International Journal of Research and Review
Vol. 10; Issue: 11; November 2023
Website: www.ijrrjournal.com
Research Paper *E-ISSN: 2349-9788; P-ISSN: 2454-2237*

Analysis of the Distribution of English Test Scores in the Vietnamese National High School Graduation Examination in 2023

Thi Loan Ly

Faculty of English Studies, East Asia University of Technology, Hanoi, Vietnam

DOI: https://doi.org/10.52403/ijrr.20231144

ABSTRACT

English has become the global language and plays an important role in order for people to communicate and fit into the real world. Learning English at high schools in Vietnam has become a necessary need of students. The aim of this paper is to analyse the distribution of English test scores in the Vietnamese National High School Graduation Examination in 2023. English test scores were first collected from 876,102 examinees and used for the analysis. The distribution of English test scores in 2023 was then identified using statistical analysis (histogram and descriptive statistics). It was found that the distribution of English test scores in the Vietnamese high school graduation examinations in 2023 was skewed to the right. Over 392,000 examinees accounting for 44.8% of the examinees, scored below 5 in which a total of 494 examinees had a perfect score of 10. A large number of examinees having scores of from 6.0 to 9.0 was also identified. English scored lowest among Vietnam's high school graduation examination subjects. Findings in this study suggest that, generally, the English tests in 2023 were suitable for the goal of graduation recognition.

Keywords: Distribution, English test scores, High school graduation examination, Vietnam.

INTRODUCTION

Vietnam is a multi-ethnic and multilingual country, with 54 different ethnic groups who speak more than 100 different languages. English is a foreign language in Vietnam, with no official status in government and law (1). English education in Vietnam has been booming in recent years as the country strives for further integration into the regional and global arena (2–4). As a result, increased attention has been paid to English assessment by various stakeholders including but not limited to teachers, educational administrators, researchers, policymakers as well as students and their parents (5). Some high schools offer bilingual programs for students to learn French, Chinese, and Japanese in addition to Vietnamese, but most students spend at least three years in high school studying English (6). In addition, according to the national strategy period 2011-2025 goal for Vietnamese EFL education, English is taught as a compulsory subject across primary to tertiary education (7). English is today without a doubt the most often used foreign language due to Vietnam's rapid economic integration with the rest of the world. It is a compulsory subject in the curriculum of secondary education, i.e., from Grade 6 (12 years old), across Vietnam and is also introduced as a subject from Grade 3 (9 years old) in many provinces (8). Increased awareness of the benefits of English proficiency has resulted in English becoming a medium of instruction in some areas of Vietnamese education (1). The importance of English in Vietnam is also evidenced by the increase of English language schools and private English centres across the nation and the significant financial commitment made by Vietnamese parents to pricey English-language programs. However, English language teaching (ELT) in Vietnam is

frequently of poor quality (9). It is, therefore, understanding the distribution of English test scores in the Vietnamese National High School Graduation Examination play an important role in the improvement of English study for students in Vietnam.

English is currently the most popular foreign language taught and learned in Vietnam. The Vietnamese government has implemented many changes to improve the English competencies of the Vietnamese people (10). It is a compulsory subject from Grade 3 to Grade 12 of the Vietnamese general education system (1). To complete Grade 12, according to the Ministry of Education and Training, more than one million high-school students across Vietnam took part in the national high school graduation exam for the academic year 2022- 2023 on June 27-30 (11). A total of 1,025,166 candidates registered for the exam, including 94.51 percent of students registering online and 5.49 percent of students registering directly. Of the total, the number of free candidates is 48,309. Around 73,232 students took the examination for graduation and 34,203 candidates registered for recruitment into universities. The number of students registered to take exams for both graduation and admission to universities and colleges was 917,731 (11). The annual high school graduation exam is the largest in Vietnam and is considered the most important. Its results will decide if a student can get into their desired university. English foreign language is one of the nine subjects that students took in 2023. Understanding the distribution of English test scores can help to improve English study for Vietnamese students. It is; therefore, this paper will attempt to analyse the distribution of English test scores in the Vietnam's National High School Graduation Examination in 2023 using statistical analysis such as histogram and descriptive statistics.

MATERIALS & METHODS

Materials

Over a million students participated in Vietnam's high school graduation examinations on June 28 and June 29 in 2023 were reported to take tests in nine subjects including Math, Literature, Physics, Chemistry, Biology, History, Geography, Civics Education and Foreign Language (12). Among these pupils, a total of 876,102 examinees took the English test. In this study, English test scores of these examinees were used to investigate the distribution of test scores of English subjects.

Methods

In this study, the histogram was first used to study the distribution of English test scores. Descriptive statistics (average score, median, number of examinees having ETSs below 1, below 5, and the average score achieved by the most examinees) were then summarized. compiled for percentage analysis. Finally, the average score for English was also compared with those of other subjects in the Vietnamese National High School Graduation Examination in 2023.

A histogram is an approximate representation of the distribution of numerical data. The term was first introduced by Karl Pearson (13). To construct a histogram, the first step is to "bin" (or "bucket") the range of values - that is, divide the entire range of values into a series of intervals, and then count how many values fall into each interval. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent and are often (but not required to be) of equal size (14). If the bins are of equal size, a bar is drawn over the bin with height proportional to the frequency. Histograms give a rough sense of the density of the underlying distribution of the data, and often for density estimation: estimating the probability density function of the underlying variable. The total area of a histogram used for probability density is always normalized to 1. If the length of the intervals on the x-axis are all 1, then a histogram is identical to a relative frequency plot. A histogram, on the other hand, is a graph that shows the distribution of numerical data, English test

scores in this study. It is a type of bar chart that shows the frequency or number of observations within different numerical ranges, called bins. The bins are usually specified as English test scores, non-overlapping intervals of a variable. The histogram provides a visual representation of the distribution of English test scores, showing the number of examinees that fall within each bin of English test scores. This can be useful for identifying patterns and trends in the English test scores.

However, bins need not be of equal width; in that case, the erected rectangle is defined to have its area proportional to the frequency of cases in the bin. The vertical axis is then not the frequency but frequency density-the number of cases per unit of the variable on the horizontal axis. Examples of variable bin width are displayed on Census bureau data below. As the adjacent bins leave no gaps, the rectangles of a histogram touch each other to indicate that the original variable is continuous. Histograms give a rough sense of the density of the underlying distribution of the data, and often for density estimation: estimating the probability density function of the underlying variable. The total area of a histogram used for probability density is always normalized to 1. If the length of the intervals on the x-axis are all 1, then a histogram is identical to a relative frequency plot. The histogram is one of the seven basic tools of quality control. Histograms are sometimes confused with bar charts. A histogram is used for continuous data, where the bins represent ranges of data, while a bar chart is a plot of categorical variables. Some authors recommend that bar charts have gaps between the rectangles to clarify the distinction. A bar graph and a histogram are two common types of graphical representations of data. While they may look similar, there are some key differences between the two that are important to understand. A bar graph is a chart that uses bars to represent the frequency or quantity of different categories of data. The bars can be either vertical or horizontal, and they are typically arranged either horizontally or vertically to make it easy to compare the different categories. Bar graphs are useful for displaying data that can be divided into discrete categories, such as the number of students in different grade levels at a school. A histogram, on the other hand, is a graph that shows the distribution of numerical data. It is a type of bar chart that shows the frequency or number of observations within different numerical ranges, called bins. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The histogram provides a visual representation of the distribution of the data, showing the number of observations that fall within each bin. This can be useful for identifying patterns and trends in the data, and for making comparisons between different datasets.

Histograms are good for showing general distributional features of dataset variables. It can be seen roughly where the peaks of the distribution are, whether the distribution is skewed or symmetric, and if there are any outliers.

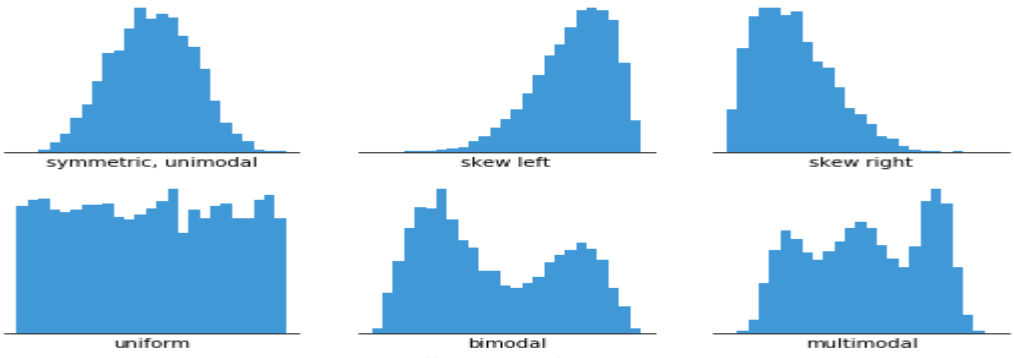

Figure 1. Different types of histograms.

Descriptive statistics:

Descriptive statistics is a branch of statistics that is concerned with describing the characteristics of the known data. Descriptive statistics provides summaries about either the population data or the sample data. Apart from descriptive statistics, inferential statistics is another crucial branch of statistics that is used to make inferences about the population data. Descriptive statistics can be broadly classified into two categories - measures of central tendency and measures of dispersion.

Descriptive statistics can be defined as a field of statistics that is used to summarize the characteristics of a sample by utilizing certain quantitative techniques. It helps to provide simple and precise summaries of the sample and the observations using measures like mean, median, variance, graphs, and charts. Univariate descriptive statistics are used to describe data containing only one variable. On the other hand, bivariate and multivariate descriptive statistics are used to describe data with multiple variables.

Descriptive statistics are simply the numerical procedures or graphical techniques used to organise and describe the characteristics or factors of a given sample (15). The main aim of descriptive statistics is to describe the midpoint of a spread of scores, usually referred to as the measure of central tendency, and the spread of scores known as the dispersion or variance (15). In addition, descriptive statistics are brief informational coefficients that summarize a given data set, which can be either a representation of the entire population or a sample of a population. Descriptive statistics are used to summarize data in an organized manner by describing the relationship between variables in a sample or population (16). Descriptive statistics are broken down into measures of central tendency and measures of variability (spread). Descriptive statistics include types of variables (nominal, ordinal, interval, and ratio) as well as measures of frequency, central tendency, dispersion/variation, and position (16). Measures of central tendency include the mean, median, and mode, while measures of variability include standard deviation, variance, minimum and maximum variables, kurtosis, and skewness. Among these measures, the mean and median have been most commonly used in many studies of quantitative research (17,18). In this study, commonly used descriptive statistics such as the mean and median were employed to measures of central tendency of English test scores.

Descriptive statistics, in short, help describe and understand the features of a specific data set by giving short summaries about the sample and measures of the data. The most recognized types of descriptive statistics are measures of center: the mean, median, and mode, which are used at almost all levels of math and statistics. The median can be viewed as the middle value for a set of numeric data. The median can be defined as the center-most observation that is obtained by arranging the data in ascending order. The mean, or the average, is calculated by adding all the figures within the data set and then dividing by the number of figures within the set. The arithmetic mean is calculated by summing all of the data values and dividing by the total number. It is normally called the mean or the average and is given by the formula:

$$x = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

All descriptive statistics are either measures of central tendency or measures of variability, also known as measures of dispersion. Measures of variability (or the measures of spread) aid in analyzing how dispersed the distribution is for a set of data. For example, while the measures of central tendency may give a person the average of a data set, it does not describe how the data is distributed within the set. If a sample is taken from a population, the sample standard deviation (SDEV) measures by how much the sample data deviates from the sample mean. The standard deviation is the positive square root of the variance. The standard deviation helps to analyze the variability in a data set in a more effective

manner as compared to the variance. It is calculated using the formula:

$$SDEV = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (2)$$

In descriptive statistics, the interquartile range (IQR) is a measure of statistical dispersion, which is the spread of the data. The IQR may also be called the midspread, middle 50%, fourth spread, or H-spread. It is defined as the difference between the 75th and 25th percentiles of the data. It is given by Q3−Q1, where:

$$Q_1 = \frac{(n+1)^{th}}{4} \quad (3)$$

$$Q_3 = \frac{3(n+1)^{th}}{4} \quad (4)$$

The Interquartile Range measures the range of the middle half of the data, and so is less affected by extreme observations. the interquartile range presents the spread of the middle half of data distribution.

RESULT & DISCUSSION

The distribution of English test scores

A histogram is a chart that plots the distribution of a numeric variable's values as a series of bars. Each bar typically covers a range of numeric values called a bin or class; a bar's height indicates the frequency of data points with a value within the corresponding bin. As discussed above, a histogram is a graph that shows the distribution of data on English test scores. It is a type of bar chart that shows the frequency or number of examinees within different numerical ranges, called bins. The histogram provides a visual representation of the distribution of the English test scores, showing the number of examinees that fall within each bin. Data from the histogram shown in Figure 2 illustrate that the distribution of English test scores in the Vietnamese high school graduation examinations in 2023 was skewed to the right. This skewness determines whether the data's distribution is symmetrical. Skewness measurement determines whether most of the distribution values lie to the left or the right of the mean. The skewness of normal distribution is zero, showing an equal amount of the data on either side of the mean. The right-skewed distribution of the English test scores means that the number of students having high scores were more than the number of students having low scores. With right-skewed distribution (also known as "positively skewed" distribution), most data on the English test scores in 2023 failed to the right, or positive side, of the graph's peak. In this case, the histogram of English test scores skews in such a way that its right side (or "tail") is longer than its left side (as shown in Figure 2).

Data from Figure 2 and Table 1 illustrate that, on a right-skewed histogram English test scores, the mean and median, and mode was different. In this case, the average was the higher point of the histogram, whereas the median and mean fall to the right of it. Data on English test scores skewed to the right is a result of a large number of examinees having scores above the average.

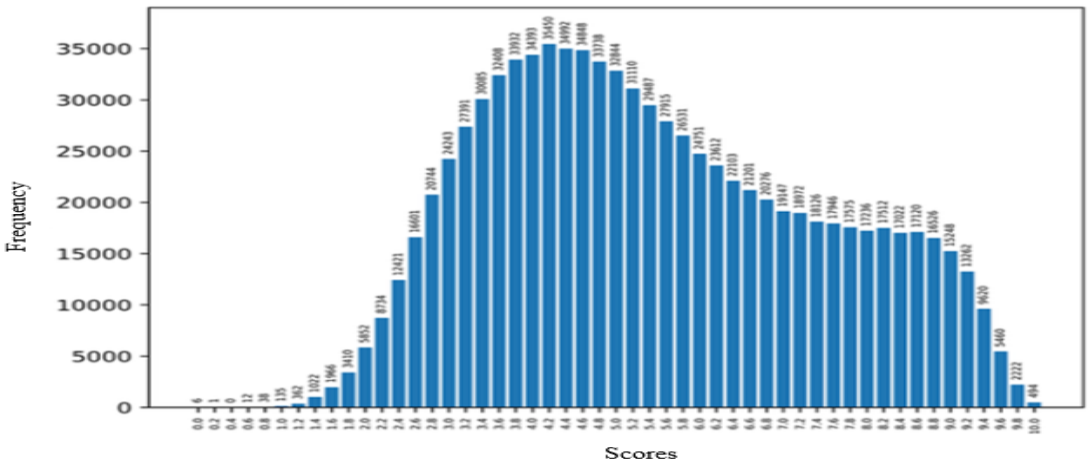

Figure 2. Histogram of English test scores.

Frequency is a count of the number of times a specific event occurs or is a number of events occurring during a given time interval (19). A frequency histogram is a bar graph that shows the frequency that data occurred within certain ranges. In this study, the frequency of English test scores was statistically summarised in Table 1. Data from Table 1 also demonstrate that the number of examinees having English scores below 5 was less than the number of examinees having English scores above 5. It can be seen from Figure 2 that the average score that most students scored was in the range of between 4.0 and 4.4. The average score achieved by the most examinees was 4.2 corresponding with 35,450 students.

Table 1. Frequency of English test scores.

Scores	Freq.	Scores	Freq.	Scores	Freq.	Scores	Freq.	Scores	Freq.
0.0	6	2.0	5,852	4.0	34,393	6.0	24,751	8.0	17,236
0.2	1	2.2	8,734	4.2	35,450	6.2	23,612	8.2	17,512
0.4	0	2.4	12,421	4.4	34,992	6.4	22,103	8.4	17,022
0.6	12	2.6	16,601	4.6	34,848	6.6	21,201	8.6	17,120
0.8	38	2.8	20,744	4.8	33,738	6.8	20,276	8.8	16,526
1.0	135	3.0	24,243	5.0	32,844	7.0	19,147	9.0	15,248
1.2	362	3.2	27,391	5.2	31,110	7.2	18,972	9.2	13,262
1.4	1,022	3.4	30,085	5.4	29,487	7.4	18,126	9.4	9,620
1.6	1,966	3.6	32,408	5.6	27,915	7.6	17,946	9.6	5,460
1.8	3,410	3.8	33,932	5.8	26,531	7.8	17,575	9.8	2,222
								10	494

Analysis of descriptive statistics of English test scores

Descriptive statistics can help describe and understand the features of a specific data set by giving short summaries about the sample and measures of the data. The mean, or the average, is calculated by adding all the English test scores within the data set and then dividing by the number of examinees within the set. According to results announced by the Ministry of Education and Training of Vietnam, a total of 876,102 examinees took the English test as shown in Table 2. Data from Table 2 illustrate that the average and median scores for English were 5.45 and 5.2, respectively. The average score achieved by the most examinees was 4.2 with 35,450 examinees. The number of examinees having scores of less than 1 was 192 (accounting for 0.022%), whereas the number of examinees having scoring less than 5 is 392,784 (accounting for 44.83%).

Table 2. Descriptive statistics of English test scores.

Number of examinees	Number of examinees having ETSs <=1	Number of examinees having ETSs < 5	Average score	The average score achieved by the most examinees	Median
876,102	192	392,784	5.45	4.2	5.2
	0.022 %	448,33			

Comparison of English test scores with others subjects

Data from Figure 3 illustrates English scores lowest among Vietnam's high school graduation examination subjects. The three subjects with the highest average scores were Civics Education (8.29), Literature (6.86) and Chemistry (6.74). A total of 876,102 examinees took the English test in 2023. Among those who were tested for English, over 392,000 examinees, or 44.8% of the examinees, scored below 5. A total of 494 examinees had a perfect score of 10 as shown in Figure 2 and Table 1. In 2022, the average English score for over 978,000 examinees was 5.15, higher than 5.02 in Biology, although Biology's mode score last year was 4.5 and English's 3.8 (12). In 2021, a total of 866.993 examinees were reported to take the English test the whole country. The average score of this exam was 5.84 in which the score achieved by the most examinees was 4.0. There were 144 examinees having paralysis scores (from 1 or less). In addition, there were 349.175 examinees having scores below 5 (20).

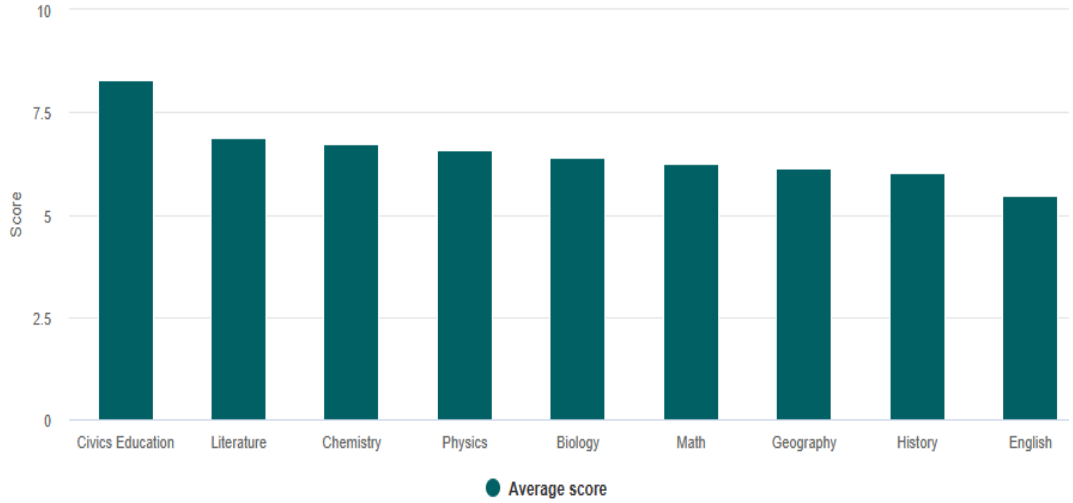

Figure 3. Average scores by subjects in Vietnam's high school graduation exams in 2023.

It can be seen that the number of examinees having scores above 8 was 131,722 in 2023, accounting for 15.03% of the total examinees. The percentages of 11.9% and 18.3% were reported in 2022 and 2021, respectively.

CONCLUSION

This paper was set out to analyse the distribution of English test scores in the Vietnamese National High School Graduation Examination in 2023. A dataset on English test scores was first collected from 876,102 examinees and used for the analysis in this study. The distribution of English test scores in 2023 was then analysed using statistical analysis (histogram and descriptive statistics). The study results showed that the distribution of English test scores in the Vietnamese high school graduation examinations in 2023 was skewed to the right. Over 392,000 examinees accounting for 44.8% of the examinees, scored below 5 in which a total of 494 examinees had a perfect score of 10. The number of examinees having scores of less than 1 was 192 (accounting for 0.022%), whereas the number of examinees having scoring less than 5 is 392,784 (accounting for 44.83%). In addition, a large number of examinees having scores of from 6.0 to 9.0 was also identified. English scored lowest among Vietnam's high school graduation examination subjects. It can be concluded that the English tests were suitable for the goal of graduation recognition in the Vietnamese National High School Graduation Examination in 2023.

Declaration by Authors

Acknowledgement: The authors thank editors and the anonymous reviewers for their careful reading of the manuscript and their many insightful comments and suggestions

Source of Funding: None

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

REFERENCES

1. Sundkvist P, Nguyen XNCM. English in Vietnam. Handb Asian Englishes. 2020;683–703.
2. Ngo XM. Sociopolitical contexts of EFL writing assessment in Vietnam: Impact of a national project. In: The politics of English second language writing assessment in global contexts. Routledge; 2018. p. 47–59.
3. Ngo XM. Vietnam's trillion-dong attempt to reform English education: A laudable reform or a costly failure? English today. 2021;37(2):115–9.
4. Nguyen H, Gu Y. Impact of TOEIC listening and reading as a university exit test in Vietnam. Lang Assess Q. 2020;17(2):147–67.
5. Ngo XM. English assessment in Vietnam: status quo, major tensions, and underlying ideological conflicts. Asian Englishes. 2022;1–13.
6. Truong TNN, Wang C. Understanding Vietnamese college students' self-efficacy

beliefs in learning English as a foreign language. System. 2019;84:123–32.
7. Hien TTT, Loan MT. Current challenges in the teaching of tertiary English in Vietnam. In: English tertiary education in Vietnam. Routledge; 2018. p. 40–53.
8. Vu D Van, Peters E. Vocabulary in English language learning, teaching, and testing in Vietnam: A review. Educ Sci. 2021;11(9):563.
9. Nguyen XNCM. Language education policy in Vietnam. In: The Routledge international handbook of language education policy in Asia. Routledge; 2019. p. 185–201.
10. Xuan Mai L, Thanh Thao L. English language teaching pedagogical reforms in Vietnam: External factors in light of teachers' backgrounds. Cogent Educ. 2022;9(1):2087457.
11. SGGP. Over 1 million students to sit for national high school graduation exam in June [Internet]. 2023. Available from: https://vietnamnet.vn/en/over-1-million-students-to-sit-for-national-high-school-graduation-exam-in-june-2143148.html
12. Anh P. English scores lowest among Vietnam's high school graduation examination subjects [Internet]. 2023. Available from: https://e.vnexpress.net/news/education/average-english-scores-lowest-among-vietnam-s-high-school-graduation-examination-subjects-4631240.html#:~:text=The three subjects with the,) and Chemistry (6.74).&text=Among those who were tested,a perfect score of 10.
13. Pearson K. X. Contributions to the mathematical theory of evolution.—II. Skew variation in homogeneous material. Philos Trans R Soc London(A). 1895;(186):343–414.
14. Howitt D, Cramer D. Introduction to statistics in psychology. Pearson education; 2008.
15. Fisher MJ, Marshall AP. Understanding descriptive statistics. Aust Crit care. 2009;22(2):93–7.
16. Kaur P, Stoltzfus J, Yellapu V. Descriptive statistics. Int J Acad Med. 2018;4(1):60–3.
17. Phinyomark A, Thongpanja S, Hu H, Phukpattaranont P, Limsakul C. The usefulness of mean and median frequencies in electromyography analysis. Comput Intell Electromyogr Anal Perspect Curr Appl Futur challenges. 2012;23:195–220.
18. Rodbard D. Glucose time in range, time above range, and time below range depend on mean or median glucose or HbA1c, glucose coefficient of variation, and shape of the glucose distribution. Diabetes Technol Ther. 2020;22(7):492–500.
19. Gagnon DR, Doron-LaMarca S, Bell M, O'Farrell TJ, Taft CT. Poisson regression for modeling count and frequency outcomes in trauma research. J Trauma Stress. 2008;21(5):448–54.
20. Newspaper I. Answers for reference in English for high school graduation 2023 [Internet]. Available from: https://www.vietnam.vn/en/dap-an-tham-khao-mon-tieng-anh-tot-nghiep-thpt-2023/

How to cite this article: Thi Loan Ly. Analysis of the distribution of English test scores in the Vietnamese national high school graduation examination in 2023. *International Journal of Research and Review*. 2023; 10(11): 367-374. DOI: *https://doi.org/10.52403/ijrr.20231144*

******