ANALYSIS OF THE READING PASSAGE DIFFICULTY AND LEXICAL DEMANDS IN THE VIETNAMESE HIGH SCHOOL GRADUATION ENGLISH TESTS (2015-2025)

Cao Thi Hong Phuong[1], Thai Linh Chi[1], Nguyen Thi My Hang[2]*

[1]*Faculty of English, Hanoi National University of Education, Vietnam*
[2]*The University of Danang - University of Foreign Language Studies, Vietnam*

*Corresponding author: ntmhang@ufl.udn.vn

(Received: November 29, 2025; Revised: December 22, 2025; Accepted: January 22, 2026)

DOI: 10.31130/ud-jst.2026.24(1).617E

Abstract - This study investigates the difficulty of the reading comprehension texts in the Vietnamese High School Graduation English Test (VGET), a matter of significant interest to various educational stakeholders. Twenty two passages from 2015 to 2025 were analyzed using classic readability formulas and lexical profiling. The findings suggest a potential misalignment between the measured text difficulty and the expected language proficiency level. This is supported by low reading ease scores (often below 50) and high lexical demand, requiring approximately 3,000-5,000 word families for 95% coverage. Texts with this level of difficulty may not be optimally suited for evaluating student graduation outcomes and could potentially create an excessive burden on test-takers. The study suggests that current test administration practices be reconsidered to ensure the test's long-term validity and relevance. It also offers practical implications for both teachers and students adapting to the observed text complexity.

Key words - Graduation English Test; reading comprehension; text difficulty; readability; lexical demand

1. Introduction

In language testing and assessment, national high-stakes English tests have long been regarded as one of the most important tools for evaluating a student's language proficiency for entrance into tertiary level. In Vietnam, they also serve another purpose as a high school graduation certification. For more than a decade, continued efforts have been made by the Ministry of Education and Training (MOET) to reform and standardize English assessment, a process significantly shaped by the issuance of Circular No. 01/2014/TT-BGDĐT, which established the six-level Foreign Language Proficiency Framework for Vietnam [1]. This landmark policy aimed at aligning national student outcomes with international standards, specifically setting the B1 level (Level 3) as the expected benchmark for high school graduates.

These sustained efforts were further solidified with the introduction of the 2018 General Education English Curriculum, which emphasizes the development of communicative competence and academic abilities required for higher education [2]. Most recently, this journey reached a new milestone in 2025, when the Prime Minister approved a scheme to transition English toward a second language in schools [3]. Within this strategic framework, the assessment system is expected to be diverse, flexible, and practical, ensuring that students can use English effectively in thinking processes and across various academic subjects.

However, despite these long-term institutional goals and the established proficiency framework, the actual content of the national high school English tests remains a subject of intense debate. Throughout the 2015–2025 period, and most notably in the 2025 examination, the tests have faced criticism for their fluctuating difficulty levels. The 2025 test received particularly strong opposition from educators and the public for being disproportionately complex for upper-secondary learners. Public discourse on social platforms and insights from reputable media outlets consistently noted that the exam contained a significant amount of B2–C1 level vocabulary, whereas the intended learning outcome, as stated in the national framework, was set at B1.

This concern stems from the lack of technical analysis reports of the test's difficulty. Considering the matrix of the test papers, despite various changes in its format, the focus on two reading comprehension passages remained unchanged, accounting for at least 30% of the whole test. Consequently, students' performance on this section heavily influenced their overall score. Therefore, evaluating the difficulty of these passages is critical to better understand the challenge posed by the test and to ensure their alignment with the objectives of the MOET.

In examining the text features that have an impact on reading comprehension, readability indices and vocabulary knowledge are identified as two of the most important factors [4]. However, studies on them have largely been conducted separately, which left a gap in the theoretical framework for analyzing how readable a text is. In the context of Vietnam, text complexity gained increasing attention, but most studies only concentrated on textbook materials rather than examination texts. Due to this limitation, this paper aims to examine the readability scores and lexical demand of the reading passages in the national high school English tests, to better understand the complexity of these tests, which is deemed crucial to inform appropriate modifications in test design and provide practical applications for the classroom. Therefore, the research questions (RQs) of this paper are:

RQ1: What are the readability levels of reading passages used in Vietnamese High School Graduation English Tests (VGET) from 2015 to 2025?

RQ2: What is the lexical coverage of the texts and to what extent do they align with the recommended threshold for reading comprehension?

2. Literature review

2.1. The Vietnamese High School Graduation English Test (VGET)

The Vietnamese High School Graduation English Test has been implemented as a high-stakes test administered to senior and graduate high school students. Its purpose has undergone significant changes over the past several decades. Since 2015, the test has served a dual function: (1) certifying student's completion of high school and (2) selecting students for university [5]. Therefore, the test components are required to align with the target language proficiency level, at B1 level in the Common European Framework of Reference for Languages (CEFR), as specified in the Dispatch 7972/BGDĐT-GDTrH [6] and 2018 English Language Curriculum [2], while being able to demonstrate capability of students in mastering their English language skills to perform efficiently in their academic quest at higher educational institutions. Assessing the reading comprehension passages' difficulty provides some empirical evidence regarding the test's validity and its ability to meet the dual objectives.

2.2. Test Validity and Fairness

Validity is one of the most important components in evaluating the usefulness of a test, defined as the extent to which a test measures what it is intended to measure [7]. A validity test can ensure appropriate and meaningful interpretation of the test results, allow test-takers to demonstrate their abilities accurately and receive appropriate recognition. In doing so, test developers may help foster learners' motivation and support the development of clear learning goals.

The validity of a test can be influenced by various factors, ranging from its design and implementation to external conditions. One key factor is the alignment of test difficulty with the proficiency of the test-takers. A test that is either too difficult or too easy may reduce its construct validity, as its items do not accurately measure the knowledge or skills that it is intended to assess [8]. Given this concern, the purpose of the test and the effects of its use should be taken into consideration to determine what type of knowledge and skills should be measured. When the Vietnamese High School Graduation English Test (VGET) is used with dual purpose for certification and selection, the difficulty of its reading texts, an important component of the test, must be carefully calibrated to align with these objectives. If the reading texts contain too many academic and complex structures and vocabulary, it may fail to measure students' true language ability, resulting in invalid score interpretations and potentially negative consequences [8]. One of them is the effect on the fairness of the test. As stated in Willingham and Cole's framework, one characteristic of test fairness is the comparability or equality in outcomes of learning and opportunities to learn [9]. Therefore, when a high-stakes test contains a level of difficulty that exceeds the learning outcome specified in the curriculum, it may disproportionately disadvantage students in rural or underserved areas, who have fewer opportunities to access the relevant knowledge compared to their peers in major cities.

To evaluate the difficulty of reading passages in the high-stakes English tests, several key factors based on the theoretical background can be utilized, including readability and lexical demand.

2.3. Readability

Readability refers to the extent to which a text can be comprehended, read at an optimal pace, and perceived as interesting by its readers [10]. After the introduction of multiple readability formulas, readability can be defined as mathematical equations designed to determine or predict the level of reading competence necessary for comprehending a given text [11]. Among hundreds of formulas, traditional formulas, Flesch Reading Ease (FRE) and Flesch-Kincaid Grade Level (FKGL), are the two most popular ones [12]. These formulas mainly focused on syntactic features such as word length and sentence length as they are based on the assumptions that texts with longer words and lengthier sentences require more time to access and interpret, making them more challenging to understand [13].

The formula for the Flesch Reading Ease Score given by Flesch is:

$$FRES = 206.835 - (1.015 \times \text{Average Sentence Length}) - (84.600 \times \text{Average Syllables per Word})$$

The Flesch-Kincaid Grade Level (FKGL), originally developed by J. Peter Kincaid for the U.S. Navy in 1976, relates a U.S. school grade level to a text. The formula is expressed as:

$$FKGL = (0.39 \times \text{Average Sentence Length}) + (11.8 \times \text{Average Syllables per Word}) - 15.59$$

Despite some criticism for their ignorance of some cognitive features, they still found big success and have been applied extensively to match reading materials with appropriate proficiency levels in various fields, not only in education but also in journalism, industry and military. The extensive use of these formulas is further evidenced by their integration into popular word processing programs and services such as KWord, IBM Lotus Symphony, Microsoft Office Word, WordPerfect, WordPro, and Grammarly.

The traditional readability assessment primarily aims at predicting reading difficulty for native speakers. However, when L2 learners are considered, some other features have emerged. Beyond syntactic complexity, lexical features are proved to have substantial positive impact on readability classification [14].

2.4. Lexical demand

Vocabulary knowledge, in particular, plays a crucial role in reading comprehension performance and overall language proficiency [15]. The idea that a reader needs to know a certain proportion of words in a text in order to comprehend it is defined as lexical demand by Nation [16]. When there are too many unknown words in the text, the reader cannot comprehend it properly, therefore lexical demand of a text

also contributes to its difficulty. The theoretical framework to identify the lexical demand of texts is based on Hu and Nation's lexical coverage model [16]. According to this model, 95% is the minimal threshold for adequate comprehension, and 98% is the ideal threshold for unsupported reading and optimal comprehension.

To determine the difficulty level of texts, word-frequency levels are taken into account. If a word belongs to a lower level, the word is less frequent and therefore more difficult [15]. When the vocabulary demand falls at a lower frequency level, it means that the text is more difficult to understand. It is therefore significant to identify text difficulty through the vocabulary level and lexical coverage to inform pedagogical interventions in reading instruction. Among various word-frequency lists, this study adopted Nation's British National Corpus/Corpus of Contemporary American English 25-level (BNC/COCA) [17] relied on word family, which is widely recognized as the most comprehensive and frequently used by lexical profiling research.

To establish a benchmark for the lexical demand in the analyzed texts, the current study interprets lexical coverage in relation to language proficiency, drawing on the established correspondences between the BNC/COCA 25-level word-frequency lists and the Common European Framework of Reference for Languages (CEFR). Nation's frequency framework provides a principled basis for estimating vocabulary size through word families, with each BNC/COCA level corresponding to a distinct proficiency band [18]. For example, learners at the B1 level generally possess 2,000-3,000 high-frequency word families; the B2 level encompasses about 4,000 word families; and the C1 and C2 levels require roughly 5,000-9,000-word families.

3. Methodology

3.1. Subject of the study

A corpus of 22 reading comprehension passages from the VGET administered between 2015 and 2025 (11 academic years) was selected for analysis. Although the VGET includes multiple test codes each year, this study utilized one randomly selected code from each academic year. It should be noted that all test versions within a single administration are developed based on a standardized test matrix and designed to be equivalent in terms of content complexity and difficulty level. Consequently, the selection of a single code per year is considered sufficient to represent the linguistic characteristics of the entire examination for that period. The tests were sourced from publicly available documents online (in .docx format). Only the two main reading comprehension passages from each selected exam were extracted and converted into plain text (.txt) files for corpus processing. The texts were subsequently divided into two distinct sub-corpora, Text 1 and Text 2, based on their sequential order within the original examination paper.

3.2. Instruments

For the analysis of text readability, the Flesch Reading Ease Score (FRES) and Flesch-Kincaid Grade Level (FKGL) were calculated using the specialized software Readability Formulas [19]. This tool was selected primarily for its accessibility and reliability, which allows users to analyze texts of unlimited length in its free version.

To determine the lexical coverage of the reading texts, AntWordProfiler (version 2.2.1) [20] was utilized. The adopted reference list was the BNC/COCA 25-level word family list, supplemented by four auxiliary lists: Proper Nouns (PNs), Marginal Words (MWs), Transparent compounds (TCs), and Acronyms. AntWordProfiler was chosen as it quantifies the percentage of word family coverage across different vocabulary levels, which is central to the aims of this research.

3.3. Analysis procedures

The study comprised two phases. In phase 1, the reading texts were profiled using Readability Formulas to calculate their FRES and FKGL indices. The resulting data were then analyzed descriptively using Microsoft Excel to calculate the mean and standard deviation. This phase aims at providing an overall trend of the texts' readability, focusing on syntactic features such as sentence length and word length. Subsequently, Phase 2 involved analyzing the corpora with AntWordProfiler (version 2.2.1). The aim was to identify the number of word families (the lexical demand) that a learner would need to attain to achieve the established comprehension coverage thresholds (i.e., 95% and 98% coverage).

4. Result and discussion

4.1. Readability scores in the VGET reading comprehension texts

In response to RQ1, this section presents the analysis of the Flesch Reading Ease Score (FRES) and Flesch-Kincaid Grade Level (FKGL) of the VGET reading comprehension texts. In the Flesch Reading Ease index, lower scores indicate material that is more difficult to read. Meanwhile, the Flesch-Kincaid Grade Level aligns the text with the corresponding grade level in the United States school education system. Thus, the higher the score the text has, the higher education attainment it corresponds to.

Table 1. *Flesch Reading Ease Score and Flesch-Kincaid Grade Level of reading passages in 2015-2025 VGETs*

Year	Flesch Reading Ease Score (FRES)		Flesch-Kincaid Grade Level (FKGL)	
	Text 1	Text 2	Text 1	Text 2
2015	48.00	32.00	11.40	13.80
2016	61.00	22.00	9.14	16.23
2017	69.00	42.00	7.84	12.21
2018	56.00	55.00	9.62	10.80
2019	53.00	43.00	9.95	12.42
2020	61.00	65.00	9.25	8.80
2021	64.00	62.00	9.17	9.70
2022	60.00	85.00	9.66	5.39
2023	59.00	52.00	9.67	12.30
2024	58.00	54.00	9.06	10.90
2025	56.00	30.00	9.70	13.74

As can be seen from Table 1, most VGET reading passages scored between 50 and 70 on the FRES scale and were appropriate for 9th to 10th grade students in the U.S educational system. This finding is consistent with those of the Korean College Scholastic Ability Test (CSAT), where passages show similar readability levels [21]. This suggests that texts generally fall within the "Fairly Difficult" to "Standard" range, according to Flesch's style description. On the other hand, some reading passages had FRES below 50, fitting students at 13th to 16th grade levels (in 2015, 2016, 2017, 2019, 2025). These texts appear to be too challenging with long and complex sentence structures and vocabulary, which may exceed the reading competence of upper secondary students. Therefore, they might not be able to properly determine a student's levels of achievement in reference to the learning outcome at the B1 level. Notably, this trend is likely to continue in the coming year, after an abrupt increase in the difficulty level of 2025 texts.

Table 2. *Comparison of the average readability levels of Text 1 and Text 2 in 2015-2025 VGETs' Mean (M) and Standard Deviation (SD)*

Index	Text 1		Text 2	
	Mean	**SD**	**Mean**	**SD**
Flesch Reading Ease Score (FRES)	58.64	5.55	49.27	18.04
Flesch-Kincaid Grade Level (FKGL)	9.50	0.85	11.48	2.89

From the descriptive statistics (Table 2), it is evident that the average FRES of Text 1 (*M*=58.64) was considerably higher than that of Text 2 (*M*=49.27). This means that the VGET is often designed with two texts at different levels of readability, which is a key to differentiate learners' ability, serving the higher education selection purpose. This finding suggests that test-takers should read the passages in the order in which they appear in the test paper, as their difficulty level increases sequentially. The low standard deviation for Text 1 (*SD* = 5.55) indicates a high level of consistency across the years. In contrast, the considerably larger standard deviation for Text 2 (*SD*=18.04) suggests a substantial inconsistency in its difficulty level across the different examination years.

4.2. Lexical demand of the VGET reading texts

This section examines the lexical coverage of the corpus and determines the word family size needed by learners to achieve the two established comprehension thresholds. A summary of the cutting point of the cumulative coverage was demonstrated in Table 3.

Table 3 demonstrates the lexical profile of the reading texts based on the BNC/COCA wordlists, including the supplementary lists: Proper nouns, Marginal words, Transparent compounds, Acronyms and off-list words. It can be observed from the collected data that these supplementary lists are essential for reaching the 95% and 98% comprehension threshold. Out of the 22 texts analyzed, 16 texts required a vocabulary size of 2,000-3,000-word families for 95% lexical coverage, with 9 of these texts also reaching the 98% threshold at this level. This means that students are required to know a minimum of 3,000 word families to get adequate comprehension of the VGET reading passages. This vocabulary size, approximately equivalent to the CEFR B1 level, suggests that these 16 texts are suitable for determining whether students have achieved the expected learning outcomes at the end of upper-secondary education. However, for the remaining texts, particularly Text 2 items from 2015, 2016, 2017, 2018, and 2025, this K-3 vocabulary load is demonstrably insufficient. To attain 95% coverage, learners would need a vocabulary size of approximately 3,000-5,000 word families, and to reach 98% coverage, 6,000-10,000 or even 14,000-word family levels were required. These lexical demands are apparently beyond the expected level of the average upper-secondary students, as they correspond to the C1-C2 levels of the CEFR. The excessively high lexical demands might cause unnecessary pressure on test-takers, especially within a limited amount of time. However, in a vision to make English the second language, it is expected that students can become autonomous readers. At the tertiary level, they need to be able to use English efficiently for learning, researching, completing assignments and projects; therefore, they need sufficient proficiency to read academic materials. Consequently, the Vietnamese Graduation English Test faces a fundamental challenge: while the test assesses basic outcomes (B1), its selective function for tertiary education necessitates assessing students' capability to engage with academic vocabulary and read without assistance, which inherently demands a higher level of text difficulty. Previous studies indicate that a 8,000-9,000 word-family vocabulary is essential for covering 98% of written texts [22], and a minimum of 5,000-word level for reaching the 95% line [23]. In Malaysia, test reading texts require at least a 6,000-word-family threshold, which is deemed appropriate for selecting university entrants [24]. These findings highlight a paradoxical situation in the current VGET texts, where the assessment simultaneously attempts to meet two conflicting objectives. This suggests that separating the two purposes of the exam could enhance both its validity and fairness for all test-takers.

Table 3. *Summary of the Cutting Point of the Cumulative Coverage for VGET Reading Texts*

Year	Text 1		Text 2	
	WFL-95%	**WFL-98%**	**WFL-95%**	**WFL-98%**
2015	K-5 (96.55)	K-6 (98.74)	K-3 (96.59)	K-6 (98.18)
2016	K-6 (95.37)	K-10 (98.85)	K-3 (95.11)	K-6 (98.38)
2017	K-2 (96.49)	K-4 (98.72)	K-4 (96.53)	K-5 (98.51)
2018	K-4 (97.22)	K-5 (98.62)	K-3 (94.66)	K-7 (98.48)
2019	K-3 (98.56)	K-3 (98.56)	K-3 (94.20)	K-4 (97.98)
2020	K-3 (99.63)	K-3 (99.63)	K-3 (94.30)	K-7 (98.71)
2021	K-2 (99.36)	K-2 (99.36)	K-3 (98.05)	K-3 (98.05)
2022	K-3 (95.22)	K-5 (98.70)	K-1 (95.71)	K-3 (99.01)
2023	K-3 (99.61)	K-3 (99.61)	K-2 (96.32)	K-3 (99.08)
2024	K-2 (95.58)	K-3 (100)	K-3 (97.97)	K-3 (97.97)
2025	K-4 (95.48)	K-6 (98.25)	K-5 (95.42)	K-14 (98.26)

Note. K= '000, WFL= Word Frequency Level

The noticeable increases in lexical demand, particularly in 2018 and 2025, are closely linked to shifts in the national assessment context. The 2018 examination followed a year of relatively high scores in 2017, which had prompted test developers to enhance the test's discriminative power for university selection by introducing more complex reading passages. For the 2025 test, the surge in complexity reflects a strategic alignment with the 2018 General Education Curriculum and the Prime Minister's recent scheme to transition English toward a second language in schools.

These policy-driven adjustments, however, are consistent with previous evaluations of the Vietnamese pedagogical landscape, which highlight a systemic gap. Specifically, research on the new series of high-school English textbooks in Vietnam revealed that reading passages are often overloaded with novel words that lack importance for comprehension and rarely reoccur to facilitate acquisition [25]. When students are primarily exposed to instructional materials that already pose significant lexical challenges with limited contextual clues, the even higher linguistic thresholds found in the 2018 and 2025 VGETs create an overwhelming barrier. This suggests that neither the pedagogical input nor the high-stakes assessment is yet fully aligned with the realistic vocabulary gain or the proficiency levels of the majority of learners.

5. Conclusion and Implication

5.1. Conclusion

The results of this study offer a deeper understanding of the difficulty level of the VGET texts, particularly in terms of linguistic and lexical demand. The analysis confirms a potential mismatch between the target language proficiency level (B1) and the actual difficulty of the reading texts, which highlights the challenges in balancing between the two purposes of the current High School Graduation Test. The present study stresses on the importance of utilizing technical analysis in test design and emphasizes the significance of textual features and vocabulary knowledge in determining reading comprehension difficulty. Acknowledging the structural challenges revealed in these high-stakes assessments, the study concludes that immediate modifications in test administration by developers and strategic adjustments in preparation by teachers and students are crucial.

5.2. Implication

The implications of this study emphasize the urgent need for a more coherent alignment between curriculum design, textbook development, and national assessment. Given that previous research on the new series of English-language textbooks for Vietnamese high school students found reading passages to be overloaded with novel words that lack sufficient recurrence and contextual clues [25], test developers must be cautious not to worsen the existing gap in teaching materials by selecting reading passages that significantly exceed the B1 threshold. To make the reading tests more relevant to the national context, it is recommended that assessment tasks prioritize the "usefulness" of vocabulary rather than mere "novelty". Instead of introducing an excessive number of low-frequency words that may not have been adequately reinforced in classroom settings, test designers should focus on evaluating how students utilize high-frequency word families in authentic contexts.

Test developers are strongly recommended to recognize the necessity of administering two distinct tests that serve different purposes, as this approach can enhance validity, fairness, and the quality of university admissions. While ensuring the examination aligns with the specified B1 proficiency level in the 2018 English Language Curriculum remains critical, strengthening overall assessment practices is equally essential. This reform is especially vital given the nation's vision for higher education to become a leading pillar in the development of the English language ecosystem. Furthermore, the employment of computational tools and technical analyses is essential to establish official datasets and objective guidelines for test construction, thereby reducing potential negative washback.

The complexity of the VGET's reading texts also highlights the need to improve reading skill instruction at school. Teachers may consider the various learning strategies as well as add extra materials to aid reading comprehension. First, students should be provided with authentic reading materials such as journals, newspapers, or narratives to be familiar with long and complex texts. Besides, expanding vocabulary is equally significant. Since deliberate vocabulary teaching at class is deficient due to time constraints, teachers should promote incidental vocabulary learning and student autonomy to effectively broaden their vocabulary size. One method is promoting the use of graded readers, which are designed to expose students to vocabulary up to the 8,000 word-frequency level. Furthermore, students should not learn vocabulary items in isolation but rather in word families, including their associations, derivations, inflections, synonyms, antonyms, and collocations. Last but not least, teachers would need to provide opportunities for students to meet the target words repeatedly to enhance their learning and help learners develop the fluency in using these words. For example, organizing productive activities or using flashcards can increase exposure to vocabulary and foster vocabulary retention. In summary, strategic selection of diverse teaching materials and instructional techniques is paramount to effectively enhance students' reading proficiency.

REFERENCES

[1] Ministry of Education and Training, *Circular on promulgating the six-level foreign language proficiency framework for Vietnam*, No. 01/2014/TT-BGDDT, 2014.

[2] Ministry of Education and Training, *Circular on English language curriculum*, No.32/2018/TT-BGDDT, 2018.

[3] Prime Minister, *Decision on Approving the Scheme on introducing English as a second language in schools (2025-2035, Vision to 2045),* No.231/QĐ-TTg, 2025.

[4] E. Rafatbakhsh and A. Ahmadi, "Predicting the difficulty of EFL reading comprehension tests based on linguistic indices", *Asian-Pacific Journal of Second Foreign Language Education,* vol. 8, no. 1, p. 41, 2023. doi: 10.1186/s40862-023-00214-4.

[5] Ministry of Education and Training, *Decision on Approval of the national high school examination from 2015,* No. 3538/QĐ-BGDĐT, 2014.

[6] Ministry of Education and Training, *Dispatch on Guidelines for implementing the pilot English curriculum at secondary level under the Project "Teaching and learning foreign languages in the national education system, 2020"*, No. 7972/BGDĐT-GDTrH, 2013.

[7] A. Hughes, *Testing for Language Teachers, 2nd edition.* Cambridge: Cambridge University Press, 2010. doi: 10.1017/CBO9780511732980.

[8] S. Messick, "Validity", *ETS Research Report Series*, vol. 1987, no. 2, 1987. doi: 10.1002/j.2330-8516.1987.tb00244.x.

[9] W. W. Willingham and N. S. Cole, G*ender and Fair Assessment. Mahwah,* NJ: Lawrence Erlbaum Associates, 1997.

[10] J. S. Chall, *Readability Revisited : the New Dale-Chall Readability Formula.* Cambridge, Mass. : Brookline Books, 1995.

[11] R. Akbari, M. R. Atae, and H. Marefat, "The role of discourse elements in determining the readability of texts", *TESL Report*., vol. 32, no. 1, pp. 16–27, 1999.

[12] G. R. Klare, "The measurement of readability: Useful information for communicators", *ACM Journal of Computer Documentation (JCD),* vol. 24, no. 3, pp. 107–121, 2000. doi: 10.1145/344599.344630.

[13] 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. doi: 10.3758/BF03195564.

[14] M. Xia, E. Kochmar, and T. Briscoe, "Text readability assessment for second language learners", *in Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications, San Diego, CA: Association for Computational Linguistics,* 2016, pp. 12–22. doi: 10.18653/v1/W16-0502.

[15] I. S. P. Nation, L*earning Vocabulary in another Language, 3rd edition*. Cambridge: Cambridge University Press, 2022.

[16] M. H. Hsueh-chao and P. Nation, "Unknown vocabulary density and reading comprehension", *Reading in a Foreign Language*, vol. 13, no. 1, pp. 403–430, 2000. doi: 10.64152/10125/66973.

[17] I. S. P. Nation, "The BNC/COCA level 6 word family lists", *wgtn.ac.nz*, 2017. [Online]. Available: https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs [Assessed October 26, 2025].

[18] I. S. P. Nation, "Vocabulary, the CEFR levels, and word family size", *wgtn.ac.nz.* [Online]. Available: https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists [Assessed October 23, 2025].

[19] S. Brian, "Readability scoring system PLUS v3", *ReadabilityFormulas.com*, 2024. [Online]. Available: https://readabilityformulas.com/readability-scoring-system.php [Accessed October 20, 2025].

[20] L. Anthony, "AntWordProfiler (Version 2.2.1)", *laurenceanthony.net*, 2024. [Online]. Available: https://laurenceanthony.net/software/antwordprofiler/ [Accessed November 01, 2025].

[21] Y. Kim and J. H. Ma, "Text readability of the college scholastic ability test and high school English II textbooks", *English Teaching*, vol. 67, no. 4, pp. 195–218, 2012. doi: 10.15858/engtea.67.4.201212.195.

[22] I. Nation, "How large a vocabulary is needed for reading and listening?", *Canadian Modern Language Review*, vol. 63, no. 1, pp. 59–82, 2006. doi: 10.3138/cmlr.63.1.59.

[23] B. Laufer, "*What percentage of text-lexis is essential for comprehension"*, in Special Language: From Humans Thinking to Thinking Machines, C. Lauren & M. Nordman, Multilingual Matters, 1989, pp. 316–323.

[24] M. Benedict, "Word frequency level and lexical coverage in the reading comprehension texts of the Malaysian university English test", *PASAA,* vol. 61, no. 1, pp. 33–60, 2021. doi: 10.58837/CHULA.PASAA.61.1.2.

[25] C.-D. Nguyen, "Lexical features of reading passages in English-language textbooks for Vietnamese high-school students: Do they foster both content and vocabulary gain?," *RELC Journal*, vol. 52, no. 3, pp. 509–522, 2021, doi: 10.1177/0033688219895045.