

2014

# BioTechnology

*An Indian Journal*

FULL PAPER

BTAIJ, 10(19), 2014 [11426-11430]

## Study on the assessment of college students English proficiency based on cluster analysis

Wu Li

School of Foreign Language Studies, Shandong Jiaotong University, Jinan, 250357, (CHINA)

### ABSTRACT

Cluster analysis method has a very profound impact on the assessment of college students in terms of their achievements in English, in which the ability of the students can be a more accurately assessed. With regard to this aspect, this paper, through sample selection of class 2007 Students' scores in Chinese universities, in which involves 206 colleges and universities. This research first sort the students' achievement squintly, and then further cut its initial classification as smaller samples according to a certain percentage. Using cluster analysis through appropriate statistical software to assess the data effectively, in order to draw out the perspectives that students need to improve, and it has specific classification of the students who share the similarities. This can make the process of group teaching in college English education explicitly and scientifically. Through research and discussion in this paper, allows the performance of the ability of students to be able to get specific embodiment, which leading a positive impact on enhancing the overall student achievement, while the process of cultivation of students will be continuously strengthened in order to achieve efficiency and comprehensive development. In the hope that this paper can provide the appropriate basis for the theory and practice of college English teaching and idealistic and scientific assessment methods.

### KEYWORDS

Cluster analysis; College students; English proficiency, Assessment methods.



## INTRODUCTION

In the research of this article, it has been explored the advantages of using the cluster analysis to examine the college students' English proficiency assessment. In this paper, the concept of cluster analysis and classification, cluster analysis method and quantitative assessment methods will be explained, and the application of cluster analysis to the case of university English language proficiency assessment methods will be discussed in three corresponding parts. The results of this study are scientifically and rationality guaranteed with a clear paper structure.

### THE CONCEPT AND CLASSIFICATION OF CLUSTER ANALYSIS

Cluster analysis is to classify variable and data statistics effectively, and cluster analysis mainly has two ways, one is a hierarchical clustering method and the other is an iterative clustering method. However, the group also known as the point group analysis or analysis Cluster analysis.

**Classification Clustering is shown as blew:**

#### (1)Direct hierarchical clustering

First of all, according to the fundamental principle of minimum distance, separate the different categories formation to independent classification, and establish a combination of classified objects into a new category. Among this classification if the two belong to the same corresponding classification, and then establish these two into new category. However, classification of objects belonging to two different classes, be grouped in the same category. However, after carrying out the merger to the classification category object columns and columns in order to delete counterparts, so that after merging of every category, merged cluster pedigree diagrams will be created.

#### (2)The shortest distance clustering method

It is to measure the distance between non-diagonal elements of the two original classification objects, classify two categories of objects, then calculate its effective distance by the relevant equation. This creates a new matrix, however, this research select the smallest from the matrix among of the effective selection<sup>[2]</sup>, then formats a new matrix after the merger among the minimum, thus establish a new category, then after the process of calculating the distance among categories and new categories, until guiding classified objects can be divided into the same category.

#### (3)The longest distance clustering method

The thought of the longest distance clustering method is more or less the same as the one of the shortest distance clustering method, but the biggest difference is the presence of the formula for the calculation process for selection, the longest distance clustering method of calculation is to measure the distance that exists between the sample size with a maximum distance.

### CLUSTER ANALYSIS AND QUANTITATIVE ASSESSMENT METHODS

Students always mentioned about the assessment as exams, which is indispensable to teaching and learning activities. However, in terms of the assessment, it has two aspects, the first is from the prospective of teachers that to inspire the students to get full potential of their ability, reading and teaching can provide effective feedback; secondly, from the perspective of the students, learning outcomes can be obtained affirmation of their own and others, and to transit effective studying methods and habits effectively<sup>[3]</sup>. However, in the current school assessment there are mainly two, one is the true score theory test method, the other is the principle of the theoretical test. However true score theory test is more comprehensive compare to the theoretical principles exam, in which the ability of the students will get embodied, however, there is higher requirements for the proposed university teaching hardware in the majority of the universities, the implementation of the examination method is not realistic. From the outcome of teaching point of view, group teaching has a positive effect for the continuous improvement of the teaching quality, however, to conduct group teaching of college English, to group the students according to their ability reasonably, and also conducting assessment on the ability of students effectively.

The above discussion shown that it cannot be ignored to adopt reasonable assessment of college English teaching, no matter it is in normal teaching activities or grouping teaching. However, grouping according to the ability of the students, the appraisal process has not been paid sufficient attention, conversely, there are too many attentions and efforts were made to the exams, thus there's no quantified assessment in terms of the ability to process more scientifically. This article introduces and explains cluster analyses, which will make this quantitative assessment method functioning in the research of College English.

From the foregoing discussion process, you can learn the process of clustering analysis is multivariate statistical analysis, which defines the main effective thing is the accordance with property classification process, and then there will be the relationship between the properties of the analysis, such is the basic process of cluster analysis<sup>[3]</sup>. However, cluster analysis is based on the effective classification of relationships between the variables, usually regarded a sample or a variable

as point in the multidimensional space, and then to calculate the distance between two points. Yet another method is to describe the degree of similarity between the elements of the samples selected according to its closeness degrees. In this paper, according to the former corresponding cluster analysis and data for the specific assessment of college students' English assessment, in order to get the cluster diagram pedigree, this can be established by the spectral characteristics of the student's ability and knowledge of the structure with the same effective classification.

### CASE STUDY ON THE APPLICATION OF CLUSTER ANALYSIS METHOD TO COLLEGE ENGLISH ASSESSMENT

In this conducting research in this article, it selected 206 universities or colleges from all the national universities, with the purpose of examining the students' English test scores of class 2007, including 81 first batch undergraduate universities, of which the total number of students was 20749, 23 second batch undergraduate universities, the number of students was 8299, and 102 colleges, total number of students was 20,513.

#### Methodology

According to the scores of the students as sequence, a sequence of the students from the first batch undergraduate university, a sequence of the students from the second batch undergraduate universities and a sequence of the students from the third batch undergraduate colleges, formed as 3 levels<sup>[4]</sup>. The proportions of the sample accounted of the total number of respective batch are: 25%, 50%, and 25%. As result, we can draw 9 different types of universities. The symbol L indicates the first batch undergraduate universities, and L1-L3, respectively; represent the 3 different levels (Upper, middle and lower) of the first batch universities. L4-L6 represent the 3 different levels of the second batch undergraduate universities, L7-L9 represent 3 different levels of colleges, by this process, this research made the cluster analysis on college English assessment.

Calculated the average score of the students from the 206 universities, formulate i a gradual process of clustering analysis on the assessment of the college English achievement.

Using professional statistical software to analyze the data.

#### Results of the research

##### The result of the original 9 types of universities

From the data in TABLE 1, there are 6 levels in the 9 types of universities, the best scores are mainly witnessed the best universities of the first batch undergraduate universities. Then the best of the second batch universities and the middle ones of the second batch universities followed as the second level. And the middle ones of the second batch universities and the colleges are belongs to the third level. The worst ones of the first batch universities are the forth level<sup>[5]</sup>. The middle ones of the colleges and the worst of the second batch universities are the level five, and the rest are the level six.

TABLE 1 : The result of the original 9 types of universities

Original 9 types	L1	L2	L3	L4	L5	L6	L7	L8	L9
Number of the universities	15	38	28	6	12	5	30	49	23
Average score	118.67	103.49	90.78	108.36	99.06	82.53	95.86	84.84	68.81
Rate of the scores	0.79	0.69	0.61	0.72	0.66	0.55	0.64	0.57	0.46

##### The result of the clustering analysis on the section scores of English

The college English test mainly reflects the eight sub-items, including a big question, total 8, the specific content of the test was illustrated by Figure 1 and TABLE 2.

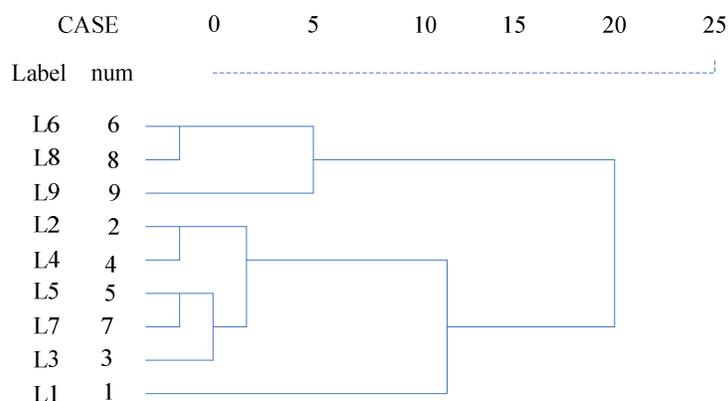


Figure 1 : The result of the clustering analysis on the section scores of English

**TABLE 2 : The eight test item**

Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
0.75	0.69	0.58	0.80	0.63	0.73	0.61	0.45

There are 9 levels of the 206 universities, and the result of the sub-items was as followed TABLE 3.

**TABLE 3 : The 9 levels of sub-items**

Type of the university category	No. of the sub-items								
	1	2	3	4	5	6	7	8	Total
1st level	0.94	0.91	0.78	0.98	0.91	0.88	0.88	0.63	0.84
2nd level	0.84	0.78	0.66	0.90	0.74	0.85	0.75	0.62	0.74
3rd level	0.86	0.81	0.68	0.92	0.77	0.84	0.75	0.50	0.73
4th level	0.81	0.75	0.61	0.89	0.70	0.82	0.70	0.41	0.67
5th level	0.78	0.72	0.60	0.83	0.67	0.73	0.65	0.53	0.66
6th level	0.76	0.68	0.57	0.82	0.62	0.76	0.60	0.38	0.62
7th level	0.71	0.64	0.53	0.75	0.58	0.71	0.57	0.47	0.59
8th level	0.65	0.58	0.49	0.68	0.50	0.62	0.47	0.37	0.52
9th level	0.57	0.50	0.43	0.57	0.39	0.50	0.36	0.28	0.44

**TABLE 4 : School cluster survey data**

University types	The original 9 types								
	L1%	L2%	L3%	4%	5%	L6%	L7%	L8%	L9%
One	53.3	0	0	0	0	0	0	0	0
Two	0	5.3	0	50	8.33	0	6.7	0	0
Three	33.3	23.7	0	16.7	0	0	0	0	0
Four	0	47.7	7.15	0	0	0	3.3	0	0
Five	6.7	2.6	0	33.3	83.34	0	43.3	0	0
Six	6.7	21.1	46.4	0	0	0	6.7	14.3	0
Seven	0	0	21.4	0	8.33	40	36.7	49.0	0
Eight	0	0	17.9	0	0	40	3.3	30.6	34.8
Nine	0	0	7.15	0	0	20	0	6.1	65.2
Total	100	100	100	100	100	100	100	100	100

As we can be seen from the above data, in the first stage of a total number of eight schools, although this stage are placed at the front of the undergraduate group, in all the universities of the second batch of universities, and includes a middle ones of the first batch universities, Then the best of the second batch universities and the middle ones of the second batch universities followed as the second level. And the middle ones of the second batch universities and the colleges are belongs to the third level, the number of this level is 15, individually is 5 and 9. However, there are totally 21 of the fourth level of universities. This level stage is a mostly middle reaches of the schools, the number is 18. Among the rest four distributions of levels of universities, the seventh is more complicated, it has a different analogy among each level. There are 27 universities in level fifth, mainly are the middle one of the second batch universities and best ones of the colleges, number is 10 and 13. And there are 31 universities in the sixth level, which involves a number of various categories of undergraduate, and also the first 2 categories of the colleges, the number was 1, 8, 13, 2 and 7. With regard to the seventh and eighth level, mainly are colleges, a total number of institutions in the seventh level is 44, while the former two categories, respectively, colleges occupy 11 and 24. Among eighth level universities involved a total of 31, which are mainly from the middle and lower class of the colleges, number is 15 and 8 respectively.

As can be seen from TABLE 4, the process of a university ranked in the forefront of the school during the school clustering mainly distributed over the third level, and this proportion has been reached 86.6%, however, the middle ones during clustering process mainly distributed between the second level to the sixth grade in school and nearly half of these universities are in the fourth level. The universities came in a thick cluster of schools in the process mainly distributed between the sixth level to ninth level, and in this part, there 85.7 percent were divided to the sixth to eighth level of

universities. However, there are 50 percent forefront of the universities can be included in the second level during the clustering process, and the one third of these universities are placed as seventh level. The the middle ones, are mainly distributed in the fifth level, while the rest of the schools were all in the second to the seventh level. Among which there are schools, mainly are in between the sixth to ninth grade, and proportions of the fifth level and seventh level has reached 79.69 percent, but the rest universities are mainly distributed in the last two levels, in which the last level accounted t two-thirds of the total.

### Discussion

Based on the cluster analysis on the nine original classification, the results show an upper, middle and upper schools in two of the first two levels of the hierarchy, the ability of the students' of learning English is higher. Only students from the upper levels reached average eligibility requirements in the colleges, which reflected that the general level of English learning is relatively low.

TABLE 4 shows the various ability levels of students, in which, the third, eighth items are relatively low, while it has many reasons for the poor performance of the item three, one important reason is that the understanding of English essay instructions, especially the lack of deeper understanding on the structure. The eighth item is aiming at testing English writing skills, in which students should integrate English knowledge and express themselves in a proper way, however, it is the disadvantages of college English learning, so it is not surprising that they got lower scores in this item<sup>[6]</sup>.

TABLE 4 shows that, although a school students' English learning levels are relatively high, the phenomenon of the visibility of distribution various in multiple levels, so this kind of imbalance balance needs to be improved, and some relatively good universities have poor performance of English learning achievements. There are 13 universities did not meet the eligibility requirements of English learning, thus enhancing the English teaching in these universities is a priority; the performance of the upper level colleges English teaching also showed greater variability, which distributed in different hierarchical levels.

### CONCLUSION

Above is the exploration on cluster analysis on assessment of college students' English assessment. In this research, we tested the English proficiency of the selected sample, and then according to the students ability conducted effective assessment and classification process by cluster analysis. Classification of the students is more clearly, and also makes the target of improve the student's ability more strengthened

### REFERENCE

- [1] Jing Liangyou; Impact of heterogeneous cooperative learning groups on the students with learning difficulties in English studing, *Teaching and Performance Management: Theory Version*, (4), 72-74 (2013).
- [2] Xie Qun, Zhang Yubin, Zhang Jibiao, etc; Cluster analysis and comprehensive evaluation on summer coastal marine environment quality of Leizhou Peninsul, *Marine Environmental Science*, 33(4), 543-549 (2014).
- [3] Zhao Dongling, Feng Yanruo, Pan Zhengyun; Research on clustering analysis based Algorithm K- grid, *Journal of Technology*, 30(7), 175-178 (2014).
- [4] Gong Liyan, Meng Xianjun; Quality evaluation of apple processing based on Principal Component Analysis and Cluster Analysis, *Agricultural Engineering*, 30(13), 276-285 (2014).
- [5] Zouyao, Qi Guinian, Liu Tingting, etc; Ingredients and cluster analysis on the principal componentr of Sichuan ottawa heap of black tea quality, *Food Industry Technology*, 35(15), 304-307 (2014).
- [6] Wang Yang, Zhang Mingsheng, Li Xiangdong, etc; Cluster analysis on different forms of sorghum germplasm diversity and morphological markers, *Guizhou Agricultural Sciences*, 42(6), 1-5 (2014).
- [7] Wei Sen; Cluster analysis on problems of the small and medium leisure farms in Nanjing, *Jiangsu Agricultural Sciences*, 42(6), 430-434 (2014).