DISTRIBUTIONAL SHIFTS IN ENGLISH SCORES AFTER EXAMINATION REFORM: A SECONDARY QUANTITATIVE ANALYSIS OF THE 2025 AND 2026 VIETNAMESE NATIONAL HIGH SCHOOL GRADUATION EXAMINATIONS

Nguyen Quoc Khanh[1*], Vo Nhat Quang[2]
[1]The University of Da Nang, Da Nang, Vietnam
[2]Ho Chi Minh University of Banking, Ho Chi Minh City, Vietnam
Email address: av32.1b2cva1d047@oncce.udn.vn
**Corresponding author: Nguyen Quoc Khanh*
https://doi.org/10.51453/3093-3706/2026/1486

ARTICLE INFO

Received: 25/06/2026
Revised: 01/08/2026
Published: 28/08/2026

KEYWORDS

Secondary data analysis; Score distribution; English examination; Assessment reform; High-stakes testing.

ABSTRACT

This study examines changes in the English score distribution of the Vietnamese National High School Graduation Examination from 2025 to 2026 using secondary aggregate data. The study was designed as a descriptive quantitative analysis of officially released national score-distribution indicators. It focused on three dimensions of distributional change: central tendency, dispersion, and lower- and upper-tail score outcomes. The analysed indicators included the number of candidates, mean score, median score, modal score, standard deviation, median absolute deviation, the proportion of candidates scoring below 5, the proportion scoring 7 or above, and the number and rate of perfect scores. The results show that the 2026 distribution shifted downward in central tendency compared with 2025. The mean decreased from 5.38 to 5.07, the median from 5.25 to 4.75, and the mode from 5.25 to 4.25. At the same time, dispersion increased, with both the standard deviation and median absolute deviation rising. The lower-score group expanded substantially, as the proportion of candidates scoring below 5 increased from 38.22% to 50.25%. However, the upper extreme also increased, with perfect scores rising from 141 to 311. These findings suggest that the 2026 distribution should not be interpreted simply as a uniform decline. Rather, it showed a lower central location, wider spread, and a wider and less centrally concentrated score profile. The study contributes by demonstrating how official score-distribution data can be used as a secondary quantitative dataset for examining national examination outcomes while remaining cautious about causal interpretation.

DỊCH CHUYỂN PHÂN PHỐI ĐIỂM TIẾNG ANH SAU CẢI CÁCH KỲ THI: PHÂN TÍCH ĐỊNH LƯỢNG THỨ CẤP CÁC KỲ THI TỐT NGHIỆP TRUNG HỌC PHỔ THÔNG VIỆT NAM NĂM 2025 VÀ 2026

Nguyễn Quốc Khánh[1*], Võ Nhật Quang[2]
[1]Đại học Đà Nẵng, Đà Nẵng, Việt Nam
[2]Trường Đại học Ngân hàng Thành phố Hồ Chí Minh, TP. Hồ Chí Minh, Việt Nam
Địa chỉ email: av32.1b2cva1d047@oncce.udn.vn
**Tác giả liên hệ: Nguyễn Quốc Khánh*
https://doi.org/10.51453/3093-3706/2026/1486

THÔNG TIN BÀI BÁO

Ngày nhận bài: 25/06/2026
Ngày hoàn thiện: 01/08/2026
Ngày đăng: 28/08/2026

TỪ KHÓA

Phân tích thứ cấp;
Phân phối điểm;
Kỳ thi tiếng Anh;
Cải cách đánh giá;
Thi cử hệ trọng

TÓM TẮT

Nghiên cứu này xem xét sự thay đổi trong phân phối điểm môn Tiếng Anh của Kỳ thi tốt nghiệp trung học phổ thông Việt Nam từ năm 2025 đến 2026 bằng dữ liệu tổng hợp thứ cấp. Nghiên cứu sử dụng thiết kế định lượng mô tả, phân tích các chỉ báo phân phối điểm cấp quốc gia được công bố chính thức theo ba phương diện: xu hướng trung tâm, độ phân tán và kết quả ở phần đuôi thấp và đuôi cao của phân phối. Các chỉ báo gồm số thí sinh, điểm trung bình, trung vị, mốt, độ lệch chuẩn, độ lệch tuyệt đối trung vị, tỷ lệ thí sinh dưới 5 điểm, tỷ lệ đạt từ 7 điểm trở lên, cùng số lượng và tỷ lệ điểm tuyệt đối. Kết quả cho thấy phân phối năm 2026 dịch xuống về xu hướng trung tâm: điểm trung bình giảm từ 5,38 xuống 5,07, trung vị từ 5,25 xuống 4,75 và mốt từ 5,25 xuống 4,25. Đồng thời, độ phân tán tăng; tỷ lệ dưới 5 điểm tăng từ 38,22% lên 50,25%, trong khi số điểm tuyệt đối tăng từ 141 lên 311. Vì vậy, kết quả năm 2026 không nên được diễn giải đơn giản như một sự suy giảm đồng đều, mà phản ánh vị trí trung tâm thấp hơn, độ phân tán rộng hơn và mức độ tập trung quanh trung tâm thấp hơn. Nghiên cứu cho thấy giá trị của dữ liệu phân phối điểm chính thức đối với phân tích định lượng thứ cấp, đồng thời duy trì sự thận trọng trong diễn giải nhân quả.

1. INTRODUCTION

1.1. Background and context

The Vietnamese National High School Graduation Examination is one of the most consequential assessment events in Vietnam because its results are used not only for graduation recognition but also for admission-related purposes in higher education. For this

reason, publicly released score distributions are not merely descriptive summaries; they become part of how candidates, parents, universities, colleges, and policymakers interpret examination outcomes. In 2026, the Government Electronic Newspaper reported that the English score distribution was officially released by the Ministry of Education and Training on July 1 and that the score distributions were published to provide information for candidates, parents, higher education institutions, colleges, and relevant agencies in admissions and examination evaluation (Báo Điện tử Chính phủ, 2026).

The 2025 examination also marked an important structural transition. From 2025 onwards, candidates take four subjects: two compulsory subjects, Literature and Mathematics, and two elective subjects selected from the subjects studied in Grade 12, including Foreign Languages. For non-Literature subjects, the examination is administered in multiple-choice format, with most subjects having a 50-minute testing time (Báo Điện tử Chính phủ, 2025). This means that English score distributions from 2025 onward need to be interpreted in a new post-reform context, where the candidate pool and examination structure may not be directly comparable with earlier periods.

This reform context is linked to the broader implementation of Vietnam's 2018 General Education Curriculum. Circular No. 32/2018/TT-BGDĐT set out the new general education programme and its implementation roadmap, with Grade 12 scheduled for implementation in the 2024-2025 school year (Ministry of Education and Training, 2018). Therefore, the 2025 and 2026 English examinations are especially important for assessment research because they represent the first years in which national examination outcomes can be considered in relation to the new curriculum cycle.

1.2. Problem statement, urgency, and significance

Public discussion of national examination results often focuses on isolated headline indicators such as the mean score, the number of candidates scoring below the midpoint, or the number of perfect scores. However, such isolated indicators can lead to oversimplified conclusions. A lower mean score, for example, does not necessarily mean that an examination was uniformly more difficult; similarly, a higher number of perfect scores does not necessarily mean that the examination was easier. A score distribution must be examined through multiple indicators, including central tendency, dispersion, lower-tail outcomes, and upper-tail outcomes.

This issue is urgent because the interpretation of score distributions has real educational consequences. AERA's position statement on high-stakes testing emphasizes that high-stakes tests carry serious consequences for students and that score-based decisions should be supported by appropriate validity evidence, meaningful opportunity to learn, and alignment between the test and the curriculum (American Educational Research Association, 2000). In the Vietnamese context, where score distributions are publicly released and used to support admissions-related decisions, a systematic analysis of distributional change is therefore educationally and socially significant. The urgency is heightened in the 2025-2026 period because these are the first adjacent years after the implementation of the new

examination structure. In such a transitional context, public interpretation of score distributions may influence how stakeholders understand the effects of examination reform, candidate readiness, and English learning outcomes.

The present study does not attempt to judge whether the 2026 English examination was valid, invalid, easy, or difficult. Instead, it treats the official score distributions as secondary aggregate data and asks a narrower quantitative question: how did the English score distribution change from 2025 to 2026? This distinction is important because contemporary validity theory emphasizes that test-score interpretations and uses, rather than tests in isolation, require evidence-based evaluation (American Educational Research Association et al., 2014; Kane, 2013).

The significance of this study is therefore educational, analytical, and policy-relevant: it provides evidence for understanding national English examination outcomes, demonstrates the value of multi-indicator score-distribution interpretation, and supports more cautious public discussion of high-stakes examination results.

1.3. Research gap and novelty

Previous studies have examined English score distributions in the Vietnamese National High School Graduation Examination. Ly (2023) analysed the 2023 English score distribution using histograms and descriptive statistics, while Ly and Nguyen (2023) compared English test scores across 2020-2023 using frequency distributions and histograms. These studies established that national English score distributions can be studied quantitatively, but they focused on examination cycles before the 2025 structural reform.

Other studies have examined Vietnamese English examination performance from different perspectives. Duong et al. (2024) analysed provincial English test-score data from 2019 to 2021 and linked provincial performance to socioeconomic and school-level variables. Nguyen and Dao (2025) examined washback effects of the 2025 National High School Graduation Exam, with attention to curriculum alignment, Bloom's taxonomy, and learning practices. These studies are useful for understanding English examination outcomes in Vietnam, but they do not provide a focused national comparison of English score-distribution indicators between the first two years after the 2025 reform.

The novelty of the present study is therefore modest and context-specific. It does not propose a new psychometric model or claim that multi-indicator distributional analysis is methodologically new. Instead, it provides an early post-reform update by comparing the 2025 and 2026 English score distributions using publicly released national aggregate indicators. Its contribution lies in documenting how central tendency, dispersion, and selected lower- and upper-tail outcomes changed across two adjacent post-reform years, while explicitly recognising that these changes may reflect not only examination characteristics but also changes in the composition of candidates who chose to sit the English paper.

1.4. Aim, contribution, and research questions

The aim of this study is to conduct a secondary quantitative analysis of aggregate English score-distribution data from the 2025 and 2026 Vietnamese National High School Graduation Examinations. The study uses officially released score-distribution indicators to describe changes in central tendency, dispersion, and score-band outcomes.

Within this limited descriptive scope, the study makes three contributions. First, it provides an empirical contribution by documenting the 2025-2026 English score-distribution shift in the post-reform examination context. Second, it provides an analytical contribution by moving beyond a single-indicator interpretation and examining central tendency, dispersion, lower-tail outcomes, and upper-tail outcomes together. Third, it provides a policy-relevant contribution by showing how publicly released aggregate score distributions can be used as a starting point for more careful assessment research, while also making clear that such data cannot support causal claims about item difficulty, teaching quality, or curriculum alignment.

The study is guided by the following research questions:

RQ1. How did the central tendency of English examination scores change from 2025 to 2026?

RQ2. How did score dispersion change between 2025 and 2026?

RQ3. How did lower-tail and upper-tail score outcomes change from 2025 to 2026?

2. LITERATURE REVIEW

2.1. High-stakes testing and score interpretation

High-stakes tests require careful interpretation because their results can affect educational opportunity, institutional decision-making, and public perceptions of school quality. AERA (2000) defines high-stakes testing contexts as those in which test results carry serious consequences for students or educators. It also cautions that tests used for consequential purposes require sufficient reliability, appropriate validity evidence, curriculum alignment, and attention to intended and unintended consequences. These principles are directly relevant to national graduation examinations, where score distributions can shape public debate and institutional decision-making.

Validity theory provides the broader conceptual foundation for this study. The Standards for Educational and Psychological Testing state that validity concerns the evidence supporting the interpretation of test scores for intended uses (American Educational Research Association et al., 2014). Similarly, Kane (2013) argues that validation involves evaluating the plausibility of the claims made from test scores rather than treating validity as a fixed property of the test itself. In this view, aggregate score distributions can provide useful descriptive evidence about national performance patterns, but they cannot by themselves establish why those patterns occurred. Sireci and Padilla (2014) similarly frame validation as an evidence-based argument supporting the intended uses of test scores.

2.2. Secondary quantitative analysis of large-scale assessment data

Secondary data analysis is appropriate when researchers use existing data to answer new research questions. In large-scale testing contexts, testing agencies often release summary statistics at national, regional, or institutional levels, and researchers may use these data to examine broader patterns than those discussed in the original reporting. Rogers et al. (2006) note that secondary analysts can use large-scale assessment data to answer broader or deeper questions, but such analyses must remain aligned with the structure and limits of the available data.

This caution is especially important in national examination contexts. Official aggregate statistics can make national-level analysis possible without collecting primary data from hundreds of thousands of candidates. However, aggregate data cannot identify individual-level mechanisms, item-level causes, or subgroup differences unless those variables are included in the dataset. Therefore, the present study treats the released English score-distribution indicators as suitable for descriptive comparison, not for causal explanation.

2.3. Previous research on examination score distributions and reform

Research on English score distributions in Vietnam has developed gradually. Ly (2023) analysed the English score distribution in the 2023 Vietnamese National High School Graduation Examination using histograms and descriptive statistics. The study reported that the 2023 English score distribution was right-skewed and that English had one of the lowest average scores among examination subjects. Ly and Nguyen (2023) extended this line of inquiry by comparing English test scores across 2020-2023 using frequency distributions and histograms. Their study found that English score distributions during this period were generally skewed to the right and that high English scores were especially prominent in 2021.

Other studies have approached Vietnamese English examination outcomes from different angles. Duong et al. (2024) analysed provincial mean and median English test scores from 2019 to 2021 and examined their relationship with socioeconomic and school-level variables. Nguyen and Dao (2025) investigated the 2025 National High School Graduation Exam in English in relation to the 2018 General Education English Curriculum, Bloom's taxonomy, and washback effects on learning. These studies are valuable because they show that English examination outcomes can be examined statistically, socioeconomically, and pedagogically. However, they do not provide a focused national comparison of the 2025 and 2026 English score-distribution indicators under the post-reform examination structure.

International evidence also shows that examination reforms can reshape score patterns without producing a simple uniform shift. In Germany, Puhani and Yang (2020) found that centralising written high-school exit examinations increased scoring differences between higher- and lower-performing schools by about 25%. In Colombia, Riehl (2024) reported that a reform of the college-admission examination reduced score gaps between high- and low-income students by nearly half in some subjects, while also altering the exam's predictive properties. Bratti et al. (2026), examining a reform that reduced university exam-retake

opportunities in Italy, found session-specific evidence consistent with shifts in grade distributions but cautioned that changes in who sat particular exam sessions complicated interpretation. Together, these studies indicate that post-reform score patterns may reflect changes in test design, scoring rules, incentives, and examinee composition. This supports interpreting the Vietnamese 2025-2026 shift descriptively rather than attributing it to examination difficulty alone.

2.4. Conceptual framework

This study is guided by a validity-oriented secondary score-distribution framework. The framework has three components. First, the Vietnamese National High School Graduation Examination is treated as a high-stakes assessment context, meaning that score interpretation has practical consequences for candidates, institutions, and public understanding. Second, score distributions are treated as population-level outcome evidence. Indicators such as the mean, median, mode, standard deviation, median absolute deviation, and score-band proportions describe different aspects of the reported distribution. Third, the data are treated as secondary aggregate data, meaning that they are appropriate for describing national-level patterns but not for explaining individual-level or item-level mechanisms.

Within this framework, the study does not ask whether the 2026 English examination was harder or easier than the 2025 examination on the basis of aggregate indicators alone. Instead, it asks how the reported national English score distribution changed across central tendency, dispersion, and selected lower- and upper-tail indicators. This framework directly supports the three research questions and informs the structure of the methodology and results sections.

3. METHOD

3.1. Research design

This study adopted a secondary quantitative data analysis design. Secondary data analysis refers to the use of existing data collected or released by other organisations or researchers to answer a new research question (Wickham, 2019; Tripathy, 2013). This design was appropriate because the study did not collect primary data from candidates. Instead, it analysed officially released aggregate score-distribution indicators from the English component of the Vietnamese National High School Graduation Examination.

The study was descriptive and comparative. It compared the 2025 and 2026 English score distributions using selected national-level indicators. The aim was not to test causal hypotheses or to determine why the distribution changed. Rather, the study described how the distribution changed across central tendency, dispersion, and score-band outcomes.

3.2. Data sources

The main dataset consisted of aggregate English score-distribution statistics for the 2025 and 2026 Vietnamese National High School Graduation Examinations. The 2026 data were taken from the comparative English score-distribution table published by Báo Điện tử Chính phủ (2026), which reported score-distribution information released by the Ministry of

Education and Training. Although this source is presented in newspaper format, it is not an independent media reconstruction of the dataset; it is a government information portal reporting officially released examination statistics. The table provided the national-level indicators used in this study, including candidate numbers, mean, median, mode, standard deviation, median absolute deviation, selected score-band counts and percentages, zero scores, and perfect-score indicators.

The 2025 figures were taken from the same comparative 2025-2026 English score-distribution table in order to ensure that both years were drawn from a single reporting format. To strengthen data verification, two key 2025 indicators-the mean score and the number of perfect scores-were cross-checked against the Ministry of Education and Training's 2025 score-distribution briefing, which reported that English had a mean score of 5.38 and 141 perfect scores in 2025 (Ministry of Education and Training, 2025). This cross-check confirmed the consistency of these two central indicators, but the remaining 2025 indicators could not be independently verified from a full official downloadable dataset.

The comparison between 2025 and 2026 was selected because these are the first two adjacent examination years after the implementation of the new examination structure from 2025. Under this structure, candidates take four subjects, including two compulsory subjects and two elective subjects selected from Grade 12 subjects. Therefore, English score distributions from 2025 onward represent the performance of candidates who selected English as an examination subject rather than the performance of an identical compulsory-examination population. This candidate-composition issue was treated as a central interpretive limitation of the study rather than as a minor background detail.

The Ministry of Education and Training announced the structure and format of the high school graduation examination from 2025 through the Department of Quality Management (Department of Quality Management, 2023). Accordingly, the present analysis should be understood as a comparison of officially reported national aggregate indicators for two post-reform English examination cohorts, not as a comparison of the same stable candidate population over time.

3.3. Variables and indicators

Following the validity-oriented secondary score-distribution framework presented in Section 2.4, the variables were selected to represent three dimensions of distributional change: central tendency, dispersion, and tail-score outcomes. The unit of analysis was the national English score distribution by examination year. The study did not use individual-level candidate scores, item-level responses, school-level data, or provincial-level data. Therefore, all variables were treated as aggregate national indicators.

The selected indicators were grouped into three categories. First, central tendency indicators included the mean score, median score, and modal score. These indicators were used to examine where the distribution was located in each year. Second, dispersion indicators included the standard deviation and median absolute deviation. These indicators

were used to examine whether scores were more or less spread out around the centre of the distribution. Third, score-band and tail indicators included the number and percentage of candidates scoring below 5, the number and percentage scoring 7 or above, the number of perfect scores, the rate of perfect scores per 1,000 candidates, the number of zero scores, and the number and percentage of candidates scoring at or below 1.

These indicators were selected because they correspond directly to the study's three research questions: central tendency for RQ1, dispersion for RQ2, and lower-tail and upper-tail outcomes for RQ3.

3.4. Data extraction and verification

The data were extracted manually from the official comparative English score-distribution table published in the 2026 Government Electronic Newspaper report. The extracted values included the number of candidates, mean, median, mode, standard deviation, median absolute deviation, counts and percentages for selected score bands, perfect scores, zero scores, and perfect scores per 1,000 candidates.

To reduce transcription error, all extracted values were checked against the original table after entry. In addition, two 2025 indicators—the mean score and the number of perfect scores—were cross-checked with the Ministry of Education and Training's 2025 briefing. No values were estimated from graphs, and no missing values were imputed. Only indicators explicitly reported in the official score-distribution materials were included in the analysis.

3.5. Data analysis

The analysis was conducted descriptively. For each indicator, the study reported the 2025 value, the 2026 value, and the absolute change between the two years. For percentage-based indicators, changes were reported in percentage points rather than as ordinary percentage increases. Percentages were rounded consistently to two decimal places in the presentation of results. To improve interpretability, two illustrative bar charts were produced from the reported aggregate indicators. The first chart visualised changes in central tendency and dispersion, including the mean, median, mode, standard deviation, and median absolute deviation. The second chart visualised selected lower- and upper-tail indicators, including the percentage of candidates scoring below 5, the percentage scoring 7 or above, and the rate of perfect scores per 1,000 candidates. These charts were used only to visualise the reported summary indicators; they were not treated as histograms or density plots of the full score distribution.

A standardised mean difference was calculated to describe the magnitude of the change in mean scores relative to the pooled standard deviation. Effect sizes are useful because they describe the size of a difference in a standardised form rather than relying only on raw-score change (Lakens, 2013). The standardised mean difference was calculated as follows:

$$d = \frac{M_{2026} - M_{2025}}{SD_{pooled}}$$

where:

$$SD_{pooled} = \sqrt{\frac{(n_{2025} - 1)SD_{2025}^2 + (n_{2026} - 1)SD_{2026}^2}{n_{2025} + n_{2026} - 2}}$$

In this formula, M refers to the mean score, SD refers to the standard deviation, and n refers to the number of candidates in each examination year. A negative value of d indicates that the 2026 mean was lower than the 2025 mean. Following conventional interpretation, a value around 0.20 was treated as a small standardised mean difference. Therefore, the effect size was used only to describe the magnitude of the mean-score shift, not to support a claim of substantial educational decline.

No inferential significance test was conducted. The reason is not only that the candidate numbers were very large, but also that the analysis used officially reported national aggregate indicators for the examined cohorts rather than a probability sample drawn from a larger target population. In this context, a significance test would not answer the main research question, because the study aimed to describe the reported national distributions rather than estimate an unknown population parameter from sample data.

The study did not calculate skewness or kurtosis because the publicly reported dataset did not provide the full individual-level score vector or a complete machine-readable frequency distribution. For this reason, the study avoids making formal claims about distributional pattern. Terms such as “more dispersed,” “lower central tendency,” and “mixed tail movement” are used to describe the available summary indicators, while stronger terms such as “polarisation” or “change in skewness” are avoided unless supported by future studies using full frequency data.

3.6. Scope and limitations of the method

Because the study used aggregate secondary data, it could describe national-level distributional change but could not explain individual-level or item-level mechanisms. The data did not include candidate background variables, school-level characteristics, provincial socioeconomic indicators, item facility values, item discrimination indices, distractor analysis, response-time evidence, or candidate-processing data. Therefore, the study did not attempt to identify the causes of the observed distributional changes.

The most important comparability limitation concerns candidate composition. From 2025, English was part of a restructured examination system in which candidates selected two elective subjects in addition to Literature and Mathematics. As a result, the English examination cohorts in 2025 and 2026 may not represent the same underlying population of learners. Differences in candidate motivation, university-admission strategy, English proficiency, regional access to English learning, and subject-selection behaviour may all have influenced who chose to sit the English paper in each year. Therefore, the observed score-

distribution differences should not be interpreted as evidence that the same population of candidates performed worse or better in 2026.

A second limitation concerns data verification. The study relied primarily on the comparative score-distribution table published by Báo Điện tử Chính phủ (2026), supplemented by cross-checking selected 2025 indicators with the Ministry of Education and Training's 2025 briefing. Although the main source reported officially released statistics, the study did not have access to the full candidate-level dataset or a complete official downloadable frequency table for independent recalculation. This limits the extent to which additional distributional statistics, such as skewness and kurtosis, could be computed. It also prevents the construction of true histograms or density plots based on the full score-frequency distribution, although illustrative bar charts can still be produced from the reported aggregate indicators.

These limitations define the appropriate scope of inference. The study provides a descriptive comparison of officially reported English score-distribution indicators across two adjacent post-reform years. It does not provide a causal explanation of examination difficulty, teaching quality, curriculum alignment, or candidate ability.

3.7. Ethical considerations

The study used publicly released aggregate data. No personal information, individual candidate scores, identifiable student records, or school-level confidential data were used. Because the dataset contained only national-level summary statistics, the analysis did not involve human-subject recruitment, consent procedures, or privacy-sensitive information. The ethical responsibility of the study was therefore to report the official data accurately, avoid unsupported causal claims, and interpret the findings within the limits of aggregate secondary data.

4. RESULTS

4.1. Overview of selected indicators

This section presents the selected national English score-distribution indicators for 2025 and 2026. The indicators include candidate numbers, central tendency measures, dispersion measures, selected score-band outcomes, and perfect-score indicators. As described in Section 3.2, the values were taken from the official comparative score-distribution material published by Báo Điện tử Chính phủ (2026), with selected 2025 indicators cross-checked against the Ministry of Education and Training's 2025 score-distribution briefing.

Table 1. Central Tendency and Dispersion of English Scores in 2025 and 2026

Indicator	2025	2026	Change
Number of candidates	351,848	334,547	-17,301
Mean score	5.38	5.07	-0.31
Median score	5.25	4.75	-0.50
Modal score	5.25	4.25	-1.00

Standard deviation	1.45	1.68	+0.23
Median absolute deviation	1.16	1.37	+0.21

Note: Values for 2025 and 2026 are from the comparative English score-distribution table published by *Báo Điện tử Chính phủ* (2026). Change values were calculated by the author.

4.2. Changes in central tendency and dispersion

Table 1 addresses the first two research questions by presenting indicators of central tendency and dispersion. The mean, median, and modal scores were used to examine the location of the distribution, while the standard deviation and median absolute deviation were used to examine the spread of scores around the centre.

The number of candidates taking the English examination decreased from 351,848 in 2025 to 334,547 in 2026. The mean score decreased from 5.38 to 5.07, while the median score decreased from 5.25 to 4.75. The modal score shifted from 5.25 in 2025 to 4.25 in 2026. In contrast, the standard deviation increased from 1.45 to 1.68, and the median absolute deviation increased from 1.16 to 1.37. Based on the reported means, standard deviations, and candidate counts, the standardized mean difference between 2026 and 2025 was approximately d = -0.20. Figure 1 visualises these changes across the five central tendency and dispersion indicators. The figure shows that all three location indicators moved downward, while both dispersion indicators increased in 2026.

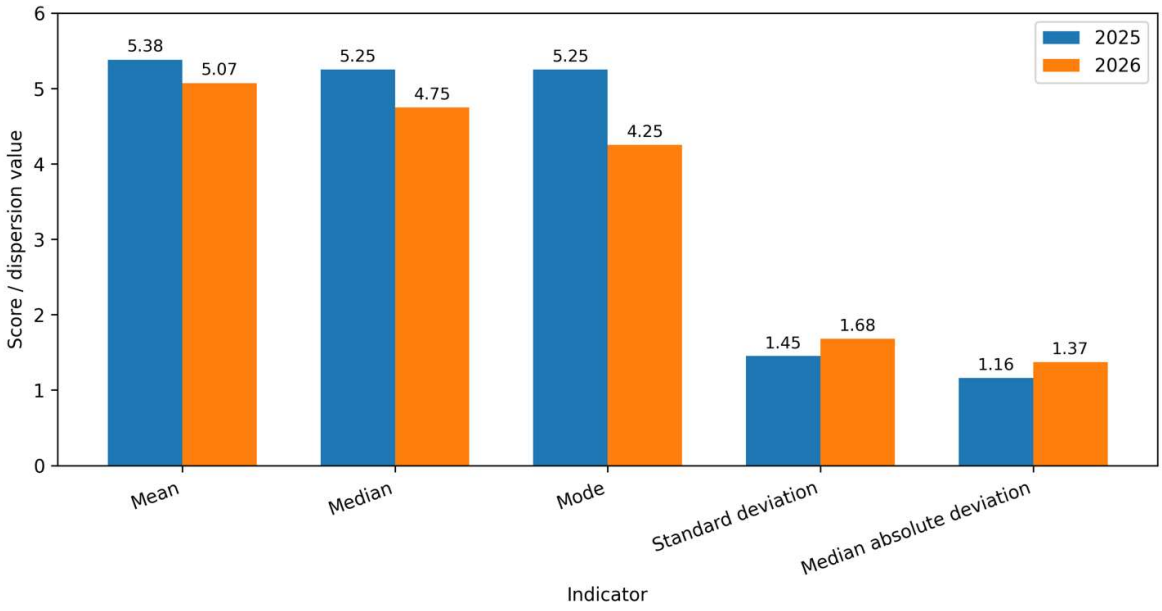


Figure 1. Central Tendency and Dispersion Indicators in 2025 and 2026

Note: Values are based on the reported aggregate indicators in Table 1. The figure is an illustrative bar chart of summary indicators, not a histogram of the full score distribution.

4.3. Lower-tail and upper-tail score indicators

Table 2 addresses the third research question by presenting lower-tail and upper-tail indicators. These indicators were included because changes in the mean score alone cannot show whether movement occurred mainly among lower-performing candidates, higher-performing candidates, or both ends of the score distribution.

Table 2. Threshold and Tail-Score Indicators in 2025 and 2026

Indicator	2025	2026	Change
Scores below 5	134,478	168,105	+33,627
Percentage below 5	38.22%	50.25%	+12.03 pp
Scores ≥ 7	53,251	53,107	-144
Percentage ≥ 7	15.14%	15.87%	+0.74 pp
Perfect scores	141	311	+170
Perfect scores per 1,000 candidates	0.4007	0.9296	+0.5289
Zero scores	2	0	-2
Scores ≤ 1	28	61	+33
Percentage ≤ 1	0.01%	0.02%	+0.01 pp

Note: pp = percentage points. Values for 2025 and 2026 are from the comparative English score-distribution table published by *Báo Điện tử Chính phủ* (2026). Change values were calculated by the author.

The number of candidates scoring below 5 increased from 134,478 in 2025 to 168,105 in 2026. In percentage terms, the below-5 group increased from 38.22% to 50.25%, corresponding to an increase of 12.03 percentage points. The number of candidates scoring 7 or above decreased slightly from 53,251 in 2025 to 53,107 in 2026; however, the percentage of candidates scoring 7 or above increased from 15.14% to 15.87%.

At the upper extreme, the number of perfect scores increased from 141 in 2025 to 311 in 2026. The rate of perfect scores per 1,000 candidates increased from 0.4007 to 0.9296. At the lower extreme, the number of zero scores decreased from 2 to 0, while the number of candidates scoring at or below 1 increased from 28 to 61. In percentage terms, the ≤1 group increased from 0.01% to 0.02%.

Figure 2 visualises selected threshold and tail indicators. The figure highlights the increase in the below-5 group, the slight increase in the percentage of candidates scoring 7 or above, and the rise in perfect-score rate per 1,000 candidates.

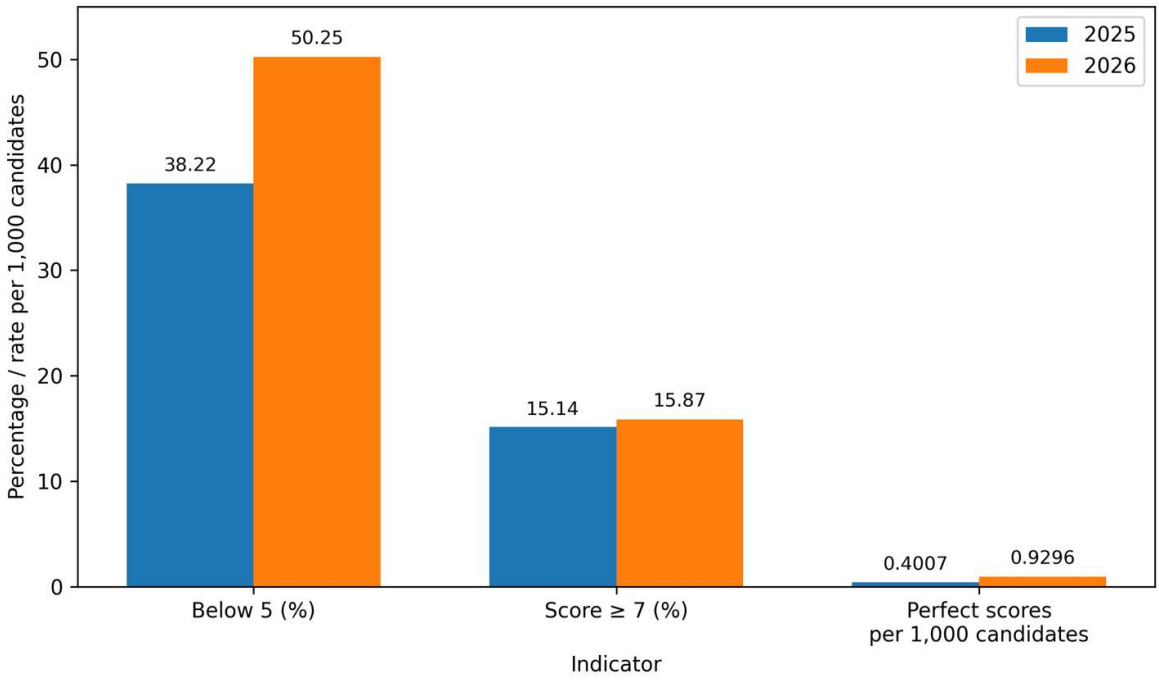


Figure 2. Selected Threshold and Tail-Score Indicators in 2025 and 2026

Note: The below-5 and ≥7 indicators are percentages. Perfect scores are reported as the number of perfect scores per 1,000 candidates. The figure is intended to visualise selected reported indicators and does not represent the full score-frequency distribution.

4.4. Summary of descriptive changes

Across the selected indicators, four descriptive changes were observed from 2025 to 2026. First, all three central tendency indicators decreased: the mean declined by 0.31 points, the median by 0.50 points, and the mode by 1.00 point. Second, both dispersion indicators increased: the standard deviation rose by 0.23, and the median absolute deviation rose by 0.21. Third, the below-5 group increased by 33,627 candidates, or 12.03 percentage points. Fourth, the upper extreme also increased, with the number of perfect scores rising by 170 and the rate of perfect scores per 1,000 candidates increasing by 0.5289.

The provincial top-10 tables were not included in the main results because the present analysis focuses on national-level distributional changes rather than regional variation.

5. DISCUSSION

5.1. A lower centre, but not necessarily a lower-performing stable population

RQ1 asked how the central tendency of English examination scores changed from 2025 to 2026. The results show that the centre of the reported English score distribution moved downward: the mean, median, and mode were all lower in 2026. However, this pattern should not be interpreted as evidence that the same national population of learners became weaker in English. The more defensible interpretation is that the reported English-

examination cohort in 2026 had a lower score location than the reported English-examination cohort in 2025.

This distinction is essential because English was no longer interpreted in the same way as a compulsory examination subject for all candidates. Under the post-2025 structure, candidates selected two elective subjects in addition to Literature and Mathematics. Therefore, the English score distribution reflects not only examination performance but also subject-selection behaviour. A change in the English cohort may occur if different groups of candidates decide to choose or avoid English in different years. For example, candidates with stronger English may still select English for admission purposes, while candidates with weaker English may select other subjects when possible. Conversely, some candidates may select English because it remains strategically useful for particular university programmes. Without candidate-level background variables, the present study cannot separate score change from cohort-composition change.

Therefore, the downward movement in central tendency should be read as a post-reform distributional signal rather than a direct measure of national decline in English proficiency. This interpretation is more consistent with validity theory, which requires score interpretations to be supported by evidence about the tested population, test use, and contextual conditions (American Educational Research Association et al., 2014; Kane, 2013).

5.2. Increased dispersion and the possibility of stronger differentiation

RQ2 asked how score dispersion changed between the two years. Both the standard deviation and median absolute deviation increased in 2026, indicating that scores were more spread out around the centre. The standardised mean difference was approximately $d = -0.20$, which is conventionally small. This means that the downward shift in the mean was modest in standardised terms, even though it may still be practically meaningful in a high-stakes admission context.

The increase in dispersion is more informative than the mean shift alone. A wider distribution suggests that the 2026 English paper, the 2026 English candidate cohort, or both produced stronger differentiation among candidates. In a high-stakes graduation and admission-related examination, differentiation is expected to some extent. However, dispersion alone cannot show whether the test differentiated candidates fairly or whether the wider spread was associated with uneven preparation, differential access to English learning resources, or changes in the linguistic demands of the paper.

This finding points to the need for item-level evidence. If future data show that certain item types, reading passages, or vocabulary demands produced much lower facility values, the dispersion pattern could be linked more directly to test design. If subgroup data show that dispersion increased mainly across regions, school types, or candidate groups, the pattern could instead indicate inequality in preparation. The present aggregate data cannot choose between these explanations, but it can identify increased dispersion as a pattern that deserves closer investigation.

5.3. Mixed lower-tail and upper-tail movement

RQ3 asked how lower-tail and upper-tail outcomes changed from 2025 to 2026. The most visible change was the expansion of the below-5 group, which increased from 38.22% to 50.25%. This indicates that a larger share of English candidates scored below the midpoint of the scale in 2026. In a high-stakes context, this is educationally important because scores around and below the midpoint may affect graduation-related confidence, admission options, and public perceptions of English learning outcomes.

At the same time, the upper end of the distribution did not disappear. The number of perfect scores increased from 141 to 311, and the rate of perfect scores per 1,000 candidates also increased. This pattern does not support a simple conclusion that the 2026 English paper was uniformly inaccessible. A more cautious interpretation is that the reported indicators show mixed tail movement: a larger lower-score group together with a more visible extreme upper group.

The term “polarisation” should be used cautiously here because the study does not calculate skewness, kurtosis, or full distributional density. The available indicators are sufficient to show lower central tendency, increased dispersion, and changes in selected tail outcomes, but they are not sufficient to establish the exact shape of the full score distribution. Future research using the full score-frequency table should calculate skewness and kurtosis and include histogram or density visualisation to determine whether the distributional pattern changed in a statistically describable way.

5.4. Implications for interpreting post-reform score distributions

The findings suggest that post-reform English score distributions should be interpreted through both statistical and policy lenses. Statistically, a lower mean score alone is insufficient. The 2026 distribution also showed a lower median and mode, increased dispersion, a larger below-5 group, and more perfect scores. These indicators together show that the change was not one-dimensional.

From a policy perspective, however, the most important issue is comparability. Because English was selected by candidates rather than taken by an identical compulsory population, year-to-year comparisons may be affected by changes in who entered the English examination. Public commentary should therefore avoid statements such as “students became weaker in English in 2026” or “the 2026 test was harder” unless additional evidence is available. More precise wording would be that “the reported English-examination cohort in 2026 had a lower central score location and wider score dispersion than the 2025 cohort.”

For assessment research, this means that score-distribution analysis should be combined with at least three additional forms of evidence. First, item-level facility and discrimination indices are needed to identify which parts of the English paper contributed to score spread. Second, curriculum-alignment analysis is needed to examine whether the linguistic and cognitive demands of the paper corresponded to the 2018 General Education Curriculum. Third, candidate-composition data are needed to determine whether changes in

score distribution reflected changes in subject-selection behaviour, candidate readiness, regional access, or admission strategy.

For English teaching, the findings have a more limited but practical implication. Teachers and schools should not treat the lower national mean as evidence that all learners require the same remedial response. The simultaneous increase in score dispersion and in the proportion of candidates scoring below 5 suggests that instructional planning may need to account for wider differences in learner readiness. Classroom- and school-level diagnostic evidence can therefore be used to identify learners who need additional support while maintaining appropriately challenging instruction for stronger learners. Because the present data are national aggregates, however, these teaching responses should be guided by local learner evidence rather than inferred directly from the national score distribution.

5.5. Contribution and limits of the study

The contribution of this study is not that it introduces a new method of distributional analysis. Rather, it provides an early post-reform account of the 2025-2026 English score distributions using officially reported national aggregate indicators. Its value lies in showing why a single headline figure, such as the mean score or the number of perfect scores, is insufficient for interpreting examination outcomes.

At the same time, the study remains limited by its data structure. It uses aggregate reported indicators rather than individual-level scores or item-level response data. It also compares only two adjacent post-reform years. Therefore, the study cannot determine whether the 2026 pattern was a temporary fluctuation, a consequence of candidate-composition change, a result of examination design, or part of a longer-term trend. These limitations do not invalidate the descriptive findings, but they restrict the claims that can be made from them.

Future studies should extend this work by incorporating the full score-frequency distribution, item-level statistics, subgroup information, and multi-year comparisons including pre-reform years such as 2023. Such research would allow stronger conclusions about whether the 2026 distribution was unusual relative to earlier patterns and whether the post-reform structure changed the population of candidates selecting English.

6. CONCLUSION

This study examined changes in the English score distribution of the Vietnamese National High School Graduation Examination from 2025 to 2026 using secondary aggregate data. The purpose was not to determine the cause of score changes, but to describe how the national distribution shifted across key indicators of central tendency, dispersion, and score-band outcomes. Because English was selected under the post-2025 examination structure, the comparison should be understood as a comparison of two reported English-examination cohorts rather than as evidence of change in an identical candidate population.

The findings show that the 2026 distribution differed from the 2025 distribution in a systematic way. Overall score location moved downward, while score spread became wider. At the same time, the distribution did not change uniformly across all performance levels. The

lower-score group became larger, but the number and rate of perfect scores also increased. This suggests that the 2026 English results should not be interpreted simply as evidence of a uniformly harder examination. Rather, they point to lower central tendency, wider dispersion, and mixed lower-tail and upper-tail movement among the reported English-examination cohort.

The study contributes to examination research by demonstrating the value of analysing public score-distribution data beyond headline figures. Instead of focusing only on the mean score or the number of perfect scores, the study considered multiple indicators together. This provides a clearer description of how national English examination outcomes changed between the two years and offers a useful basis for further assessment-related inquiry. In this sense, the study makes an empirical contribution by documenting the 2025-2026 shift, an analytical contribution by showing the value of multi-indicator distributional interpretation, and a policy-relevant contribution by encouraging more cautious public use of examination score data.

However, the findings must be interpreted within the limits of aggregate secondary data. The analysis cannot identify which test items, task types, candidate characteristics, or instructional factors contributed to the observed distributional pattern. It also cannot support causal claims about test difficulty, teaching quality, or curriculum alignment. The results should therefore be viewed as descriptive evidence of distributional change rather than an explanation of why that change occurred.

Future research should extend this analysis by using item-level statistics, multi-year score-distribution data, curriculum-alignment evidence, and learner-level data. Such evidence would allow researchers to move from describing national score patterns to explaining the mechanisms behind them. In this way, secondary score-distribution analysis can serve as a starting point for more comprehensive research on English assessment, educational equity, and high-stakes examination outcomes in Vietnam.

REFERENCES

[1]. American Educational Research Association. (2000). *AERA position statement on high-stakes testing in pre-K-12 education*. American Educational Research Association. https://www.aera.net/About-AERA/AERA-Rules-Policies/Association-Policies/Position-Statement-on-High-Stakes-Testing

[2]. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

[3]. Báo Điện tử Chính phủ. (2025, February 16). *Cấu trúc đề thi tốt nghiệp THPT áp dụng từ năm 2025*. Structure of the high school graduation examination applied from 2025. Cổng Thông tin điện tử Chính phủ. Retrieved November 1, 2026, from https://xaydungchinhsach.chinhphu.vn/cau-truc-de-thi-tot-nghiep-thpt-tu-nam-2025-119240308200554932.htm

[4]. Báo Điện tử Chính phủ. (2026, July 5). *Phổ điểm môn Tiếng Anh, kỳ thi tốt nghiệp năm 2026*. English score distribution in the 2026 graduation examination. Cổng Thông tin điện tử Chính phủ. Retrieved November 1, 2026, from https://xaydungchinhsach.chinhphu.vn/pho-diem-mon-tieng-anh-ky-thi-tot-nghiep-nam-2026-119260701100900533.htm

[5]. Bratti, M., Granato, S., & Havari, E. (2026). Another chance: Number of exam retakes and university students' outcomes. *European Economic Review*, 183, 105222. https://doi.org/10.1016/j.euroecorev.2025.105222

[6]. Department of Quality Management. (2023, December 29). *Cấu trúc định dạng đề thi tốt nghiệp THPT từ năm 2025.* Structure and format of the high school graduation examination from 2025. Ministry of Education and Training. Retrieved November 1, 2026, from https://vqa.moet.gov.vn/vi/news/thong-bao/cau-truc-dinh-dang-de-thi-tot-nghiep-thpt-tu-nam-2025-74.html

[7]. Duong, T. T. D., Huynh, U. T., & Pal, N. (2024). A statistical analysis of Vietnamese high school English test score data. *Vietnam Journal of Education, 8*(1), 1-17. https://doi.org/10.52296/vje.2024.334

[8]. Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73. https://doi.org/10.1111/jedm.12000

[9]. Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

[9]. Ly, T. L. (2023). Analysis of the distribution of English test scores in the Vietnamese National High School Graduation Examination in 2023. *International Journal of Research and Review, 10*(11), 367-374. https://doi.org/10.52403/ijrr.20231144

[10]. Ly, T. L., & Nguyen, T. N. (2023). A comparative analysis of English test scores in the Vietnamese National High School Graduation Examination during the period 2020-2023. *International Journal of Research and Review, 10*(12), 665-672. https://doi.org/10.52403/ijrr.20231267

[11]. Ministry of Education and Training. (2018). *Thông tư 32/2018/TT-BGDĐT ban hành Chương trình giáo dục phổ thông.* Circular No. 32/2018/TT-BGDĐT promulgating the general education curriculum. Retrieved November 1, 2026, from https://thuvienphapluat.vn/van-ban/Giao-duc/Thong-tu-32-2018-TT-BGDDT-Chuong-trinh-giao-duc-pho-thong-403454.aspx