

# An Analysis of the Distribution of VNU High School Competency Assessment Results

Thi Nga Le<sup>1</sup>, Thi Loan Ly<sup>2</sup>

<sup>1,2</sup>Faculty of English Studies, East Asia University of Technology, Hanoi, Vietnam

Corresponding author: Thi Nga Le

DOI: <https://doi.org/10.52403/ijrr.20240158>

## ABSTRACT

Assessment is an essential feature of the competency-based educational model because only by means of evaluation we can verify achievement of specified learning outcomes. This is especially important in the context of professions education for high school students. This study aims to investigate the distribution of VNU HSA results. The HSA cores collected from 87,095 examinees taking part in the VNU HSA exam in 2023. Descriptive statistics will be first used to measure central tendency including the mean and median and variability including standard deviation, minimum and maximum variables. Histograms are then used to identify the distribution of VNU HSA results. Finally, the study results and main findings will be discussed and summarized. VNU HSA results showed that the lowest and highest VNU HSA scores were 31 and 133, respectively. Whereas, the VNU HSA values of the mean and median were 77.1 and 77.0, respectively. VNU HSA scores in 2023 were normally distributed. The mean value of 29.5 was highest the section of qualitative thinking, followed by 26.5 and 21.2 for sections of quantitative thinking, and natural and social sciences, respectively. Findings in this study provide an insight how effectively improve the high school student competency in Vietnam.

**Keywords:** Relationship, English Test Scores, Vietnamese National High School Graduation Exam, VNU Competency Assessment Results.

## INTRODUCTION

Many studies define "competency" (also known as ability, capacity, capability, etc.) as the subject's capacity to do a particular task. This capacity manifests in the subject's

physical, psychological, knowledge, and skill characteristics. Numerous Vietnamese linguists define "competence" as the ability, subjective or natural conditions available to perform a certain activity; it may be also a psychological and physiologic quality that enables a person to perform a certain type of activity with high quality" (1). As stated in the Encyclopedia, "competence is the ability to be formed and developed, allowing people to achieve success in a physical, intellectual, or occupational activity; It is expressed in the ability to perform an activity or a task" (2). A study by (3) has the same view, noting that "competency is understood as a set of knowledge, attitudes, skills, or thinking strategies that are core and important for creating important outputs." Therefore, "competency" is defined as the knowledge, skills, and basic values that each individual demonstrates in their actions (4).

The higher education system in Vietnam is made up of universities, colleges, and academies structured according to their industry of expertise or management hierarchy (local and central). Vietnam now has 224 universities, with over 30% of those specializing in social sciences, according to the statistics from the Ministry of Education and Training (5). Most universities in Vietnam use the results of the Vietnamese National High School Graduation Exam to evaluate the abilities of high school students for the university entrance (6,7). This is a nationally standardized exam that determines whether students may graduate from high school and attend post-secondary education

(8). For many Vietnamese students, the national high school graduation examination is one of the most important exams in their lives, transitioning from upper-secondary education to tertiary education (9). While the importance of the exam is significant in the country where a degree is inevitable for a stable job in the future, policies on the national examination have significantly changed since 2015 (9). However, along with that, the High school Competency Assessment exam organized by Vietnam National University, Hanoi and Vietnam National University, Ho Chi Minh city, is also used as a reliable assessment result of high school students' capacity as a basis for taking the entrance exam to a prestigious university in Vietnam.

The High School Competency Assessment (HSA) exam of Vietnam National University, Hanoi (VNU) took place in eight exam sessions from March to June, 2023 (10). Examinees were from Hanoi city and northern provinces such as Thai Nguyen, Hung Yen, Thanh Hoa, Nam Dinh, Hai Phong, and Nghe An. Candidates registered for the HSA exam. In 2023, each candidate can register to take the exam in a maximum of two times. The time between two consecutive times was at least 28 days apart. The number of registered candidates was 90,045, of which 29,110 candidates registered to take the exam a second time (accounting for about 32%). A total of 37% candidates were from Hanoi; followed by Nam Dinh (7%); Thai Binh, Hai Duong, Nghe An, Thanh Hoa (5%/province); Hai Phong, Hung Yen, Bac Ninh, Vinh Phuc (4%/province); Bac Giang, Ninh Binh, Quang Ninh, Phu Thi, Ha Nam, and Thai Nguyen (2-3%/province) etc. A total of 96.7% of candidates successfully completed the VNU HSA exam. This study aims to investigate the distribution of VNU HSA results. The HSA cores collected from 87,095 examinees taking part in the VNU HSA exam in 2023. Descriptive statistics will be first used to measure central tendency including the mean and median and variability including standard deviation,

minimum and maximum variables. Histograms were then used to identify the distribution of VNU HSA results. Finally, the study results and main findings will be discussed and summarized.

## **MATERIALS & METHODS**

### **Materials**

VNU HSA exam aims to evaluate high school students' abilities in 3 main groups: (i) creativity and problem solving; (ii) competence in Math, Vietnamese, linguistic thinking, reasoning, logic, calculation and data processing; (iii) self-study, discovery and application of technology/science (natural - social) (11). The HSA test structure includes 3 section: qualitative thinking (50 questions, 60 minutes), quantitative thinking (50 questions, 75 minutes), and natural and social sciences (50 questions, 60 minutes). Total number of questions is 150. Time to take the HSA exam for high school students is 195 minutes. Candidates take the test on a computer and complete it in one exam session. In this study, VNU HSA cores collected from 87,095 examinees in 2023 were employed to investigate the distribution of VNU HSA results.

### **Methods**

#### ***Descriptive statistics:***

Descriptive statistics seek to describe the midpoint of a spread of scores, called the measure of central tendency, and the spread of scores which is called the dispersion, of which variance is an example (12). 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 (12). 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 (13). 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 (13). 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 (14,15). In this study, commonly used descriptive statistics such as the mean and median were employed to measures of central tendency of HSA scores. 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:

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

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 Interquartile Range (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  $Q_3 - Q_1$ , 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.

### Histograms

A histogram is a visual representation of the distribution of numeric data. The term was first introduced by Karl Pearson (16). To construct a histogram, the first step is to "bin" (or "bucket") the range of values - 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 (17). If the bins are of equal size, a bar is drawn over the bin with height proportional to the frequency - the number of cases in each bin. A histogram may also be normalized to display "relative" frequencies showing the proportion of cases that fall into each of several categories, with the sum of the heights equaling 1. 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 (18). 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 (19).

A histogram is a type of chart that shows the frequency distribution of data points across a continuous range of numerical values (20). The values are grouped into bin or buckets that are arranged in consecutive order along the horizontal x-axis at the bottom of the chart. Each bin is represented by a vertical bar that sits on the x-axis and extends upward to indicate the number of data points within that bin. Histograms were first conceived as a visual aid to statistical approximations (21). Even today this point is still emphasized in the common conception of histograms: Webster's defines a histogram as "a bar graph of a frequency distribution in which the widths of the bars are proportional to the classes into which the variable has been divided and the heights of the bars are proportional to the class frequencies". Histograms, however, are extremely useful even when disassociated from their canonical visual representation and treated as purely mathematical objects capturing data distribution approximations

In a histogram, consider a relation  $R$  with  $n$  numeric attributes  $X_i$  ( $i = 1..n$ ). The value set  $V_i$  of attribute  $X_i$  is the set of values of  $X_i$  that are present in  $R$  (21). Let  $V_i = \{v_i(k) : 1 \leq k \leq D_i\}$ , where  $v_i(k) < v_i(j)$  when  $k < j$ . The spread  $s_i(k)$  of  $v_i(k)$  is defined as  $s_i(k) = v_i(k + 1) -$

$v_i(k)$ , for  $1 \leq k < D_i$ . (We take  $s_i(D_i) = 1$ .) The frequency  $f_i(k)$  of  $v_i(k)$  is the number of tuples in  $R$  with  $X_i = v_i(k)$ . The area  $a_i(k)$  of  $v_i(k)$  is defined as  $a_i(k) = f_i(k) \times s_i(k)$ . The data distribution of  $X_i$  is the set of pairs  $T_i = \{(v_i(1), f_i(1)), (v_i(2), f_i(2)), \dots, (v_i(D_i), f_i(D_i))\}$ . The joint frequency  $f(k_1, \dots, k_n)$  of the value combination  $\langle v_1(k_1), \dots, v_n(k_n) \rangle$  is the number of tuples in  $R$  that contain  $v_i(k_i)$  in attribute  $X_i$ , for all  $i$ . The joint data distribution  $T_1, \dots, T_n$  of  $X_1, \dots, X_n$  is the entire set of (value combination, joint frequency) pairs (21).

Histograms can be of different kinds based on the frequency distribution of the data. There are different types of distribution, such as normal distribution, skewed distribution, bimodal distribution, and multimodal distribution (22). The histograms are named based on the type of data they are representing into the following types: uniform, bimodal, right and left skewed histograms as shown in Figure 1. A uniform histogram a histogram where all the bars are almost of the same height, contain the same number of elements, or have a uniform distribution. A bimodal histogram has two peaks. These points are not necessarily of the same height. The two peaks mean that the two groups in the given frequency table have the most occurrences. A right skewed histogram is skewed to the right by the name. It happens when most data is concentrated towards the end of the distribution. Whereas, a left skewed histogram a histogram that is skewed to the left. Such type of graph is found when most of the data is concentrated towards the start of the distribution.

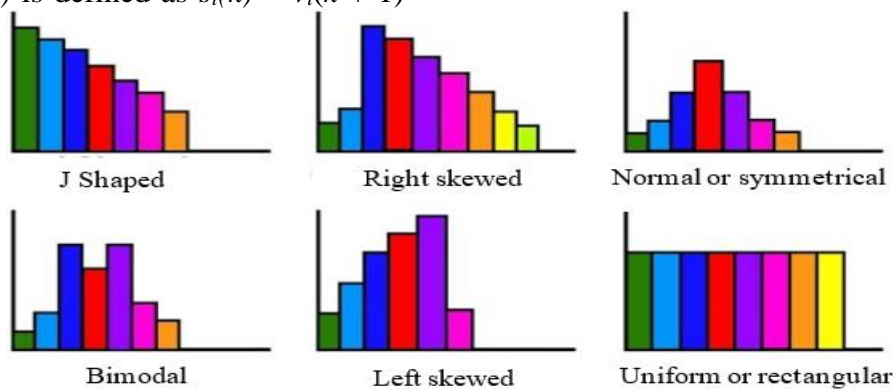


Figure 1. Main types of histograms.



## RESULTS & DISCUSSION

### Analysis of VNU HSA results

Statistics of HSA exam results in 2023 show that the lowest and highest scores were 31 and 133/150, respectively. The values of the mean and median were 77.1 and 77.0, respectively. The standard deviation was 14.0. It can be seen that HSA score follows a normal distribution (as shown in Figure 2) with mean and median values close to each other. The highest score belonged to the candidate from Hung Yen province. Whereas, the male student having the second highest score of reached 129 was from Vinh Phuc province. Another female student having the third highest score of 128 was from Thai Binh province. The fourth highest score of 128 was in Hanoi. All four top candidates took the exam twice, the first (or second) score was in the range of from 113 to 119, followed by seven candidates with the

same score of 125. A total of 58 candidates were found with scores above 120. 1.09% of the total candidates having scores of above  $\geq 110$ . 6.0% of candidates having scores of  $\geq 100$ , 19.3% achieving scores  $\geq 90$ , 27.8% of candidates having scores  $\geq 85$ ; 42.3% scored  $\geq 80$ ; 56.1% having scores  $\geq 75$ . The distribution of HSA test scores in 2023 was similar to those obtained in 2022. This year's average score decreased slightly when comparing with those obtained in 2022. This is consistent with the distribution HSA score in the March-April 2023 exams. The reason for the slight decrease in HSA scores was due to the scale of the HSA exam in 2023 increased nearly 1.5 times. In addition, the time to organize the exams was in 2.5 months (from March 10 to June 4, 2023). Whereas, the maximum number of candidates was reduced to 2 times/year.

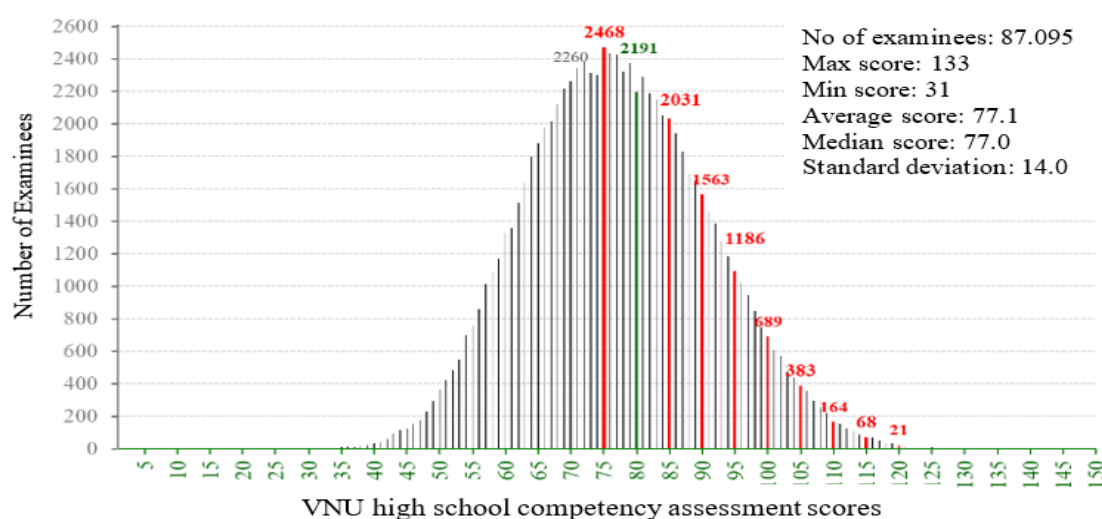


Figure 2. A histogram of VNU HSA scores.

### Analysis of VNU HSA scores for section of qualitative thinking

Statistics of exam results for the section of quantitative thinking (mathematics and data processing) in 2023 show that the lowest and highest scores were 50 and 06. Meanwhile, the average score and the median was 26.5 and 26.0. The standard deviation was 8.0. The distribution of HSA score for the section of Quantitative Thinking shows that the data follows a normal distribution with the mean

and median values of being relatively close to each other (only a difference of 0.5). Meanwhile, the standard deviation was 8.0. This indicates that a relatively large dispersion was found in the results obtained from the section of quantitative thinking. Data from the histogram in Figure 3 also demonstrates that a large number of candidates had test scores in the range between 21 and 27.

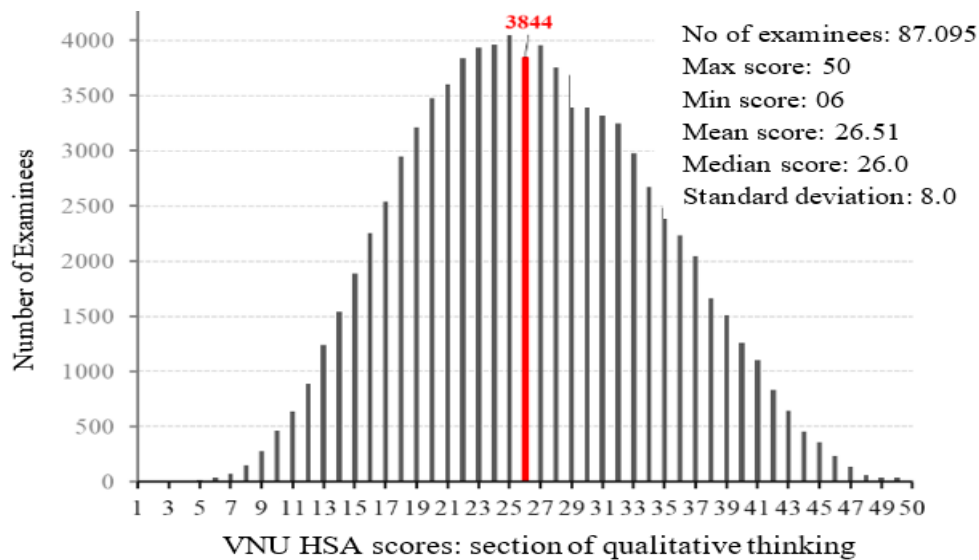


Figure 3. A histogram of VNU HSA scores: section of qualitative thinking.

### Analysis of VNU HSA scores for section of quantitative thinking

Statistical data of HSA test results in 2023 for the qualitative thinking section (literature and language) was shown in the histogram in Figure 4. Data from Figure 4 shows that the lowest and highest scores of the section was 14 and 46, respectively. Meanwhile, the mean and median scores were 29.5 and 30.0, respectively. It can be concluded that, when comparing with those obtained in section one, the mean score of candidates in section two was higher than those obtained in section one. The standard deviation was 4.7. The

distribution of HSA test scores for the Qualitative Thinking section shows that the HSA scores are normally distributed with the mean and median values of being relatively close to each other (only a difference of 0.5). However, the standard deviation of HSA score in section two was smaller than that of section two ( $4.7 < 8.0$ ). This indicates that the HSA scores in the Qualitative Thinking section were less dispersed. Data from the histogram in Figure 4 illustrate there was a large number of candidates having HSA test scores ranging from 25 to 33.

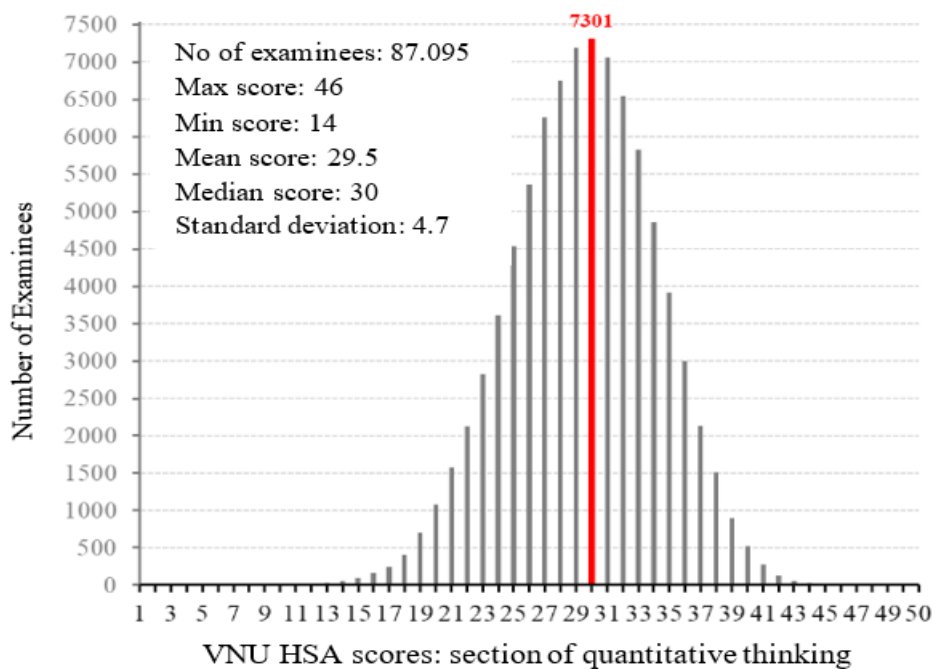


Figure 4. VNU HSA scores: section of quantitative thinking.

### Analysis of VNU HSA scores for section of natural and social sciences

The distribution of HSA exam results in 2023 for the section of natural and social sciences was shown in the histogram in Figure 5. The data in Figure 5 shows the lowest and highest scores were 07 and 45, respectively. Meanwhile, the mean and median scores were 21.2 and 21.0, respectively. It can be seen that the mean HSA test score of candidates in section three was the lowest when comparing those obtained in the first two sections. The value of standard deviation was 5.3. In addition, the distribution of HSA

test scores for the Science section shows that, similar to those obtained in the two test sections above, the HSA test scores were also normally distributed. The mean and median values were relatively close to each other (only a difference of 0.2). However, the standard deviation in the section three was larger than that of section two ( $5.3 > 4.7$ ). This shows that the HSA test results in the section three have a higher degree of dispersion than that of the section two. Data from the histogram in Figure 5 also illustrates that a large number of candidates had HSA test scores in the range of from 18 to 24.

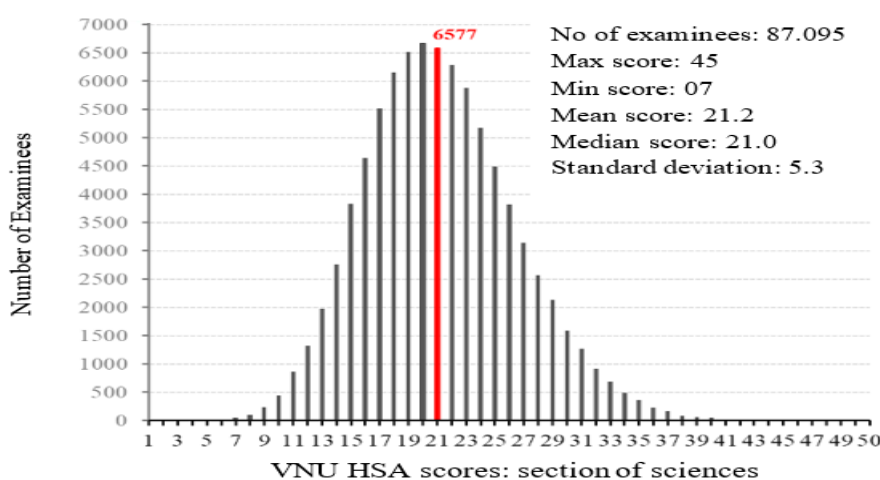


Figure 5. VNU HSA scores: section of sciences.

### CONCLUSIONS

This study was set out to study the distribution of VNU HSA results. The HSA cores collected from 87.095 examinees taking part in the VNU HSA exam in 2023. Descriptive statistics will be first used to measure central tendency including the mean and median and variability including standard deviation, minimum and maximum variables. Histograms are then used to identify the distribution of VNU HSA results. Finally, the study results and main findings will be discussed and summarized. VNU HSA results showed that the lowest and highest VNU HSA scores were 31 and 133, respectively. Whereas, the VNU HSA values of the mean and median were 77.1 and 77.0, respectively. VNU HSA scores in 2023 were normally distributed. The mean value of 29.5 was highest the section of qualitative thinking, followed by 26.5 and 21.2 for

sections of quantitative thinking, and natural and social sciences, respectively. Findings in this study provide an insight how effectively improve the high school student competency in Vietnam. The analysis of the distribution of the VNU HSA scores plays an important role in the improvement the quality of the HSA exam. This can help to understand and improve existing problems in high school students' abilities.

#### Declaration by Authors

**Acknowledgement:** The authors thank VNU, Center for Educational Testing for providing the data, 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. Phe H. Vietnamese Dictionary. Da Nang Publishing House; 1997.
2. EPH. Education Dictionary. Encyclopedia Publishing House;
3. Lam NH. Capacity model in education, training, and human resource development. 2007.
4. Nga PVTY, An MAVTT, Hang MATTT, Thuy MAVTT, Tuyen MAHTM. Foreign Language Competence Of Vietnamese Students-A Case Study At The National Academy Of Public Administration. *J Namibian Stud Hist Polit Cult.* 2023;33:5809–22.
5. MOET. List of educational institutions [Internet]. Available from: <https://moet.gov.vn/cosogiaduc/Pages/danh-sach.aspx>
6. Van Le D, Tran TQ. Economic growth and quality of education: Evidence from the national high school exam in Vietnam. *Int J Educ Dev.* 2024;104:102947.
7. Nguyen P, Nguyen P, Bruneau P, Cao L, Wang J, Truong H. Evaluation of Mathematics Performance of Google Bard on The Mathematics Test of the Vietnamese National High School Graduation Examination. *Authorea Prepr.* 2023;
8. Nguyen C. Nationally standardized policy and locally interpreted implementation: How Vietnamese school leaders enact education reform. *Int J Leadersh Educ.* 2023;26(3):397–426.
9. BAO N Van. Exploring the effects of perceived justice and motivation on student satisfaction regarding entrance exams for higher education. *KDI School;* 2021.
10. Tiên NB, Nguyen H-N, Le H-H, Trang TT, Van Dinh C, Nguyen H-N, et al. Applying Machine Learning approaches to predict High-school Student Assessment scores based on high school transcript records. *Int J Internet, Broadcast Commun.* 2023;15(2):261–7.
11. Yen GT, Huong NTQ. Vietnamese Students' Foreign Language Competence. *Migr Lett.* 2023;20(8):341–8.
12. Fisher MJ, Marshall AP. Understanding descriptive statistics. *Aust Crit care.* 2009;22(2):93–7.
13. Kaur P, Stoltzfus J, Yellapu V. Descriptive statistics. *Int J Acad Med.* 2018;4(1):60–3.
14. 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.
15. 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.
16. 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.
17. Howitt D, Cramer D. Introduction to statistics in psychology. Pearson education; 2008.
18. Freedman DA. Statistical models: theory and practice. cambridge university press; 2009.
19. Stangor C. Research methods for the behavior science. Cengage Learning; 2011.
20. Jagadish H V, Jin H, Ooi BC, Tan K-L. Global optimization of histograms. *ACM SIGMOD Rec.* 2001;30(2):223–34.
21. Ioannidis Y. The history of histograms (abridged). In: *Proceedings 2003 VLDB Conference.* Elsevier; 2003. p. 19–30.
22. Roh YJ, Kim JH, Chung YD, Son JH, Kim MH. Hierarchically organized skew-tolerant histograms for geographic data objects. In: *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data.* 2010. p. 627–38.

How to cite this article: Thi Nga Le, Thi Loan Ly. An analysis of the distribution of VNU high school competency assessment results. *International Journal of Research and Review.* 2024; 11(1): 521-528.  
DOI: <https://doi.org/10.52403/ijrr.20240158>

\*\*\*\*\*