers. Second, the study uses Coxhead's (2000) Academic Word List because the present aim is to evaluate broad academic vocabulary exposure against

an established and widely used benchmark. The Academic Vocabulary List (AVL) developed by Gardner and Davies (2014) offers an alternative corpus-based approach and may be particularly useful for studies examining contemporary academic vocabulary across disciplines or comparing discipline-specific lexical patterns. Future research could apply AWL and AVL measures in parallel to determine how different operationalizations of academic vocabulary influence the evaluation of upper-secondary reading materials. Finally, the specialized corpus focuses solely on written reading sections, excluding multimodal scaffolds, listening input, and teacher mediation, which collectively shape the ultimate learning experience.

Declaration of Generative AI in the Writing Process

The authors employed ChatGPT to improve the readability and language of the manuscript during the preparation of this work. Following its use, the authors carefully reviewed and revised the content to ensure accuracy, coherence, and alignment with the intended purpose. The authors take full responsibility for the final version of the publication.

References

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press. https://doi.org/10.1017/CBO9780511621024

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243–257. https://doi.org/10.1093/llc/8.4.243

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5–35. https://doi.org/10.5054/tq.2011.244483

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., & Dascalu, M. (2019). The tool for the automatic analysis of cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods, 51*(1), 14–27. https://doi.org/10.3758/s13428-018-1142-4

Dang, T. N. Y., & Webb, S. (2016). Evaluating lists of high-frequency words. *ITL: International Journal of Applied Linguistics, 167*(2), 132–158. https://doi.org/10.1075/itl.167.2.02dan

Evans, S., & Morrison, B. (2011). Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong. *English for Specific Purposes, 30*(3), 198–208. https://doi.org/10.1016/j.esp.2011.01.001

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378–382. https://doi.org/10.1037/h0031619

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305–327. https://doi.org/10.1093/applin/amt015

Hoang, V. V. (2015). The development of the ten-year English textbook series for Vietnamese schools under the National Foreign Language 2020 Project: A cross-cultural collaborative experience. *VNU Journal of Foreign Studies, 31*(3), 1–17.

Hoang, V. V., Hoang, T. X. H., Chu, Q. B., Vu, H. H., Hoang, T. H. H., Kieu, T. T. H., & Nguyen, T. K. P. (2022). *Tiếng Anh 10: Global Success* [English 10: Global Success]. Vietnam Education Publishing House.

Hoang, V. V., Vu, H. H., Chu, Q. B., Hoang, T. H. H., Kieu, T. T. H., & Nguyen, T. K. P. (2023). *Tiếng Anh 11: Global Success* [English 11: Global Success]. Vietnam Education Publishing House.

Hoang, V. V., Vu, H. H., Chu, Q. B., Hoang, T. H. H., & Nguyen, T. K. P. (2024). *Tiếng Anh 12: Global Success* [English 12: Global Success]. Vietnam Education Publishing House.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly, 41*(2), 235–253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Jin, T., Lu, X., & Ni, J. (2020). Syntactic complexity in adapted teaching materials: Differences among grade levels and implications for benchmarking. *The Modern Language Journal, 104*(1), 192–208. https://doi.org/10.1111/modl.12622

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. https://doi.org/10.57709/8501051

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly, 49*(4), 757–786. https://doi.org/10.1002/tesq.194

Kyle, K., Crossley, S. A., & Berger, C. M. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods, 50*(3), 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

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15–30. https://doi.org/10.64152/10125/66648

Mai, N. H., Lien, N. N., & Nguyen, H. T. (2024). Lexical demands and features of English textbooks for Vietnamese 10th graders: An in-depth comparison of listening sections. *SAGE Open, 14*(4), 1–15. https://doi.org/10.1177/21582440241299247

Ministry of Education and Training. (2018). *Thông tư số 32/2018/TT-BGDĐT ban hành Chương trình giáo dục phổ thông* [Circular No. 32/2018/TT-BGDĐT promulgating the general education program].

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Pham, T. T. T., Tu, T. L., & Nguyen, T. H. N. (2025). An investigation into English vocabulary and lexical activities in "Global Success" English textbooks (Grade 3, 6, 10). *HNUE Journal of Science: Journal of Educational Science, 70*(4), 97–105. https://doi.org/10.18173/2354-1075.2025-0073

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Shao, Z., Zhang, H., Zhang, J., Zhong, Y., & Xu, X. (2022). Phrasal complexity in English argumentative writing: Variations across Chinese STEM versus English majors' production and EFL textbook essays. *Journal of English for Academic Purposes, 55*, Article 101070. https://doi.org/10.1016/j.jeap.2021.101070

Vuong, K. T. M., Trakulkasemsuk, W., & Zahabi, A. (2025). English textbooks in the modern world: Cultural representation in an EFL textbook from Vietnam. *PASAA, 70*, 204–238. https://doi.org/10.58837/CHULA.PASAA.70.7

Yu, M., & Renandya, W. A. (2021). A corpus-based study of the vocabulary profile of high school English textbooks in China. *LEARN Journal: Language Education and Acquisition Research Network, 14*(1), 28–49.

Zhang, X., & Lu, X. (2025). Aligning linguistic complexity with the difficulty of English texts for L2 learners based on CEFR levels. *Studies in Second Language Acquisition, 47*(5), 1407–1434. https://doi.org/10.1017/S0272263125101125

Nhat Quang Vo is an MA student in Applied Linguistics at Ho Chi Minh University of Banking. He has over a decade of English teaching experience. His mixed-methods research interests include Critical Discourse Analysis, Second Language Acquisition, Semantics, and language pedagogy, with attention to their applications in English language education.
ORCID: https://orcid.org/0009-0006-6908-5163

Quoc Khanh Nguyen is an IELTS instructor and an English Linguistics undergraduate at The University of Da Nang. He has over ten years of English teaching experience. His mixed-methods

research interests include Critical Discourse Analysis, TESOL, and Corpus Linguistics, with a particular focus on their applications to English language education.
ORCID: https://orcid.org/0009-0007-0171-7942

Thi Thuy Loan To is a faculty member at the Faculty of Education, University of Social Sciences and Humanities, Vietnam National University Ho Chi Minh City. With over nine years of academic experience, her research focuses on Education Studies, Educational Management, and online and blended learning in tertiary education contexts.
ORCID: https://orcid.org/0009-0002-2999-5286

Appendix

Data Exclusion Rubric and Decision Rules

The following table details the criteria used for the manual cleaning of the reading sub-corpora from the *Global Success* series. These elements were systematically removed to ensure that the multi-dimensional computational analyses, spanning lexical diversity (TAALED), sophistication (TAALES), syntactic complexity (TAASSC), and text cohesion (TAACO), accurately reflect generalizable academic input rather than topic-specific noise, punctuation distortions, or pedagogical framing.

Category	Decision Rules	Rationale & Examples
Proper Nouns and Acronyms in Lexical Analyses	**Excluded:** Names of individuals, geographic locations, and organizational acronyms.	To prevent the artificial inflation of Lexical Diversity (MTLD) scores and isolate generalizable lexis from highly topic-bound entities. For TAASSC and TAACO analyses, sentence-level structure was preserved to avoid disrupting dependency parsing and cohesion patterns. *e.g., Nam, Vietnam, Hanoi, UNICEF, ASEAN.*
Task Instructions & Rubrics	**Excluded:** Meta-discourse used for classroom management or exercise directions.	To isolate authentic reading discourse from pedagogical framing and task-related repetitions. *e.g., "Read the passage and choose...", "Match the words".*
Numbers & Symbols	**Excluded:** Isolated digits, dates, percentages, and mathematical notations.	These are parsed as unique tokens by computational algorithms, distorting total token, type counts, and phrasal dependency ratios. *e.g., 2025, 50%, $, +.*
Non-English Annotations	**Excluded:** Any Vietnamese words, phrases, or translational scaffolding found within the textbooks.	To maintain a pure English lexical dataset for accurate baseline profiling across reference lists. *e.g., Vietnamese translations in glossaries.*
Conversational Fragments	**Excluded:** Spoken-discourse markers found in interview-style or	To maintain the focus on written academic register and prevent distortion of clausal complexity and academic word frequency indices.

Category	Decision Rules	Rationale & Examples
	dialogue-based reading passages.	*e.g., Oh, well, wow, hi.*