# The Evaluation of English Learning Level about Non-English Majors Based on Canonical Correlation Analysis

## Su Ran Kong [1, a], Guang Li Li [2,b] and Hui Chuan Wang [3,c*]

[1,3]Basic Science Department, Beijing University of Agriculture, Beijing, China
[2]Beijing Normal University Library, Beijing Normal University, Beijing, China
[a]Email:kongsr@126.com
[b]Email:ligl@bnu.edu.cn
[c]Email:whch56221@sohu.com
[*]Author for correspondence

**Abstract.** The English learning level about non-English majors was divided into three levels which are primary level (College entrance examination scores), secondary level (each semester final exam scores) and final level (CET-4 scores). The aim of this study is to examine the relationships between three levels by using canonical correlation analysis. The results indicated that the canonical correlation coefficients between three levels are 0.685, 0.6523 and 0.7566, which illustrated that canonical correlation relation between three levels was remarkable. In addition, the results indicated that the indicators of the level have larger load in the indicator of other levels. Students were also offered some targeted advices in proving English learning level.

## Introduction

College English is a compulsory course for college students improving their abilities of listening, speaking, reading and writing. Improving comprehensive capability and learning level are common concerns both for teachers and students. Currently, the studies for college English mainly focused on three aspects.

Study on the theory and practice of College English Teaching. Most of those studies were countermeasures and suggestions proposed by educators who working at College English. Those suggestions are unconvincing because they were resulted from experience or impression lacking of theory evidence.

Study on learning effect and influencing factors of College English. Tao Tao and Li Guang made qualitative analysis about various factors influencing the effect of English learning [1, 2]. Wei Feng et al analyzed the relationship between learning effect and time, learning interest by conducting a multivariate regression analysis [3]. Statistical method was used by analyzing the learning effect and influencing factors rather than the learning level of College English.

Study on learning level of College English [4,5]. Previous studies mainly focused on qualitative evaluation or simply quantitative evaluation. Qualitative evaluation to learning level of College English was just subjective description lacking of objective evidence which is likely to be questioned by others. There were some statistical methods such as variance analysis and cluster analysis in simply quantitative evaluation, but only a few factors were considered.

On the basis of findings by predecessors, the influencing factors model of learning effect was established by using canonical correlation analysis. An empirical study was taken as Beijing

University of agriculture. The results will help to stimulate students' interest in learning and to improve teachers' teaching quality.

**Method**

Since Hotelling (1936) developed Canonical Correlation Analysis (CCA), it has been a useful tool in a wide variety of research areas including economics, medical sciences and education etc. [6]. CCA is an appropriate statistical technique for analysis of relationships between two sets of variables [7]: one set of variables is referred to as independent and the other as dependent [8]. First of all, a pair of linear combination that has maximum correlation coefficient is found. Secondly, a pair of linear combination that has maximum correlation coefficient is selected from the linear combination of irrelevant from initial combination. The rest can be deduced from this. The selected pair of linear combination is called the canonical variable. The correlation coefficient between them is called the canonical correlation coefficient [9].

The English level of non-English Majors was divided into N levels, named as $X_1, X_2, \cdots, X_n$. Let $X_i = (X_{i1}, X_{i2}, \cdots, X_{ip})^T$ and $X_j = (X_{j1}, X_{j2}, \cdots, X_{jq})^T$. When $p, q > 1$, it can be transformed correlation between new aggregate variables from correlation between $X_i$ and $X_j$ by principal component. That is to solve $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_p)^T$ and $\beta = (\beta_1, \beta_2, \cdots, \beta_q)^T$, which make maximum possible correlation between

$$V = \alpha_1 X_{i1} + \alpha_2 X_{i2} + \cdots \alpha_p X_{ip} = \alpha^T X_i; \tag{1}$$

$$W = \beta_1 X_{j1} + \beta_2 X_{j2} + \cdots \beta_q X_{jq} = \beta^T X_j. \tag{2}$$

If there are $\alpha_1 = (\alpha_{11}, \alpha_{21}, \cdots, \alpha_{p1})^T$ and $\beta_1 = (\beta_{11}, \beta_{21}, \cdots, \beta_{q1})^T$, they make

$$\rho(\alpha_1^T X_i, \beta_1^T X_j) = \frac{Cov(\alpha_1^T X_i, \beta_1^T X_j)}{\sqrt{Var(\alpha_1^T X_i)}\sqrt{Var(\beta_1^T X_j)}} = \max_{Var(\alpha_1^T X_i)=1, Var(\beta_1^T X_j)=1} \rho(\alpha^T X_i, \beta^T X_j). \tag{3}$$

Then variables $\alpha_1^T X_i, \beta_1^T X_j$ are called the first canonical variables and the vectors $\alpha_1$ and $\beta_1$ are the canonical coefficients.

If there are $\alpha_k = (\alpha_{1k}, \alpha_{2k}, \cdots, \alpha_{pk})^T$ and $\beta_k = (\beta_{1k}, \beta_{2k}, \cdots, \beta_{qk})^T$, they make

(1) $\alpha_k^T X_i$ and $\beta_k^T X_j$ are irrelevant to previous canonical correlation variables;

(2) $Var(\alpha_k^T X_i) = 1, Var(\beta_k^T X_j) = 1$;

(3) $\alpha_k^T X_i$ and $\beta_k^T X_j$ have the largest correlation coefficient.

Then variables $\alpha_k^T X_i, \beta_k^T X_j$ are called the $k$ pairs of canonical correlation variables and the vectors $\alpha_k$ and $\beta_k$ are the canonical coefficients, $k = 2, 3, \cdots p$.

**An empirical study**

**Data Collection.** Based on scientific, objective and easily manageable principles, the indicators of primary level, secondary level and final level were designed. The indicators of initial level were language scores in the college entrance examination (X11), math scores in the college entrance examination (X12) and English scores in the college entrance examination (X13); The indicators of secondary level were the first semester scores (X21), the second semester scores (X22), the third semester scores (X23) and the fourth semester scores (X24); The indicators of final level were learning scores in CET-4(X31), reading scores in CET-4(X32), writing scores in CET-4(X33) and comprehensive scores in CET-4(X34).

**Data Statistical Description.** The data were analyzed with descriptive statistics by SPSS 17.0. It showed that each indicator of learning level has followed the normal distribution.

Table 1   Descriptive statistics for the indicators of learning level

| Variables | Mean | Standard Deviation | Minimum | Maximum | Coefficient of Variation | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|
| $X_{11}$ | 106.745 | 7.591 | 77 | 131 | 7.112 | -0.082 | 0.111 |
| $X_{12}$ | 98.483 | 14.00 | 38 | 135 | 14.216 | -0.351 | 0.538 |
| $X_{13}$ | 82.906 | 17.676 | 35 | 139 | 21.322 | 0.682 | 0.483 |
| $X_{31}$ | 134.016 | 25.627 | 87 | 230 | 19.123 | 0.563 | 0.196 |
| $X_{32}$ | 151.048 | 21.892 | 98 | 224 | 14.493 | 0.161 | -0.111 |
| $X_{33}$ | 77.567 | 12.322 | 49 | 124 | 15.886 | 0.431 | 0.146 |
| $X_{34}$ | 42.057 | 11.809 | 26 | 69 | 20.971 | 0.499 | -0.302 |
| $X_{21}$ | 52.297 | 11.809 | 14 | 84 | 22.58 | -0.008 | -0.327 |
| $X_{22}$ | 56.193 | 10.463 | 23 | 86 | 18.620 | -0.041 | -0.235 |
| $X_{23}$ | 55.572 | 10.466 | 20 | 94 | 18.835 | 0.065 | 0.084 |
| $X_{24}$ | 58.308 | 12.938 | 22 | 95 | 22.188 | 0.139 | -0.241 |

**Findings.** Table 2 shows the relationship between indicators of primary level, secondary level and final level estimated by canonical correlation analysis.

Table 2   Canonical correlation analysis and canonical variables structure

| Variables 1 | Variables 2 | Canonical correlation coefficient | P value | canonical function |
|---|---|---|---|---|
| Primary level | Final level | 0.6523 | < 0.001 | $V1=0.1998 X_{11}+0.1377 X_{12}+0.9807X_{13}$<br>$W1=0.2320 X_{31}+0.2564 X_{32}+0.5383X_{33}+0.2176 X_{34}$ |
| Primary level | Secondary level | 0.6850 | < 0.001 | $V1=0.3006 X_{11}+0.1576 X_{12}+0.9536 X_{13}$<br>$W1=0.5254 X_{21}+0.1711 X_{22}+0.4344 X_{23}-0.0247 X_{24}$ |
| Secondary level | Final level | 0.7566 | < 0.001 | $V1=0.3784 X_{21}+0.2368 X_{22}+0.3753 X_{23}+0.1863 X_{24}$<br>$W1=0.4802 X_{31}+0.2252 X_{32}+0.4305 X_{33}+0.0847 X_{34}$ |

Table 2 shows that the canonical correlation coefficient between primary level and final level is 0.6523 and p value is less than 0.001. It shows that there exists a significant canonical correlation between the primary level and the final level of English learning. From the linear expressions of

canonical variables, we can see that English scores and writing scores have larger load than other indicators in the first pairs of canonical variables. We can also see that the indicators of primary level are positively correlation with the indicators of final level.

Table 2 shows that the canonical correlation coefficient between primary level and secondary level is 0.6850 and p value is less than 0.001. It shows that there exists a significant canonical correlation between the primary level and the secondary level of English learning. From the linear expressions of canonical variables, we can see that English scores in primary level, first semester scores and third semester scores have larger load than other indicators in the first pairs of canonical variables. We can also see that the indicators of primary level are positive correlation with first semester scores, second semester scores and third semester scores, but the indicators of primary level are negative correlation with fourth semester scores.

Table 2 shows that the canonical correlation coefficient between secondary level and final level is 0.7566 and p value is less than 0.001. It shows that there exists a significant canonical correlation between the secondary level and the final level of English learning. From the linear expressions of canonical variables, we can see that first semester scores ($X21$), third semester scores ($X23$), learning scores and writing scores have larger load than other indicators in the first pairs of canonical variables. We can also see that the indicators of secondary level are positively correlation with the indicators of final level.

The results showed that, first of all, English scores in the college entrance examination have more influence on CET-4 scores than language scores and math scores. This is not incompatible with reality. The students who had got higher English scores in the college entrance examination had got higher CET-4 scores. On the contrary, the students who had got lower English scores in the college entrance examination had got lower CET-4 scores; Secondly, English scores in the college entrance examination have more powerful connection with the first and third semester scores, but have more weaker connection with the second and fourth semester scores. The relevance decreased year by year which means the influence of primary level is gradually decreased. Therefore, students should lose the burden of "the basis is not good, must learn bad" and study hard for improving English comprehensive ability; Finally, each semester scores have more powerful connection with learning and writing scores in CET-4, but have more weaker connection with reading and comprehensive scores. It shows that students should strengthen reading and comprehensive training and teachers should pay attention to improving students of reading and comprehensive ability by cultivating students' interest. Only by doing this, student may use English freely and there is no problem in improving students' scores and performance.

**Conclusion**

The three levels indicators which have direct or indirect impact on English learning level are not separate index. They have relationship between them. All the indicators of primary level, secondary level and final level have impact on English learning level. Thus, we should evaluate English learning level based on many indicators instead of one indicator. An appropriate evaluating method can promote learning and teaching. An appropriate evaluating method can help students to improve English learning level and then to improve the ability of learning, speaking, reading and writing English.

As an English teacher, we should try different education methods according to the different types of students. For example, students who speak English difficultly should be encouraged to participate in more social English activity in order to have more chances to speak; Students who read English poorly should be encouraged to read more English materials in order to improve

literary accomplishment [10]; Students who have good foundation but short of efforts should be cultivated their interest, enthusiasm and good habits for learning English. Teach students in accordance with their aptitude. So we can totally improve English learning level of non-English Majors.

**References**

[1] Tao Tao, Analysis of influencing factors and Countermeasures of English Autonomous Learning, J. Theory and Practice of Contemporary Education. vol, 4(1)(2012) 128-130.

[2] Li Guang, Analysis of affecting factor about College Students' English learning, J. Mudanjiang Normal University. 152(2009) 74-75.

[3] Wei Feng, Jin Hua-wang. Analysis of affecting factor about College Students' English learning Results, J. Higher Agricultural Education. 11(2012) 66-69.

[4] Shang Xiao-hua, Wang Hai-hua. Study on correlation between English learning strategy use and CET scores. J. Foreign Language Education. vol, 31(2)(2010) 54-56.

[5] Yang Yue. Application of the Clustering Analytical Method in Students' English Language Performance Analysis. J. Education Science. vol, 24(6)(2008) 50-53.

[6] Hye-seune Lee. Canonical Correlation Analysis Using Small Number of Samples. Communications in Statistics—Simulation and Computation, 36(2007) 973–985, ISSN: 0361-0918 print/1532-4141 online.

[7] Gao Hui-xuan. Applied Multivariate Statistical Analysis, Peking University Press, Beijing, 2005.

[8] Alamgir Kabir, Rebecca D. Merrill et. Canonical Correlation Analysis of Infant's Size at Birth

and Maternal Factors: A Study in Rural Northwest Bangladesh. Information on http://www.plosone.org.

[9] Richard A.Johnson, Dean W.Wichern. Applied Multivariate Statistical Analysis, sixth ed, Tsinghua University Press, Beijing, 2008.

[10] Li Jie. Factors Affecing Passing Rate of College English Test Band Four and Strategies. J. Journal of Tianjin Vocational Technical Teachers' College. vol, 12(2)(2002) 53-55.