Biplot analysis for marking the quality of schools which compatible with national examination score

F Fahrurrazi1*, R Rohiat2, M Somantri2 and T Turdja'i2

1Doctorate Programme Student, University of Bengkulu, Indonesia
2Department of Management Education, University of Bengkulu, Indonesia.

*nahdah63@yahoo.co.id

Abstract. The issue of national examinations has always been a matter of debate, because the scores are only used to determine the passing scores of students without follow-up. This study aims to map the quality of schools in accordance with national exam scores, and describe the relationship between national exam scores and influential variables. Data were taken from 22 high schools in Malang and Malang regencies with a population of 5396 people and a sample of 359 people. Data were analysed with a biplot package based on Mathematica programming, by combining scores and national deviation examination standards. Mapping the quality of schools is compatible with the national exam scores. It is obtained by the biplot with standard variables. The results obtained show that there is the compatibility of the rankings, and there is a significant positive correlation between all schools.

1. Introduction

The school quality is the result which achieved by school performance. In the academic field, the school quality is related with the quality of school graduates. The indicator of the quality of school graduates generally use the results of student achievement in the National Examination (UN), the National Final Examination (UAN) and School Final Examination (UAS), related to the quality of school graduates (output), it can be explained that school output is said to be of high quality, if school achievement, especially student learning achievement, shows high achievement in academic ability results. They are exam scores such as UAN or UAS.

Responding to these problems, in the 2017/2018 Academic Year, the government was still holding the National Examination by making a number of changes at the elementary level, while junior and senior high schools were still carrying out the National Examination as like as previous years. Based on Government Regulation No.19 of 2005 and Minister of National Education Regulation No. 37/2007 in the 2017/2018 Academic Year, the UASBN was be held by elementary / MI / SDLB students. There was a change in objectives in the UASBN. The results of the final exam were used as a tool to map the education unit. In other word, the decision of student's graduation, the criteria of graduation were set by each school / madrasah whose students attending the UASBN. This research will obtain a description of the mapping and the weaknesses and the advantages of each school based on the variables which influencing the UN score.

Principal's leadership is the principal's ability to influence the behaviour of teachers and students to achieve school goals [1]. The school principal who successful and effective in his leadership is
indicated to influence achievement in learning achievement. The Characteristics of successful school principals include: 1) having a strong vision of the future of the school and encouraging students, and their staff to work for realizing the vision, 2) having high expectations of student achievement and staff performance, 3) monitoring teachers in the classroom and providing some inputs to solve problems in teaching improvement, 4) Being able to create a safe learning environment [2]. While the categories of effective headmaster include: 1) being able to increase awareness and having an active role about the need of school improvement and high achievement, 2) being able to monitor the progress of student achievement, 3) being able to create a reward system for successful students and teachers, 4) being able to obtain learning equipment’s and resources, 5) being able to create safe learning environment and orderly [3].

The selection process for entering high school generally only uses the National Examination (UN) SMP. The average score of students’ national examination are different. Each school is different, this is possible because it depends on the quality of the school and the score standard that have been formed in the previous years. The quality of schools is often related with the score of national examination that students accepted [4,5]. The quality of high school graduates produced is generally in line with the national examination of students at the junior high school level.

Boxplot diagrams are one of some techniques to give description about the location of centralized data, the range of spreading and the sloping of the distribution pattern. The description is shown in the form of a box (rectangle) on both sides of the line. The length of the boxes and lines is based on a summary of 5 digit: the minimum score, the first quartile, the median or the second quartile, the third quartile, and the maximum score of the sorted data [6].

![Boxplot Diagram](image)

Figure 1. The summary of position 5 numbers in a line box diagram.

The position of Q1 and Q3 limits the box while the median (Me) is inside the box, it shows that 50% of data is spreading in the box and the rest is divided equally around the top and bottom lines of the box. Data located above the Upper Limit (BA) or below the Lower Limit (BB) is called outlier, with BB = Q1 - 1.5 (Q3 - Q1) and BA = Q3 + 1.5 (Q3 -Q1) [7].

Biplot analysis was introduced by [4]. This analysis is known as one of the statistical techniques by presenting through graphs from the matrix of data into a plot by combining or overlapping vectors in small dimensions. Generally it uses two dimensions to represent line vectors (for example, as objects) and column vectors (variable figures). Biplot analysis is based on the Decomposition Singular Value (DNS) or Singular Value Decomposition (SVD) from the matrix of data which has been corrected to its average [8]. For example matrix X* is a matrix of double variable data consisting of n objects and p variables. Then the matrix X* is transformed to the average value which obtained by the matrix

\[
x = x^* - \frac{1}{n} (1 x^*)
\]

where 1 is a matrix where all the elements are worth 1 [9,10].

The Pearson correlation coefficient (r) is used to search the compatibility of the configuration of the origin certain variable data with its projection at the variable vector in the biplot.
\[ r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \]

; Where, \( x_i \): the origin certain variable data to-\( i \), \( y_i \): the projection of the object to-\( i \) at the variable vector in the biplot, \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \), and \( \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \).

For the compatibility of the ranking of the object, it is used the proportion from objects whose ranking is suitable with the ranking of the certain variable data [11,12]. While the compatibility of the ranking of the correlation coefficient variable to a certain variable, It is used the proportion from variables to biplot which the ranking of correlations toward certain variables is suitable with the ranking of the correlation of variable toward certain variables in the original data.

2. Methods

The research was conducted on third grade students from 22 High Schools in Malang City and Malang District. Samples were taken randomly from each high school. The determination of sample size is based on the Krejcie table with a real level of 5%, so that of the total students of 5396 students were taken as a sample of 359 students. While the size of sample of each high school is based on the proportion of the number of students of each high school. Analyses the data using the model of Path Analysis [1]. The results of the analysis obtained are directly or indirectly relationships of each variable. In addition, the criteria obtained generally from the quality of each variable (very low, low, medium, high, and very high).

The variables are used in this research:
- X1: Principal's Leadership,
- X2: Teacher's Ability to Teach,
- X3: Parents' Social Economy,
- X4: Input Quality,
- X5: Student Motivation,
- X6: Learning Facilities, and
- X7: Graduates Quality.

3. Results and discussion

Boxplot diagrams are one of the techniques to provide a description about the location of the centralized data, the range of distribution, and the slope of the distribution pattern. A description of the variables X1-X7 is presented in Figure 2.

![Boxplot Diagram](image-url)
Based on Figure 2, the slope of the pattern of data distribution variable X1 approaches symmetry or its median is almost as same as the average. The 12th object is an upper outline of variable X1, it means that the value of the object is quite large compared to the value of other objects. Variables X2, X4, and X5, the slope of the pattern of data distribution is negative. It shows that the average of the three variables is below the median. The 17th object becomes down outline of the variable X5, while the 2nd and 3rd objects become upper outline. On the variable X4, the 17th object is down outline. In Figure 2 it can also be seen that the variables X3, X6, and X7 have a slope of positive data distribution patterns, It indicates that the data of these variables are much below the average. Objects 2nd and 3rd become outline for the variable X7, it means that these two objects have a large difference in value when they are compared to the average or the value of its under object. Based on the original data, the value of the 2nd object and 3rd objects are 72.94 and 75.39, respectively, while the average of variable X7 is 65.01. From Figure 2, it is also obtained that X6 has a larger variance, while the variance of X5 is smaller than other variables.

The relationship between variables can be seen in Table 1. The significance of the correlations in the table based on the p-value are all less than 1% (Appendix 4). The biggest correlation occurred between variables X3 and X6 of 0.98. While variable X7 (graduate quality) has the biggest correlation with variable X4. It shows that the average of the graduate’s quality were achieved by schools has the highest correlation with the quality of input.

| Variable | X1   | X2   | X3   | X4   | X5   | X6   |
|----------|------|------|------|------|------|------|
| X2       | 0.62**|      |      |      |      |      |
| X3       | 0.74**| 0.91**|      |      |      |      |
| X4       | 0.70**| 0.79**| 0.84**|      |      |      |
| X5       | 0.55**| 0.81**| 0.77**| 0.69**|      |      |
| X6       | 0.80**| 0.90**| 0.98**| 0.83**| 0.75**|      |
| X7       | 0.74**| 0.83**| 0.93**| 0.94**| 0.72**| 0.93**|

*p-value <0.01*

Based on the configuration of original data variable X7 with its projection toward the vector of variable X7 in the biplot, it was obtained r of 0.96 for all tested values of α. Likewise, the rank has not suitable yet with the actual ranking.

Figure 3. The Biplot description of the object’s position and variables using α = 0.
Object configuration uses $\alpha = 0$. The analysis was obtained as follows: - using the value $\alpha = 0$. This selection is based on the previous description that for the value $\alpha = 0$, it was obtained the correlation between variables, in addition the quadratic of the Euclid distance between objects on the biplot is proportional to the quadratic of the Mahalanobis distance between objects from the original data, - doubling of standard deviation of variable $X_7$. The doubling coefficient that was tried: \{1.1, 1.2, \ldots\} until the expected results were obtained. The results obtained are summarized.

**Table 2.** The compatibility of Biplot, the object configuration and the rank of variables correlation based on some coefficients of the doubling of standard deviation $X_7$ and $\alpha = 0$.

| Compatibility       | The coefficient of doubling standard deviation |
|---------------------|-----------------------------------------------|
|                     | 1.1  | 1.2  | 1.3  | 1.4  | 1.5  | 1.6  | 1.7  |
| GGF Data (%)        | 90.36| 90.45| 90.56| 90.70| 90.85| 91.02| 91.21|
| Variable (%)        | 99.54| 99.56| 99.56| 99.57| 99.59| 99.61| 99.62|
| Object (%)          | 58.10| 57.85| 57.60| 57.35| 57.10| 56.86| 56.62|
| $r$                 | 0.96 | 0.97 | 0.97 | 0.97 | 0.98 | 0.98 | 0.98 |
| The rank of object (%) | 77.27 | 77.27 | 77.27 | 100.00 | 100.00 | 100.00 | 100.00 |
| The rank of variable correlation (%) | 66.67 | 66.67 | 66.67 | 66.67 | 66.67 | 33.33 | 33.33 |
Based on the results were given in Table 2, the compatibility of the configuration of object projection for the variable vector X7 based on the ranking, is obtained on the doubling of the standard deviation of X7 by 1.4, 1.5, 1.6, and 1.7. But the correlation level of variable with variable X7, the standard deviation 1.4 and 1.5 is better than 1.6 and 1.7. On the other hand, the value of r for the doubling of standard deviation 1.5 tends to be larger than the doubling 1.4.

**Table 3.** The compatibility of Biplot, object configuration and the rank of variables correlation based on several coefficients of doubling standard deviation X7 and $\alpha = 0$.

| Compatibility          | The coefficient of doubling standard deviation |
|------------------------|-----------------------------------------------|
|                        | 1.8 | 1.9 | 2   | 2.1 | 2.2 | 2.3 | 2.4 | 2.5 |
| GF Data (%)            | 91.41 | 91.63 | 91.86 | 92.10 | 92.34 | 92.58 | 92.83 | 93.07 |
| Variable (%)           | 99.64 | 99.66 | 99.68 | 99.70 | 99.72 | 99.74 | 99.76 | 99.77 |
| Object (%)             | 56.39 | 56.17 | 55.97 | 55.77 | 55.58 | 55.39 | 55.22 | 55.05 |
| r                      | 0.98 | 0.99  | 0.99  | 0.99  | 0.99  | 0.99  | 0.99  | 1.00  |
| The rank of object (%) | 90.91 | 90.91 | 90.91 | 90.91 | 90.91 | 90.91 | 81.82 | 81.82 |
| The rank of variable correlation (%) | 33.33 | 33.33 | 33.33 | 33.33 | 33.33 | 33.33 | 33.33 | 33.33 |

From Table 3, doubling of standard deviation 2.5 is obtained by r of 1, but the compatibility of object configuration is based on rank and correlation of variables to variable X7.

![Figure 6](image_url)  
**Figure 6.** Biplot description of the object’s position and variables using $\alpha = 0$ also standard deviation X7=2.5

From Figure 6 and Figure 7, we get a correlation description between variables in the doubling of standard deviation 1.5 is better than doubling 2.5. So that the doubling of standard deviations 1.5 describes the mapping of school quality that compatible with the National Exam scores better than the doubling 2.5. A description of the position of objects and variables on the biplot based on standard deviation X7 1.5 can be seen in Figure 7.
Based on Figure 5, when it was viewed based on the variable vector $X_7$, all variables are correlated positively with this variable. It shows that the quality of graduates which were achieved by a school correlates with the leadership of successful and effective principal, the ability of teachers’ teaching effectively in carrying out their duties, the level of parents ‘s education and parents ‘s economic, the high and low quality of input, student’s motivation, and completeness of learning facilities which owned by student. However, if it was observed further, variable $X_7$ has a larger correlation with variables $X_4$ and $X_6$. It means that the quality of input and learning facilities which owned by students are variables which having a larger correlation with the quality of graduates were produced. Meanwhile, variables $X_3$ and $X_6$ have a large positive correlation. They are shown by the large of angles are formed between the two variable vectors relatively small. It describes that the completeness of learning facilities which owned by students has large correlation with the level of education of. Students’ s parents and economics of students’ parent

General description of the quality of schools for the City and Malang Regency based on the quality of graduates, there are 8 schools which having graduate quality above the average, each region places 4 schools. However, the quality of schools in Malang City tends to be better than Malang Regency when it was viewed from the national examination average and the school rank. In biplot, the proximity of an object to a variable is indicated by the location of the object toward variable vector. If the object is one-sidedly with the direction of the vector, then the object has the value above the average. Conversely, if the location is opposite, it means the value is below average. While the proximity of an object with another shows the similarity between objects. Figure 4 provides a description of the object’s position and variable vectors in the biplot. Based on the proximity between objects and the proximity of objects to variables, these objects can be grouped into 7 groups, They are:

- **Group 1**, This school gets the highest quality input among other schools. It shows that SMAN 3 Kota Malang has an image as an excellent school for Malang City and Malang Regency. An Image as being an excellent school has emerged generally in the community, so, only students whose having high UN average are accepted at the school. Based on the original data, almost all other variables are also higher than other schools.

- **Group 2**, SMAN 2 Malang and SMAN 12 Malang Regency are described as having similar achievements, while SMAN 4 Malang left behind the two schools. Teacher performance and student motivation from these three schools still need attention in efforts to improve the quality of their schools.

![Figure 7. Biplot description of the object’s position and variables using $\alpha = 0$ also standard deviation $X_7 = 1.5$.](image-url)
• Group 3, It shows that for Malang Regency, SMAN 1 Batu and SMAN 2 Batu, the quality of schools is almost the same.

• Group 4, While the achievements of the quality of their graduates are still a little below average. These Differences in the variables are still below the average of these schools: SMAN 5 Kota Malang the levels of education students’ parents of and the levels of economic students’ parents, SMAN 6 Kota Malang, the leadership of principals, and SMAN Tumpang, the completeness of learning facilities which owned by students.

• Group 5, The position of this school is on the biplot has opposite quadrant to the variable vectors X1, X4, X6, and X7, so that on the one hand there is a fairly good student motivation but the average of input quality, graduate quality, student learning facilities, and the leadership of principal are still below average.

• Group 6, It means that these schools are characterized by having unsuccessful principals, the quality of input, the quality of graduates, and learning facilities which owned by students are also still less.

• Group 7, It shows that these schools in this group are characterized by: teacher performance ineffective in carrying out their duties, the level of education and economics of students ‘parents is below average, and students' learning motivation is still low.

4. Conclusions

Biplot can give description the relationship of schools with variables which influencing the results of the National Examination, so that the weakness and the advantages of schools are obtained based on the score of the National Examination. Based on the original area of SMAN, the quality of schools in Malang City tends to be higher than Malang Regency.

References

[1] Idris J 2005 Critical Analysis of Education Quality (Yogyakarta: Suluh Press)
[2] Pakpahan R 2001 Survey pelaksanaan Ebtanas SMP, SMA, dan SMK Tahun Pelajaran 1999/2000
[3] Farkas N and Nagi V 2008 Student assessment of desirable technical skills: a correspondence analysis approach Budapest: Acta Polytechnica Hungarica 52
[4] Gabriel K R 1971 The biplot- graphic display of matrices with application to principal component analysis Biometrika 58 453-467
[5] Gabriel K R 2002 Goodness of fit of biplots and correspondence analysis Biometrika 89(2) 423-436
[6] Jolliffe I T 2002 Principal Component Analysis 2nd ed. (Berlin: Springer-Verlag)
[7] Lipkovich I and Smith E P 2002 Biplot and singular value decomposition macros for Excel Journal of statistical software 7(5) 1-15
[8] Johnson RA and Wichern DW 2002 Applied Multivariate Statistical Analysis. 5th ed. (USA: Prentice-Hall, Inc.)
[9] Siswadi and Suharjo B 1999 Analisis Eksplorasi Data Peubah Ganda (Bogor: FMIPA IPB)
[10] Ardana N K K 2008 BiplotPack Ver. 3.2-A Mathematica Package for Multivariate Data Visualization (Bogor: IPB)
[11] Ardana N K K and Siswadi 2005 Biplot Dan Implementasinya Dengan Pemrograman Fungsional Mathematica JMA 4(2) 21-31
[12] Ivy J 2001 Higher Education Institution Image: A Correspondence Analysis Approach The International Journal of Educational Management 15 276-282