

EXAMINING CEFR READING COMPREHENSION TEXT COMPLEXITY IN VIETNAM'S 2025 NATIONAL ENGLISH EXAM

Nguyen Thi Khanh An^{1*}, Tran Trong Thanh²

¹*University of Languages and International Studies, Vietnam National University, Hanoi, Vietnam*

²*East Asia University of Technology, Hanoi, Vietnam*

*Corresponding author: Nguyen Thi Khanh An, e-mail: khanhan@vnu.edu.vn

Received: March 9, 2026. Revised: April 14, 2026. Accepted: June 9, 2026.

Abstract. This study examines the complexity of reading comprehension passages in the English component of the 2025 National High School Graduation Exam. It estimates the CEFR-aligned difficulty levels of passages from the 2020-2025 examinations using quantitative measures and qualitative CEFR (Common European Framework of Reference for Languages) descriptors and compares them against the B1 benchmark specified in the national curriculum. The findings suggest a substantial increase in difficulty over time, with passages in the 2025 examination exhibiting complexity characteristics associated with C1-C2 levels in terms of lexical and readability demands. These levels considerably exceed the B1 requirement outlined in the 2018 General Education Curriculum. The results suggest a growing mismatch between curriculum standards and examination demands, raising concerns about validity, equity, and fairness. The study contributes to international discussions on high-stakes language testing by highlighting the importance of curriculum-assessment alignment and transparent test design in promoting pedagogical coherence and educational justice.

Keywords: curriculum-assessment alignment, text complexity, high-stakes assessment, CEFR, language policy

1. Introduction

Public concern has been raised about the difficulty of Vietnam's 2025 National High School Graduation English Exam (NHSGE) (Do, 2025). Although the exam format followed the publicly released sample, the reading passages were reported to include unfamiliar topics, a high density of difficult vocabulary, and complex or confusing content, making them challenging for students (Thanh Niên, 2025). While these concerns highlight potential mismatches between test difficulty, expected proficiency levels, and curriculum alignment, they are largely based on perceptions, with limited empirical evidence on the actual complexity of NHSGE reading texts in relation to CEFR benchmarks.

This study examines whether the 2025 exam assessed the intended reading proficiency or presented material beyond the expected level and curriculum coverage. Specifically, it analyses the difficulty of the 2025 reading passages and compares them with those from 2020-2024 and against the proficiency expectations specified in the official Ministry of Education and Training (MOET) curriculum. The NHSGE English test is a large-scale, high-stakes assessment administered annually in Vietnam, serving both to certify high school completion and as a key pathway to university admission. With a large number of candidates each year (Hoang Trang & Duc Tuan, 2025), the exam plays a major gatekeeping role. Consequently, even small changes in exam design can significantly influence classroom practices, student motivation, and perceptions of fairness. To address this, the study analyzes the complexity of reading passages in NHSGE papers from 2020-2025 using quantitative indices and qualitative CEFR-aligned descriptors, to

estimate their CEFR-aligned difficulty levels and compare them with the B1 benchmark specified in the national curriculum. Rather than directly determining curriculum-assessment alignment, text complexity is treated as an indicator of potential (mis)alignment with the intended proficiency level.

2. Background

2.1. Policy intentions and curriculum expectations

English language instruction in Vietnam has changed significantly in recent years. The MOET's endeavor to align English instruction and assessment with globally accepted standards, especially the Common European Framework of Reference for Languages (CEFR), is at the center of these changes. Clear expectations were outlined in the 2018 General Education Curriculum issued under Circular 32/2018/TT-BGDĐT. Students are expected to achieve Level 3 of the Vietnamese Six-Level Foreign Language Proficiency Framework by the end of upper secondary education, which is officially in line with CEFR B1 (MOET, 2018). According to the policy, a B1-level student should handle common real-life situations and understand main ideas in familiar contexts (MOET, 2014). Therefore, the curriculum stresses functional and communicative abilities rather than just structural knowledge. The current approved textbook series support these goals by developing real-world communication while reinforcing B1-level vocabulary and grammar.

Ensuring that the English NHSGE aligns with curriculum objectives is challenging. Misalignment can lead to negative washback, where teachers focus on test strategies and memorization instead of communicative language skills (Kuang, 2020). To address this, the MOET proposed revising the exam to better assess students' ability to use English in real-life contexts and to move toward a competency-based approach consistent with communicative language testing and principles of authenticity in language assessment (Liao et al., 2023).

2.2. Text complexity: Indicators and interpretation

Text complexity is a multifaceted concept that includes lexical, syntactic, and discourse-level elements that affect how a text is interpreted and processed (Spencer et al., 2018). Text complexity encompasses deeper elements, such as cohesiveness, organization, and knowledge demands in addition to surface-level linguistic traits (Fisher et al., 2012). Different parts of this construct are captured by the quantitative indicators utilised in this study: Lexile measurements provide an overall estimate of difficulty based on word frequency and syntactic properties, while readability indices like Flesch Reading Ease primarily reflect sentence length and word length. In turn, vocabulary profiling provides information about the distribution of CEFR levels within a text as well as lexical sophistication. Accordingly, the interpretation of alignment in this study is predicated on contrasting the expected CEFR B1 level indicated in the curriculum with the observed complexity of reading texts as expressed through several indicators. Alignment is inferred from the convergence of quantitative and qualitative data rather than from a single measure, using methods that integrate multiple instruments to assess CEFR-related text difficulty (Natova, 2019).

2.3. Recent changes in exam format

Recent modifications to the exam's structure and regulations have made the issue more complex. Beginning in 2025, MOET has made significant changes to the NHSGE format in accordance with the 2018 curriculum overhaul. English used to be one of three compulsory examination subjects but is now optional. Students are required to take only Literature and Mathematics as compulsory subjects, and they may choose two other subjects from a list that includes history, foreign languages, and several science and technology disciplines (MOET, 2024a). In addition to subject selection changes, the exam format has shifted. In particular, the following are the main changes for 2025 (MOET, 2024a; MOET, 2024b): (1) a mix of item types,

such as multiple choice, matching, and short-answer questions (and in the case of English, only multiple choice questions are employed); (2) a more competency-based evaluation approach that seeks to evaluate higher-order thinking skills instead of rote recall; and (3) new scoring guidelines and marking schemes, which include the elimination of the automatic conversion of foreign language certificates to 10/10 scores. These changes are part of the efforts to improve alignment with the new curriculum and Resolution 29 (MOET, 2013) on comprehensive education reform, as stated by MOET. However, the importance of English in the classroom may be compromised if it is no longer emphasized as a necessary subject, especially in institutions where students are motivated primarily by exams. Furthermore, these changes also raise an important equity concern. Students in rural or non-specialized schools, who rely mainly on classroom instruction, could be disadvantaged compared to students in urban or specialized schools with better support, leading to an unequal testing environment that contradicts the curriculum's goal of educational equity.

2.4. The central role of Reading in the exam

While the exam format has undergone several adjustments this year, reading comprehension remains of central importance. In the 2024 exam, reading comprehension accounted for 12 out of 50 questions, whereas in the 2025 version, it increased to 18 out of 40 questions, highlighting a notable rise in its proportion within the overall test. Reading is a key component of language learning, and reading comprehension tasks can assess important aspects of students' language proficiency. Reading comprehension involves understanding word meanings and sentence structure to build coherence, and deeper comprehension occurs when readers connect the text with their prior knowledge and experiences (Van Dijk & Kintsch, 1983, as cited in Choi & Zhang, 2018). To understand a text, readers must possess a range of skills and knowledge such as word recognition, linguistic knowledge, reading strategies, inferring, etc. (Grabe, 2014). Given that reading can assess these diverse aspects of language proficiency, it has been a key component and consistently included in the NHSGE.

Reading comprehension in the NHSGE is assessed through formal measures. While formal measures serve important summative and comparative functions, concerns regarding their potential to oversimplify reading comprehension remain as standardized test formats may encourage strategic reading rather than deep comprehension (Farr et al., 1990). Nonetheless, reading sections in the NHSGE continue to be in a standardized format, reflecting the prevailing approach to large-scale assessment in Vietnam. Therefore, the purpose of this study is to provide empirical evidence on the difficulty of reading passages in the NHSGE and their alignment with curricular expectations. The overarching aim of this study is to examine the complexity of reading passages in selected NHSGE papers from 2020 to 2025 and to explore their alignment with the CEFR B1 level specified in the national curriculum. To achieve this aim, the study pursues the following objectives: to analyze the complexity of NHSGE reading passages using quantitative measures and qualitative CEFR-aligned descriptors, and to compare the complexity of reading passages in 2025 with those from 2020 to 2024 in relation to the CEFR B1 target.

Accordingly, the study addresses the following research questions:

- (1) What is the CEFR-aligned complexity level of reading passages in selected NHSGE papers from 2020 to 2025?
- (2) How does the complexity of reading passages in the 2025 NHSGE compare to those from 2020 to 2024 in relation to the CEFR B1 level?

3. Methodology

3.1. Research design and data collection

To answer the questions, the study adopts a corpus-based approach, analyzing the reading texts from six consecutive years of the NHSGE (2020-2025). Each year's reading section is

examined individually using both qualitative CEFR-aligned descriptors and quantitative measures, before being compared to identify changes in reading text complexity. CEFR levels were assessed using Natova's proposal, in which Lexile measures are mapped to CEFR levels using published correspondence ranges and overall CEFR judgments are derived through a triangulation model combining Lexile measures, Flesch Reading Ease scores, vocabulary profiles, and qualitative text-complexity assessments (Natova, 2019). The exam papers were publicly released after each test administration and were retrieved from online sources for analysis. Although each year's exam includes multiple test versions, they are designed to be equivalent in terms of difficulty level and topic coverage. The exam codes and selected passages analyzed in this study were as follows: 2020 (Code 402; Passage 1: Questions 39-43, Passage 2: Questions 44-50), 2021 (Code 401; Passage 1: Questions 31-35, Passage 2: Questions 36-42), 2022 (Code 405; Passage 1: Questions 31-35, Passage 2: Questions 36-42), 2023 (Code 401; Passage 1: Questions 39-43, Passage 2: Questions 44-50), 2024 (Code 401; Passage 1: Questions 31-35, Passage 2: Questions 36-42), and 2025 (Code 1102; Passage 1: Questions 1-8, Passage 2: Questions 20-29).

This study employs a purposive sampling approach, selecting official NHSGE reading tests as they represent high-stakes, standardized assessments aligned with CEFR benchmarks. The selected timeframe (2020-2025) is intentional, with 2025 marking the first implementation of the new curriculum (MOET, 2024a; MOET, 2024b). The inclusion of the immediately preceding years allows for a comparative perspective to identify potential shifts in text complexity. Although multiple test versions are administered each year and are designed to be comparable in content and difficulty, one version per year was selected for analysis. Consequently, the findings should be interpreted as indicative rather than fully representative of all NHSGE versions. This sampling decision may limit the generalizability of the findings and is acknowledged as a limitation of the study.

3.2. Text analysis criteria and tools

The tools for assessing the CEFR level of texts in this study follow Natova (2019)'s proposal. Regarding qualitative data, reading texts were analyzed based on four features of text complexity (Meaning/Purpose, Structure, Grammatical Complexity and Knowledge Demands) specified for each level. To determine the CEFR levels of grammatical categories in the reading texts, the database of the free online resource associated with the CEFR, English Grammar Profile, was used (EnglishProfile, 2025a), which maps grammatical forms to CEFR levels. For quantitative data, several complementary indicators of text complexity were employed. The Lexile Text Analyzer (Lexile, 2025) estimates overall reading difficulty based on word frequency and sentence length. The Flesch Reading Ease reflects surface-level readability, focusing on sentence and word length. Vocabulary profiling was performed using the Text Inspector (EnglishProfile, 2025b), which provides information on word frequency bands and CEFR levels, indicating lexical sophistication, while type-token ratio (TTR) measures lexical diversity.

3.3. Procedure

The study analysed NHSGE exams from 2020 to 2025, focusing on reading texts; cloze tasks were excluded as they are not continuous texts. Each text was examined using both qualitative and quantitative criteria. Qualitative analysis followed Natova (2019), covering four dimensions: meaning/purpose, structure, grammatical complexity, and knowledge demands. Meaning and purpose were identified based on communicative function, level of abstraction, and clarity. Structure was analysed in terms of organisation, coherence, sequencing, and flow. Grammatical features were identified and mapped to CEFR levels using the English Grammar Profile. Knowledge demands were assessed based on topic familiarity, required background knowledge, specialised vocabulary, and cultural/academic content. Overall CEFR levels were then estimated

by aligning these features with CEFR descriptors. To enhance analytical reliability, selected passages were independently reviewed by a second researcher using the same CEFR-aligned criteria. Any discrepancies in interpretation were discussed until consensus was reached. For quantitative analysis, Lexile Analyzer was used to estimate reading difficulty and map scores to CEFR levels. The Flesch Reading Ease index (calculated in R) measured sentence and word length complexity, and Text Inspector was used for vocabulary profiling to determine the distribution of word levels.

4. Results and discussion

4.1. Results

Using both quantitative metrics and a qualitative approach, we assessed the CEFR-aligned difficulty of NHSGE reading passages from 2020 to 2025. A summary of the findings is given below, and Table 1 contains specific statistics.

Table 1. CEFR Reading Comprehension Text Complexity

Year	Text no.	Word count	Quantitative assessment of CEFR reading text complexity					Qualitative assessment of CEFR reading text complexity				Estimated overall CEFR level
			Lexile measure		Flesch reading ease		Vocabulary profile	P	S	G	K	
			Lexile units	CEFR level	Index	CEFR level	Type-token ratio					
2025	1	286	1200L-1400L	C1+	52.62973	C1	0.72	C1	C1	B2+	C1	C1
	2	343	1210L-1400L	C1+	28.44123	C2	0.76	C1+	C1	B2+	C1	C1+
2024	1	251	610L - 800L	B1	72.86954	B1	0.54	B1	B1	B1	B1	B1
	2	361	1210L -1400L	C1	51.19727	C1	0.60	C1	C1	B1	C1	C1
2023	1	255	1010L -1200L	B2+	58.15654	C1	0.65	C1	C1	B2	B2	C1
	2	325	1210L -1400L	C1+	49.35889	C2	0.57	B2+	C1	B2	B2+	C1
2022	1	225	1010L -1200L	B2+	60.79227	B2	0.63	A2+	A2+	B1	A2+	B2
	2	303	610L - 800L	A2+	84.54941	A2	0.49	B2	B2	B1	B1+	B1
2021	1	155	1010L -1200L	B2+	60.46194	C1	0.60	A2+	A2	A2+	A2	B2
	2	298	1010L -1200L	B2+	61.02118	C1	0.49	C1	C1	B2	B2	B2
2020	1	284	1210L -1400L	C1	49.10794	C2	0.61	B2	B2+	B1	B1	B2
	2	389	810L - 1000L	B1+	62.12631	B2	0.56	B2	B2	B1+	B2	B2

Note: P = Meaning/Purpose; S = Structure; G = Grammatical Complexity; K = Knowledge Demands. Qualitative CEFR descriptors and Lexile-CEFR correspondence ranges were adapted from Natova (2019).

4.1.1. Quantitative analysis

Across the six-year period (2020-2025), the overall word count of NHSGE reading passages remained relatively stable (155-389 words). However, other indicators reveal a marked increase in text complexity over time (see Table 1).

Each quantitative metric captures a different aspect of text complexity. Flesch Reading Ease mainly measures surface readability through sentence and word length, Lexile scores reflect overall processing difficulty based on word frequency and sentence length, and vocabulary profiling (including TTR) reflects lexical sophistication and diversity rather than overall difficulty. In the earlier exams (2020-2022), Lexile scores ranged from 610L to 1210L, corresponding to CEFR levels between A2+ and C1, and Flesch Reading Ease scores were moderate (49-62), indicating moderately challenging passages. Lexical diversity was also lower

during this period, with type-token ratios (TTR) ranging from 0.49 to 0.63. From 2023 onward, passages became significantly harder, with Lexile scores consistently above 1010L and most passages at C1 level, while Flesch scores dropped below 59. This trend reaches its peak in 2025, where both passages fall within 1200L-1400L (C1+), include the lowest Flesch score observed (28.44, corresponding to C2), and show the highest lexical diversity (0.72-0.76). Overall, these measures suggest that 2025 may be the most linguistically demanding exam year in the dataset, in terms of lexical and readability indicators. Regarding Research Question 1, these findings show that the CEFR-aligned complexity of the 2025 texts is primarily at C1 level or higher. Regarding Research Question 2, the quantitative data demonstrate that the 2025 texts routinely surpass the curriculum's B1 threshold and are more challenging than most texts from 2020-2024.

4.1.2. Qualitative analysis

Analysis of NHSGE reading texts (2020-2025) shows considerable variation across years, with the 2025 passages standing out as consistently demanding across both quantitative and qualitative indicators. Using Natova's (2019) four qualitative dimensions, passages were classified by CEFR level, showing both within-year variation and a steady increase over time. In 2020-2021, most texts remained within the B1-B2 range, broadly aligned with curriculum expectations. From 2022 onward, difficulty increased, with greater variation across passages. By 2023-2024, several texts reached B2-C1 levels, particularly in Structure and Meaning/Purpose, despite some internal inconsistency. The 2025 exam was the most demanding, with both passages reaching B2+-C1 across dimensions and higher conceptual density, especially in Meaning/Purpose and Knowledge Demands. Overall, the findings indicate a growing divergence from the CEFR B1 target, with texts increasingly reflecting upper B2-C1 requirements. Regarding the research questions, the 2025 passages align with higher B2-C1 levels and demonstrate a clear increase in complexity compared to previous years, particularly in discourse-level features.

4.1.3. Integration of findings

Overall, both strands of analysis indicate a general increase in reading text complexity over time, with the selected 2025 passages consistently reaching levels above the CEFR B1 target. Quantitative measures, particularly Lexile scores and Flesch Reading Ease indices, suggest a shift toward higher-level texts (B2-C1), while qualitative descriptors similarly point to increased demands in terms of structure, meaning, and knowledge requirements. Nonetheless, there is some variation between the two kinds of measurements, particularly in previous years, when quantitative indices may indicate greater difficulty than qualitative assessments. This disparity demonstrates that text complexity is a multifaceted concept and that various instruments capture different facets of reading difficulty. Despite these differences, the judgment that recent exam passages, especially those from 2025, typically exceed the intended B1 level is supported by the general convergence of the evidence. When combined, these results offer a more cohesive response to the research questions by demonstrating how text complexity has evolved over time as well as how several indicators collectively point to a possible mismatch with curriculum objectives. Nevertheless, given the limited sample and the focus on text complexity alone, the results should be interpreted as indicative rather than conclusive.

4.2. Discussion

4.2.1. Curriculum Reform and Exam Misalignment

The increased difficulty of the 2025 NHSGE reading passages can be contextualised within the 2018 education reforms (Circular No. 32/2018/TT-BGDĐT), which introduced a competency-based curriculum with revised objectives, content, and assessment guidelines. While the reform aims to move beyond knowledge recall, its first full implementation in 2025 appears to reveal misalignment between curriculum, instruction, and assessment, as well as inconsistencies between sample and official exams, which may contribute to confusion among teachers and students (Nguyen et al., 2023; Dân trí, 2025). Reports suggest that the examination

places greater emphasis on academically demanding texts and tasks with fewer basic items and greater emphasis on higher-order thinking (Phụ nữ Việt Nam, 2025). Evidence suggests that some reading sections approach C1, exceeding the intended B1 level.

From a policy perspective, curriculum-assessment alignment is essential for ensuring equity and instructional coherence (Porter, 2006). When high-stakes exams exceed prescribed levels, they risk compromising validity and public trust (Haladyna, 2006). In this case, the 2025 NHSGE reflects grammatical, lexical, and conceptual demands closer to B2-C1, indicating limited alignment with curriculum expectations. Several factors may explain this shift. First, policymakers may implicitly benchmark the NHSGE against internationally oriented assessment standards (e.g., TOEFL, IELTS) to enhance perceived rigour, although such alignment requires explicit policy support and curriculum adjustment (Wang, 2019). Second, increased difficulty may serve as a mechanism to differentiate high-achieving students in competitive admissions, potentially conflating certification and selection purposes (Nguyen & Cho, 2022). Finally, limited coordination between curriculum developers and test designers may lead to inconsistent interpretations of the B1 target.

4.2.2. Consequences for validity and fairness

Test validity, the extent to which an exam measures its intended construct, is threatened when test content exceeds policy targets. Under Messick's (1989) unified validity framework, such misalignment introduces irrelevant construct variance. If the NHSGE is intended to certify CEFR B1 proficiency, the inclusion of C1-level reading texts shifts the exam toward a selection function rather than certification. The increasing difficulty of the NHSGE also has important consequences. High-stakes pressure is associated with student stress and mental health risks (Ngo & Pham, 2021), while it encourages teaching-to-the-test and increases instructional demands on teachers (Nguyen et al., 2023). It may also widen educational inequalities, as students in urban areas can access private tutoring, whereas rural ones often rely solely on classroom instruction (Nguyen & Do, 2020). Although educational equity is a key principle of Vietnam's reforms (Resolution 29-NQ/TW), increased test difficulty may disadvantage underprivileged learners. Access to the additional support needed for higher-level tasks is uneven, giving advantaged students greater opportunities to succeed. Such disparities threaten both distributive justice and procedural fairness in high-stakes assessment contexts (Jerrim et al., 2024; Johnson et al., 2021).

4.2.3. Washback, cognitive load and the need for transparent design standards

The linguistic and conceptual demands of the 2025 reading passages may also increase cognitive load, which can reduce learner performance when intrinsic and extrinsic demands are too high (Sweller, 1988). Moreover, heightened exam difficulty may lead to greater anxiety, reduced motivation, and negative attitudes toward English learning (Khan, 2022), as reflected in student reactions reported in the media (Thanh Niên, 2025). To prevent such misalignment, policy should establish clear test design specifications aligned with curriculum objectives, including transparent test blueprints, item-writing guidelines, and representative sample items (Hargreaves, 2020). For Vietnam, this would involve explicitly defining vocabulary load, text difficulty ranges, and cognitive demands for B1-level tasks, alongside conducting regular alignment studies to ensure coherence between curriculum, instruction, and assessment.

4.2.4. The dual function of graduation and selection examinations

The distinction between certification (graduation) and selection (university admission) exams is crucial, as they serve fundamentally different purposes and require different design principles. Certification exams aim to verify minimum competence (e.g., CEFR B1), whereas selection exams differentiate among candidates, often requiring a wider range of difficulty. This distinction is central to interpreting the present findings. The increasing complexity of NHSGE reading passages, particularly in 2025, suggests a shift toward a selection-oriented function. If the exam is intended to certify B1 proficiency, the inclusion of higher-level (B2-C1) texts raises

concerns about alignment and score validity. However, such difficulty may be partly justified if the exam also serves as a selection tool for university admission.

Combining these two purposes in a single high-stakes exam presents challenges. Tasks designed to differentiate top performers may disadvantage students expected only to meet minimum standards, creating tensions between discrimination and fairness, especially in contexts of unequal educational access. This underscores the need for clearer policy articulation of the NHSGE's purpose and careful consideration of whether a single exam can fulfil both functions without compromising validity and equity.

5. Conclusion

The analysis of the selected NHSGE 2025 reading passages indicates a misalignment between the curriculum's proficiency target (B1) and the linguistic and cognitive demands of the exam, which in several cases approach higher levels (e.g., B2-C1). This pattern is consistently observed across both quantitative measures and qualitative CEFR-aligned descriptors, suggesting that the reading texts may exceed the intended level specified in the national curriculum.

At the same time, the implications of this discrepancy should be interpreted with care. While such a mismatch may raise concerns regarding aspects of validity and fairness, as well as potential effects on teaching practices and student experiences, these issues cannot be fully established based on the present analysis alone.

In light of these findings, regular alignment studies should be conducted to examine the relationship between curriculum objectives, classroom instruction, and assessment design. Greater transparency may also support more consistent interpretations of test results, such as clear descriptions of target competency levels and text complexity ranges. In general, the objectives of equitable and competency-based education are likely to be supported by ongoing attention to alignment across curriculum reform, teaching practice, and assessment.

Notes for contributors: MA. Nguyen Thi Khanh An is a lecturer at Faculty of English Language and Culture, ULIS – VNU, Vietnam. MA. Tran Trong Thanh is Head of the Division of English Linguistics and English Language Pedagogy at East Asia University of Technology, Vietnam. Author 1: conceptualization, methodology, data collection & analysis, writing original draft, review & editing; Author 2: conceptualization, supervision, data collection & analysis, writing original draft, review & editing, validation.

Conflict of interest: The authors declare no conflict of interest.

REFERENCES

- Choi, Y., & Zhang, D. (2021). The relative role of vocabulary and grammatical knowledge in L2 reading comprehension: A systematic review of literature. *International Review of Applied Linguistics in Language Teaching*, 59(1), 1-30. <https://doi.org/10.1515/iral-2017-0033>
- Dân trí. (2025). *Vì sao đề thi tốt nghiệp THPT 2025 khó một cách bất thường?* [Why was the 2025 high school graduation exam unusually difficult?] (in Vietnamese). <https://dantri.com.vn/giao-duc/vi-sao-de-thi-tot-nghiep-thpt-2025-kho-mot-cach-bat-thuong-20250629014446655.htm>
- Do, H. (2025). *Đề thi Tiếng Anh tốt nghiệp THPT năm 2025 quá khó: Một 'cú sốc' thật sự?* [Was the 2025 English graduation exam too difficult: A real 'shock'?] (in Vietnamese). <https://tienphong.vn/de-thi-tieng-anh-tot-nghiep-thpt-nam-2025-quua-kho-mot-cu-soc-that-su-post1756118.tpo>
- EnglishProfile (2025a). *English Grammar Profile*. <https://englishprofile.org/?menu=english-grammar-profile>
- EnglishProfile (2025b). *Text Inspector*. <http://englishprofile.org/wordlists/text-inspector>.

- Farr, R., Pritchard, R., & Smitten, B. (1990). A Description of What Happens When an Examinee Takes a Multiple-Choice Reading Comprehension Test. *Journal of Educational Measurement*, 27(3), 209-226. <https://doi.org/10.1111/j.1745-3984.1990.tb00744.x>
- Fisher, D., Frey, N., & Lapp, D. (2012). *Text Complexity: Raising rigor in reading*. Intl Literacy Assn.
- Grabe, W. (2014). *Key issues in L2 reading development*. 8-18. <https://www.nus.edu.sg/celc/wp-content/uploads/2022/09/2.-William-Grabe.pdf>
- Haladyna, T. M. (2006). Perils of standardized achievement testing. *Educational Horizons*, 85(1), 30-43. <https://www.jstor.org/stable/42925964>
- Hargreaves, A. (2020). Large-scale assessments and their effects: The case of mid-stakes tests in Ontario. *Journal of Educational Change*, 21, 393-420. <https://doi.org/10.1007/s10833-020-09380-5>
- Hoang Trang, & Duc Tuan. (2025). *Công bố phổ điểm thi tốt nghiệp THPT 2025* [Announcement of the 2025 high school graduation exam score distribution] (in Vietnamese). <https://baohinhphu.vn/cap-nhat-cong-bo-pho-diem-thi-tot-nghiep-thpt-2025-102250715141512057.htm>
- Jerrim, J., Allen, R., & Sims, S. (2024). High Stakes Assessments in Primary Schools and Teachers' Anxiety About Work. *Educational Assessment*, 29(2), 59-74. <https://doi.org/10.1080/10627197.2024.2350961>
- Johnson, A., Kuhfeld, M., & Soland, J. (2021). The Forgotten 20%: Achievement and Growth in Rural Schools Across the Nation. *AERA Open*, 7. <https://doi.org/10.1177/23328584211052046>
- Khan, M. N. (2022). Effect of online examinations on student's motivation to deeply understand subject content. *International Journal of Health Sciences*, 6(S3), 5266-5275. <https://doi.org/10.53730/ijhs.v6nS3.7082>
- Kuang, Q. (2020). A review of the washback of English language tests on classroom teaching. *English Language Teaching*, 13(9), 10-17. <https://doi.org/10.5539/elt.v13n9p10>
- Lexile (2025). *Framework for Reading*. <https://hub.lexile.com/text-analyzer/>
- Liao L, Ye S. X., and Yang, J. (2023) A mini review of communicative language testing. *Front. Psychol.* <https://doi.org/10.3389/fpsyg.2023.1058411>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). Macmillan.
- MOET. (2013). *Resolution No. 29-NQ/TW on fundamental and comprehensive reform of education and training to meet the requirements of industrialization and modernization in the context of a socialist-oriented market economy and international integration*. <https://thuvienphapluat.vn/van-ban/Thuong-mai/Nghi-quyet-29-NQ-TW-nam-2013-doi-moi-can-ban-toan-dien-giao-duc-dao-tao-hoi-nhap-quoc-te-212441.aspx>
- MOET. (2014). *Circular No. 01/2014/TT-BGDĐT on the Six-Level Foreign Language Proficiency Framework for Vietnam*. <https://thuvienphapluat.vn/van-ban/Giao-duc/Thong-tu-01-2014-TT-BGDĐT-Khung-nang-luc-ngoai-ngu-6-bac-Viet-Nam-220349.aspx>
- MOET. (2018). *Circular No. 32/2018/TT-BGDĐT on the issuance of the General Education Curriculum*. <https://luatvietnam.vn/giao-duc/thong-tu-32-2018-tt-bgdtt-ban-hanh-chuong-trinh-giao-duc-pho-thong-moi-169745-d1.html>
- MOET. (2024a). *Circular No. 24/2024/TT-BGDĐT on the regulation for the high school Graduation Examination*. <https://thuvienphapluat.vn/van-ban/Giao-duc/Thong-tu-24-2024-TT-BGDĐT-Quy-che-thi-tot-nghiep-trung-hoc-pho-thong-622931.aspx>
- MOET. (2024b). *Reference exam papers for the 2025 high school graduation exam*. <https://vqa.moet.gov.vn/vi/news/tin-tuc-su-kien/de-thi-tham-khao-ky-thi-tot-nghiep-thpt-tu-nam-2025-159.html>

- Natova, I. (2019). Estimating CEFR reading comprehension text complexity. *The Language Learning Journal*, 49(6), 699-710. <https://doi.org/10.1080/09571736.2019.1665088>
- Ngo, V. M., & Pham, T. H. L. (2021). Thực trạng lo âu, trầm cảm ở học sinh lớp 12 tại 2 trường Trung học phổ thông của thành phố Thái Bình năm 2020 [Current Situation of Anxiety, Depression, and Some Related Factors Among 12th Grade Students of Two High School in Thai Binh, 2020] (in Vietnamese). *Vietnam Medical Journal*, 506(2). <https://doi.org/10.51298/vmj.v506i2.1256>
- Nguyen, P. L., & Do, K. C. (2020). Education, Employment, and Income: An Overview of Rural-Urban Gaps in Vietnam. *Vietnam Journal of Agricultural Sciences*, 3(2), 647-661. <https://doi.org/10.31817/vjas.2020.3.2.08>
- Nguyen, T. H. N., Do, T. T. D., Luu, T. H., Nguyen, T. M. P., Tran, T. M. P., & Mai, T. V. A. (2023). Thực tiễn triển khai Chương trình giáo dục phổ thông 2018 môn Tiếng Anh: Góc nhìn từ giáo viên thực hiện chương trình [The implementation of the 2018 general education curriculum for English: Perspectives from teachers] (in Vietnamese). *Vietnam Journal of Education*, 23(5), 58-63. <https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/view/667>
- Nguyen, V. B., & Cho, Y. (2022). How the High-Stakes and College Entrance Exam Affects Students' Perception: Implication on Management Policy in Higher Education. *East Asian Journal of Business Economics (EAJBE)*, 10(2), 83-94. <https://doi.org/10.20498/EAJBE.2022.10.1.83>
- Phụ nữ Việt Nam. (2025, June 27). *Thi tốt nghiệp THPT 2025: Thí sinh than trời 'Đề thi tiếng Anh khó hơn cả thi IELTS'* [Students complain the English test is harder than IELTS] (in Vietnamese). <https://phunuvietnam.vn/thi-tot-nghiep-thpt-2025-thi-sinh-than-troi-de-thi-tieng-anh-kho-hon-ca-thi-ielts-20250627112813585.htm>
- Porter, A. (2006). Curriculum Assessment. In Green, J. L., Camilli, G., & Elmore, P. B. (Eds.), *Handbook of Complementary Methods in Education Research* (3rd ed.). Routledge.
- Spencer, M., Gilmour, A. F., Miller, A. C., Emerson, A. M., Saha, N. M., & Cutting, L. E. (2018). Understanding the influence of text complexity and question type on reading outcomes. *Reading and Writing*, 32(3), 603-637. <https://doi.org/10.1007/s11145-018-9883-0>
- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive science: A multidisciplinary approach*, 12(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4
- Thanh Niên. (2025). *Nhận xét đề tiếng Anh kỳ thi tốt nghiệp THPT 2025: Đề thi khó, tiệm cận với đề thi quốc tế* [Comments on the English test of the 2025 high school graduation exam: The test was difficult and approached international standards] (in Vietnamese). <https://thanhnien.vn/nhan-xet-de-tieng-anh-ky-thi-tot-nghiep-thpt-2025-de-thi-kho-tiem-can-voi-de-thi-quoc-te-185250627142953192.htm>
- Wang, T. (2019). Competence for Students' Future: Curriculum Change and Policy Redesign in China. *ECNU Review of Education*, 2(2), 234-245. <https://doi.org/10.1177/2096531119850905>