An Analysis of the Influencing Factors of College Students' English Achievement

Lu-tao ZHAO, Feng-yun LIANG, Zhi-gang ZHANG* and Er XU
School of Mathematics and Physics, University of Science and Technology Beijing, Beijing, China
*Corresponding author

Keywords: English achievement, Influencing factors, Principal component analysis, Multiple linear regression analysis.

Abstract. Based on the data collected from questionnaires and scores of corresponding students, this paper finds out some factors affecting students' English achievements through the subjective and objective analysis of the data. Firstly, the principal component analysis (PCA) is used to reduce the dimensions of 14 variables, and the four main factors affecting English achievement are effort, confidence, motivation and additional effort. After that, a multiple linear regression analysis is made on the four aspects. It is found that there is a non-linear correlation between students' English achievement and effort. Factors that have a greater positive impact on achievement are confidence and motivation, while the additional effort has a negative impact on students. Finally, some suggestions are provided for students' English learning accordingly.

Introduction

With the gradual development of China toward a powerful socialist country, there is an increasing need for international talents, which makes English a necessary skill for graduate students. Since English scores have accounted for a large proportion in the overall appraisal of a student, students pay more attention to it. In order to meet the employment needs of graduates, our school focuses on the training of students' English ability in the talent cultivation program. In order to help students to improve their English performance and master their English skills better, this paper designs 14 questions from four aspects: students' efforts, motivation, psychological quality and examination experience, and conducted a survey among students. A total of 528 valid questionnaires are collected in the end. This paper will make a research and analysis based on the data of these questionnaires and students' scores.

Literature Review

Because of the importance of education, some scholars have elaborated on the influencing factors of students' performance. For example, Sanzana used classification and regression tree (CART) and random forest analysis to illustrate that parents' expectation and school type affect performance [1]. Costa used path analysis to study the influence on students' learning, and then predicted the students’ achievement [2]. Migueisaa, Hamsa, Talebi and Altujjar also predicted the students’ achievement [3-6].

The principal component analysis and multiple linear regression methods have also been studied by some scholars. Çilan ued classified regression analysis to analyze the factors affecting MBA students’ academic performance [7]; Kuan used statistical analysis to establish the average regression equation of footprint and height [8]; Aramburo conducted descriptive analysis and multiple linear regression analysis of student data [9].
Data

Data Description

The data are collected from the classroom questionnaires and final English scores of the University of Science and Technology in Beijing. The average English score of the students is 72.22, and the excellent rate is 32.60%. 14 questions are included in the questionnaire, such as students' preference for English, their effort to review after class, their degree of rest before examinations, their time to recite words every day, and their intention to go abroad. A total of 528 valid questionnaires were collected in the end.

Data Preprocessing

Questionnaire data and scores are arranged in the data table. Each row represents a student's individual data. The first column is student information, the second to fifteenth columns represent options for different questions, and the last one is student's English scores. Because the options are all character data, it needs to be quantified.

Methodology

Principal Component Analysis

Principal Component Analysis (PCA) was first proposed by Hotelling in 1933. It synthesizes the indexes by dimensionality reduction and transforms the original random vectors related to their components into new random vectors irrelevant to their components by means of an orthogonal transformation. These new variables are arranged in descending order according to the variance. Keep the total variance of variables unchanged in the mathematical transformation, so that the first variable has the largest variance, which is called the first principal component. The second variable has the largest variance and is not related to the first variable, which is called the second principal component. Fourteen questions were involved in the questionnaire, which increased the calculation and complexity of subsequent analysis. Because of the obvious correlation between many variables, the information provided by the variables overlap to some extent. Therefore, it is necessary to reduce the dimension by principal component analysis and express most of the information with fewer variables.

Basic Principle

Based on the original data, new variables are obtained by principal component analysis. The new variables are the linear combination of the original variables, that is, multiple original variables are transformed into new variables by linear transformation.

Steps of Principal Component Analysis

Step 1: Standardization of indicator data

Since the principal component analysis method is influenced by the dimension and quantity of the evaluation indicator, the principal component will vary with the change of the dimension and quantity of the evaluating indicator, and some information will be lost. The equalization method is utilized to process the original data dimensionless, which can not only effectively eliminate the influence of dimension and order of magnitude, but also make the processed data contain all the information of the original data. This step can be implemented in SPSS.

Step 2: Correlation determination between indicators

Principal Component Analysis (PCA) requires a strong correlation between the indicators. KMO and Bartlett spherical tests are carried out by SPSS. When $KMO > 0.7$ and $Sig < 0.05$, principal component analysis is suitable and the effect is better.

Step 3: Determine the number of principal components $m$

If the total amount of data information contained in the first $m$ components (i.e. the cumulative contribution rate) is not less than 80%, the first $m$ components can be used to reflect the original
evaluation indicator, or the components with eigenvalue greater than 1 can be taken as the main component.

Step 4: The expression of principal component $F_i$

$F_i$ expression is written according to the factor score matrix obtained by SPSS.

Step 5: Name the principal component $F_i$

We subjectively named $F_i$ according to the coefficients of each variable in the principal component $F_i$’s expression.

**Multivariate Linear Regression Analysis**

For multivariate regression analysis, if the relationship between two or more independent variables and dependent variables is linear, it is called multivariate linear regression. The dependent variable is $y$, and the $k$ factors that influence the dependent variable are recorded as independent variable $x_1 \cdots x_k$. The corresponding model of multivariate linear regression analysis is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$  \hspace{1cm} (1)

$$\beta_0, \beta_1, \beta_2 \cdots \beta_k$$ is the undetermined coefficient.

**Results**

**Principal Component Analysis**

Using SPSS for principal component analysis, we know that $KMO = 0.72$ and $Sig = 0.00$, which indicates that principal component analysis is suitable. Then we can get four principal components and recorded them as $F_1, F_2, F_3, F_4$. The English Preference and the efforts to review after class etc. were recorded as variables $X_1, X_2 \cdots$. The expression of the principal components can be obtained as follows:

$$F_1 = - 0.015 X_1 - 0.002 X_2 - 0.016 X_3 + 0.025 X_4 + 0.010 X_5 - 0.047 X_6 - 0.076 X_7 + 0.270 X_8 + 0.305 X_9 + 0.294 X_{10} - 0.029 X_{11} + 0.208 X_{12} + 0.328 X_{13} - 0.081 X_{14}$$  \hspace{1cm} (2)

$$F_2 = 0.016 X_1 + 0.360 X_2 - 0.422 X_3 + 0.438 X_4 + 0.023 X_5 + 0.019 X_6 + 0.020 X_7 + 0.042 X_8 + 0.021 X_9 - 0.001 X_{10} - 0.046 X_{11} - 0.033 X_{12} - 0.039 X_{14}$$  \hspace{1cm} (3)

$$F_3 = 0.445 X_1 + 0.061 X_2 + 0.105 X_3 + 0.018 X_4 + 0.254 X_5 + 0.210 X_6 - 0.010 X_7 + 0.025 X_8 - 0.015 X_9 - 0.058 X_{10} + 0.360 X_{11}$$

$$F_4 = - 0.186 X_1 - 0.095 X_2 - 0.117 X_3 + 0.014 X_4 + 0.073 X_5 + 0.192 X_6 + 0.689 X_7 - 0.276 X_8 - 0.034 X_9 + 0.160 X_{10} + 0.103 X_{11} + 0.405 X_{12} - 0.108 X_{13} - 0.039 X_{14}$$  \hspace{1cm} (4)

It can be concluded from Eq.2 that the coefficients of the variables of exercise level before examination, time to recite words every day, degree of listening practice, degree of oral practice and reading effort are relatively large in Eq.2, so $F_1$ can be called effort;

$F_2$ has higher absolute coefficients of the following variables: Efforts to review after class, Number of examination participation, and self-evaluation on the spot. Therefore, $F_2$ can be called confidence. It can be found that when the number of examinations increases, confidence decreases.

In Eq.3, the absolute coefficients of English preference, extracurricular English accumulation and intention to go abroad are larger in $F_3$, so $F_3$ can be called motivation.

The absolute value of the coefficient of Extracurricular Counseling Level in $F_4$ is larger, so $F_4$ can be called additional effort.
Multivariate Linear Regression Analysis

4 principal components were obtained from 14 independent variables by principal component analysis, and these four principal components were used as new variables for linear regression analysis. $R^2 = 0.279$ and $\text{Sig} = 0.00$, which shows that the regression model performs well and the dependent variable has a significant linear relationship with the independent variable directly.

Table 1. Standardized Beta coefficients of the variables.

| Model   | Non-standardized coefficients | Standardized coefficients | t    | Sig  |
|---------|------------------------------|----------------------------|------|------|
| Constant| 79.193                       | 0.356                      | 222.364 | 0.000 |
| Efforts | 0.163                        | 0.356                      | 0.017 | 0.458 | 0.647 |
| Confidence | 4.401                       | 0.356                      | 0.458 | 12.346 | 0.000 |
| Motivation | 2.018                       | 0.356                      | 0.210 | 5.660 | 0.000 |
| Additional Efforts | -1.499                  | 0.356                      | -0.156 | -4.204 | 0.000 |

As can be seen from the table 1, $\text{Sig} > 0.05$ indicates that the relationship between effort and achievement is non-linear. This is because hard review at the beginning of class may lead to a significant increase in English achievement, but when a student's effort reaches a certain level, the space for improvement becomes smaller and the improvement of English achievement becomes more and more difficult.

The coefficient of confidence is 0.458, which is obviously higher than other variables. This indicates that besides the variable of effort degree which has no significant linear correlation with achievement, the most influential factor on students' English achievement is experiential confidence. If students review actively, they will naturally have more confidence in examinations. In addition, from the Eq.3, we can see that the number of examinations is negatively correlated with confidence. When the number of examination participation increases, confidence decreases, leading to a decline in performance. This is contrary to our long-held belief that the more number of exams we take, the more experience we will have, and therefore the more English scores we can improve. A large part of the reason is that after failing in the exams for several times, students' psychological state will be affected when they re-test.

There is a negative correlation between English achievement and extra effort, which shows that unlike the common view believed by many parents that after-school remediation will result in a higher achievement, too much after-school remediation, on the contrary, will lead to a decline in students' English achievement. There are mainly two reasons: One is that most of the students who often take extra-curricular classes usually have relatively poor learning ability or self-control, they cannot grasp the knowledge on the course well. Secondly, a misunderstanding of their English ability deficiencies results in a blind remedy. They do not conform to their own situation and the ability of internalization of knowledge is not good enough, which wastes a lot of time in remedy but fail to improve their performance.

Conclusions

From the above, it can be concluded that effort, motivation, confidence and additional effort are the main factors affecting students' English achievement. The relationship between effort and achievement is not linear, that is to say, single effort does not necessarily lead to ideal results. Therefore, teachers should not just ask students to increase their English learning time, but also cultivate students' interest in English learning to let them study English more actively and efficiently. Students should also carry out their own psychological construction to enhance their psychological quality, so as to better play their own strength in the examination. But students should not blindly carry out extra-curricular guidance, which may waste time and money, but also fail to improve
English performance. Students should find out their own problems in English learning, and then carry out targeted learning.

Acknowledgements
The authors gratefully acknowledge the financial support from the Beijing Municipal Education Commission under Grant No.CEHA17066.

References
[1] M.B. Sanzana, S.S. Garrido, Profiles of Chilean students according to academic performance in mathematics: An exploratory study using classification trees and random forests, J. Stud Educ Eval 44 (2015) 50-59.
[2] A. Costa, L. Faria, The impact of Emotional Intelligence on academic achievement: A longitudinal study in Portuguese secondary school, J. Learn Individ Differ 37 (2015) 38-47.
[3] V.L. Miguéisa, A.Freitasb, P.J.V. Garcia, Early segmentation of students according to their academic performance: A predictive modelling approach, J. Decis Support Syst 115 (2018) 36-51.
[4] H. Hamsa, S. Indiradevi, Student academic performance prediction model using decision tree and fuzzy genetic algorithm, J. Proc Tech 25 (2016) 326-332.
[5] S. Talebi, S. Davodi, A. Khoshroo, Investigating the effective component of classroom management in predicting academic achievement among English language students, J. Procd Soc Behv 205 (2015) 591-596.
[6] Y. Altujjar, W. Altamimi, I.A.Turaiki, Predicting Critical Courses Affecting Students Performance: A Case Study, J. Procedia Comput Sci 82 (2016) 65-71.
[7] Ç.A. Çilan, M. Can, Measuring Factors Effecting MBA Students’ Academic Performance by Using Categorical Regression Analysis: A Case Study of Institution of Business Economics, Istanbul University, J. Procedia Behv 122 (2014) 405-409.
[8] J. Kuan, Regression analysis estimation of stature from foot length, J. Cogn Syst Res 52 (2018) 251-260.
[9] V. Aramburo, B. Bororel, Predictive factors associated with academic performance in college students, J. Procd Soc Behv 237 (2017) 945-949.