International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-10

https://doi.org/10.70917/ijcisim-2026-1812
Article

Identifying Patterns of Learners' Cognitive Styles on Differences in English L2 Lexical Expression through Cluster Analysis

Junling Wang[1,2] and Suthagar Narasuman[1,*]

[1] Faculty of Education, Universiti Teknologi MARA, Kampus Puncak Alam, Selonger, 42300, Malaysia
[2] School of Foreign Languages, Ningxia Medical University, Yinchuan, Ningxia, 750001, China
* Correspondence author: sarahwang1234@163.com

Abstract: This study aims to investigate the pattern of association between English learners' cognitive styles and their second language vocabulary expression differences. The study used the Embedded Figures Test (EFT) to measure the cognitive styles of the subjects, categorized the subjects into two cognitive styles: field-independent and field-dependent, and used the acquisition test to explore the timely and delayed effects of the subjects' field cognitive styles on the acquisition of second language vocabulary in English. The LSD method was used to verify whether the learners' different cognitive styles made a difference in English bilingual vocabulary expression, and then the k-means clustering analysis was used to derive the pattern of difference between the two. The results show that learners can be classified into three typical patterns: "field-independent - high retention", "field-dependent - high immediacy", and "field-dependent - low efficiency". Different cognitive styles of learners result in different types of second language vocabulary expression differences. The findings of this paper provide empirical evidence and data support for the subsequent implementation of individualized vocabulary instruction and assessment.

Keywords: Embedded Figures Test (EFT); acquisition test; LSD method; k-means clustering; bilingual vocabulary expression differences

1. Introduction

In English language acquisition, learners' individual differences are important factors affecting the learning effect, and cognitive style, as a unique and stable psychological tendency shown by learners in the cognitive process, influences individuals' reception, processing, storage and application in the process of English second language vocabulary acquisition [1-4]. Traditional research on English second language vocabulary acquisition has mainly focused on the influence of teaching method improvement and other aspects, and has lacked attention to the variability of learners' intrinsic cognitive modes. As a matter of fact, learners with different cognitive styles show different behavioral patterns in terms of the variety, complexity and accuracy of lexical expressions when processing English lexical information [5-6]. Meanwhile, most of the existing studies focus on the application of cognitive load in mathematics, science and other disciplines, which makes it difficult to clearly describe the relationship between vocabulary acquisition and cognitive styles, and the related studies in the field of second language vocabulary acquisition are still fragmented.

Cluster analysis, as an exploratory data analysis method, has the core goal of dividing cases in a dataset into different categories or clusters according to some similarity metric [7-9]. Given that cluster analysis can effectively classify homogeneous groups based on similarity metrics, it is particularly suitable for identifying the patterns of cognitive styles and L2 lexical expression differences in this study. In the pattern of learners' cognitive styles on the differences in English L2 vocabulary expressions, cluster analysis can be divided into multiple clusters based on the similarity of learners' multiple cognitive styles, which is helpful to grasp the pattern of the differences in English L2 vocabulary expressions under the clusters of different cognitive styles by comparing the metrics of

richness and complexity of the vocabulary expressions of these clusters [10-13].Thus, this study aims to fill this gap by exploring the association between cognitive styles and L2 lexical expression differences through cluster analysis.

The patterns of learners' cognitive styles on the differences in English bilingual vocabulary expression were classified by cluster analysis. The study firstly categorized the subjects into field-independent and field-dependent through the Embedded Figures Test (EFT), and then explored the timely and delayed effects of the subjects' different field cognitive styles on vocabulary acquisition through the Vocabulary Acquisition Test. Then, based on LSD multiple post hoc tests, we explored whether different cognitive styles make a difference in English bilingual vocabulary expression. Finally, k-means clustering was used to classify the patterns of differences in learners' cognitive styles on English bilingual vocabulary expression.

2. Research design and methodology

2.1. Study design

(1) Subjects of the study

The participants of this study were first-year postgraduate students from six classes at a medical university. The test instrument used was Nation's Vocabulary Level Measurement Scale, adapted for medical English terminology. To ensure homogeneity in participants' medical English proficiency, a total of 240 students from the six classes completed a medical English vocabulary proficiency test. Based on k-means cluster analysis of their medical English vocabulary test scores and current semester's medical English final examination scores (with k=2, categorized into low proficiency and intermediate-advanced proficiency), 120 students with intermediate to advanced medical English proficiency were finally selected as the formal subjects of this study. The cluster analysis for subject screening adopted the same data standardization and Euclidean distance calculation methods as the subsequent cluster analysis of 110 valid samples, ensuring the consistency of research methods.

(2) Research Tools

a) Reading material and target words

The reading material for this experiment is a Greek myth entitled "*The Golden Fleece*", which is interesting and requires no background knowledge. The proportion of unknown vocabulary in the article was controlled within 5.3%. To ensure the reading material met this requirement, two measures were adopted: According to the *New College English Syllabus*, high school graduates typically master about 1,800 core English words. After four years of undergraduate study, medical students further expand their vocabulary to approximately 3,500–4,500 words. Following nearly one year of postgraduate education, participants' general and academic vocabulary size is well above 2,000 words. First, the reading material was analyzed using the lexical frequency analysis (LFP software), which categorizes all words in the material into five frequency levels. Second, a pilot study was conducted with 10 postgraduate students of similar English proficiency. They were asked to read the material and mark all unknown words. Statistical results showed that the students identified 29 unknown words, accounting for 5.3% of the total text, confirming the moderate difficulty of the material. Among these 29 words, 15 key words essential for text comprehension were selected as target words. To avoid participants' prior exposure to these target words in extracurricular reading or learning, the 15 target words were replaced with non-words (e.g., replacing "fleece" with a non-existent pseudo-word).

b) Field Cognitive Style Test

The field cognitive style test adopted in this study was the Embedded Figures Test (EFT) developed by the Department of Psychology, Beijing Normal University. The test consists of 26 complex figures and 9 simple reference figures. Participants were required to locate and mark the corresponding simple figures within each complex figure under time constraints, using a pencil or pen. The scoring criteria followed conventional standards: 0.5 points for correct answers to items 1–2; 1 point for correct answers to items 3–4; 1.5 points for correct answers to items 5–10; 0 points for incorrect answers. The total score was 24 points. Participants scoring $\geq$ 12 points were classified as field-independent (FI); those scoring < 12 points were classified as field-dependent (FD).

c) Timely and delayed vocabulary acquisition test

Vocabulary acquisition refers to the mastery of a word's morphology, meaning, and usage. During vocabulary instruction, teachers typically use teaching materials to explain target words, including word formation, multiple meanings, collocations, common expressions, idioms, and applicable contexts. In this study, vocabulary acquisition tests were used to examine differences in medical English bilingual vocabulary expression among participants. Immediately after reading the experimental material, participants completed a timely test assessing their acquisition of the 15 target words. A

delayed test with the same items (but rearranged word order) was administered to eliminate potential memory effects from the timely test. Tests were scored independently by two experienced EFL teachers, using Wesche and Paribakht's Vocabulary Knowledge Scale (VKS). The scale demonstrated high reliability, with a Cronbach's α of 0.95. Vocabulary test results were scored quantitatively and analyzed using SPSS: each target word correctly recognized and used in context was awarded 1 point (total score = 15 points), which was consistent with the VKS scale's scoring standards for vocabulary mastery. The 15 target words were selected based on their frequency, academic relevance, and importance for text comprehension. All target words were identified as high-level academic vocabulary items in the reading material.

(3) Experimental Steps

First, during the one-semester English teaching process, participants completed the Embedded Figures Test (EFT) and were divided into two groups: field-independent (FI) and field-dependent (FD) (based on the EFT scoring criteria, participants with a score ≥ 12 were classified as field-independent, and those with a score < 12 were classified as field-dependent). Both groups received the same reading task and subsequent vocabulary test simultaneously. Participants were not informed about the vocabulary test in advance. During reading, dictionary use and peer consultation were prohibited. After 20 minutes of reading, materials were collected, and a timely vocabulary acquisition test (assessing knowledge of the 15 target words) was administered to both groups. Participants were given 15 minutes to complete and submit the test. A delayed vocabulary test was administered two weeks later. This two-week interval was based on psychological principles of forgetting, aiming to reduce carryover effects from the initial test and measure long-term retention of target vocabulary. Due to attrition (10 participants failed to complete all tests, including 6 who did not attend the delayed test and 4 who submitted invalid test papers), 10 out of 120 participants did not complete all tests; thus, the final valid sample size was 110. The attrition rate was 8.33%, which is within the acceptable range in academic research and did not affect the proportional distribution of FI/FD groups (FI 68, FD 42). Meanwhile, the researcher collected the medical English final examination scores of all 240 students at the end of the semester, and used the scores of 120 formal subjects for subsequent data analysis.

2.2. Research methodology

In this study, the LSD (Least Significant Difference) method was used to test the significance of differences in medical English bilingual vocabulary acquisition between field-independent (FI) and field-dependent (FD) groups. Subsequently, K-means clustering analysis was employed to classify the participant groups based on their performance characteristics.

(1) Least Significant Difference (LSD) method

Under the premise that the F-test is significant (the F-test result of this study was $F=8.72$, $p<0.01$, indicating significant overall differences between groups), the absolute value of the difference between the means of any two groups ($\bar{X}_i$ and $\bar{X}_j$) is divided by the standard error of the difference between the means. The two-tailed probability distribution value of this t is then found and compared with the significance level. If the probability is less than the significance level, the difference between $\bar{X}_i$ and $\bar{X}_j$ is considered significant; otherwise, it is not significant. Among them:

$$t = \frac{\left|\bar{X}_i - \bar{X}_j\right|}{\sqrt{\frac{2MSE}{n}}} \tag{1}$$

MSe : Mean square error from the F-test.

n : Number of valid samples per group (in this study, n =68 for the FI group and n =42 for the FD group, based on a total valid sample of 110 participants).

When utilizing the LSD method for multiple comparisons, the following steps can be followed:

a) Create a table of multiple comparisons of means and arrange them according to their averages from largest to smallest.

b) Calculate the absolute value of the difference between the two treatment averages ($\left|\bar{X}_i - \bar{X}_j\right|$) divided by the value of the standard error of the difference between the means.

(2) Compare the calculated two-tailed probabilities with the significance levels of 0.05 and 0.01 to make statistical inferences.

(3) Cluster analysis model

The main purpose of cluster analysis in this study is to classify the 110 final valid samples of

participants, based on their medical English vocabulary proficiency and vocabulary acquisition performance, to identify the patterns of cognitive styles and L2 lexical expression differences. Its characteristics are not known in advance, so the analysis is based on the statistical data of the participants. The similarity or dissimilarity between samples is analyzed through the "distance" between objects. To distinguish between objects, samples belonging to the same class have a high degree of commonality, while the distance between different classes is relatively large. Usually, assuming that the samples of things with certain similar characteristics are $x_1, x_2, ..., x_n$, and the structure of each sample x_i is portrayed by p indexes, which is written as $x_i = (x_{i1}, x_{i2}, ..., x_{ip})$, with $i = 1, 2, ..., n$, and $j = 1, 2, ..., p$, and the basic steps of cluster analysis are as follows:

Step 1: Data standardization processing.

Due to the different statistical data scales of each indicator, it is necessary to standardize the data when conducting the research. The basic method is to first perform centering transformation processing on each variable. The formula is:

$$x_{ij}^{'} = x_{ij} - \overline{x}_j ; (i = 1, 2, ..., n; j = 1, 2, ..., p) \tag{2}$$

where x_{ij} denotes the statistics of the j th indicator for the i th sample, and $\overline{x}_j$ denotes the sample mean of the j th indicator for all samples. Then, the result of centering the variable is used to standardize the data, and the standardization formula is:

$$x_{ij}^{''} = \frac{x_{ij}^{'}}{s_j} ; (i = 1, 2, ..., n; j = 1, 2, ..., p) \tag{3}$$

where s_j is the sample standard deviation of the j th indicator ($s_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2}$). After the above standardized transformation, each variable, i.e., each column of data in the data matrix, has a mean of 0 and a variance of 1, and the statistics of each statistical indicator no longer have a scale.

Step 2: Calculate the initial distance matrix.

According to the statistical data after standardization, each sample is regarded as a class for the initial classification, which is written as $G_1, G_2, ..., G_n$ (where n =110 represents the final total number of valid participants). The distance d_{ij} between the samples G_i and G_j is calculated using the Euclidean distance formula:

$$d_{ij} = \sqrt{\sum_{k=1}^{p}(x_{ik}^{''} - x_{jk}^{''})^2} \tag{4}$$

Get the distance matrix $D^{(0)} = (d_{ij})_{n \times n}$, which at this point is equivalent to dividing the sample into n initial classes.

Step 3: Perform clustering.

According to the idea of cluster analysis, firstly, according to the initial Euclidean distance matrix D_0, find the two classes with the smallest distance and merge them into a new class for the first clustering. Let's assume that the distance between G_p and G_q is the shortest, and form them into a new class as G_{n+1}, with the other original classes unchanged. For ease of presentation, the notation is changed to: $G_1, G_2, ..., G_{n-1}$. Thus, the original n classes are clustered into $n-1$ classes.

In the later clustering process, the class may contain two and more samples. To overcome the defects of the gravity method, such as the calculation of the distance between the classes losing the information of the sample details, this paper adopts the class averaging method. The class averaging method defines the distance between two classes as the average of the Euclidean distances between all pairs of elements from the two classes, and the calculation formula is:

$$d_{G_{new}G_k} = \frac{1}{|G_{new}| \cdot |G_k|} \sum_{x \in G_{new}} \sum_{y \in G_k} d_{xy} \tag{5}$$

where $|G_{new}|$ and $|G_k|$ represent the number of samples in classes G_{new} and G_k respectively, and d_{xy} represents the Euclidean distance between sample x and sample y . Thus, the new distance

matrix D' after the first clustering is obtained according to equation (5). Then, the two classes corresponding to the minimum distance are found to be merged for the second clustering; and so on, the calculation and classification are repeated until all the elements are clustered into one class. Generally, the distance between classes can be written as:

$$d_{G_iG_j} = f\left(d_{xy}\right) \tag{6}$$

where G_i denotes the i-th class after the t-th clustering and $\left|G_i\right|$ denotes the number of elements. According to this equation (6), the distance matrix between the classes after the t-th clustering is obtained:

$$D^t = \left(d_{G_iG_j}\right)_{m\times m} \tag{7}$$

where, $m = n - t + 1$.

Step 4: Determine the number of classifications.

To get the desired k-means clustering results, it is necessary to find the optimal k value. The most common used methods to find the optimal k-value are the intra-cluster variance (SSE) method and the contour analysis method. Since the optimal k-value determined by the contour coefficient method is not necessarily optimal, sometimes it is necessary to assist in the selection based on SSE method. This study adopts the intra-cluster variance (SSE) method in combination with k-means clustering to determine the optimal number of clusters (k=3), which ensures the homogeneity of the selected sample. To verify the rationality of the optimal k-value, this study set k=2-9 for clustering (the common range of k-value in cluster analysis for small sample sizes), and compared the SSE values corresponding to each k-value to determine the optimal inflection point.

3. Results and analysis

3.1. Data collection and analysis

(1) The results of the study on the type of field cognitive style of the subjects

After the experimental subjects were given the Embedded Figures Test (EFT), the distribution results of field cognitive styles of 110 subjects were obtained by scoring the test. In order to visualize the results of the EFT and the distribution of the types of field cognitive styles of the second language learners of English in this experiment, this study entered the test data into SPSS26.0 and analyzed it with descriptive statistics, and the descriptive statistics of the results of the EFT are shown in Figure 1. The results are the final test scores (results are retained to two decimals) after the paper statistics scores were brought into the normative formula. The minimum value of the subjects' score on the EFT was 6.53, the maximum value was 20.32, and the mean score was 13.2(±3.67).

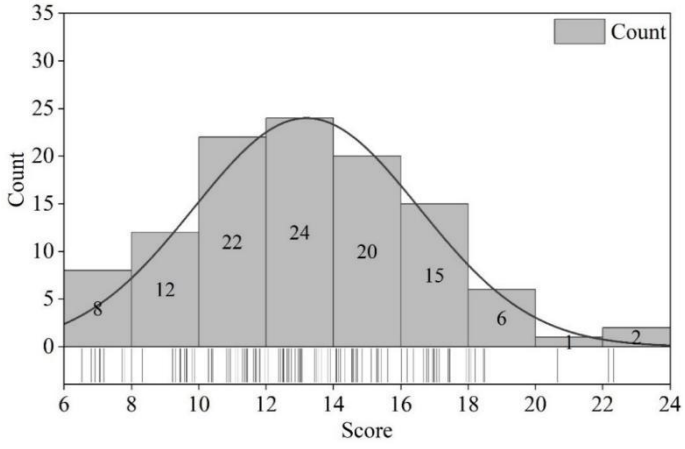

Figure 1. Descriptive statistical results of Embedded Figures Test scores

The descriptive statistics of the overall type situation of field-specific cognitive styles according to the cognitive style categorization criteria of the Embedded Figures Test (EFT) are shown in Table 1. Among the 110 experimental subjects, there were 68 subjects with field-independent cognitive style,

accounting for 61.82%, while there were only 42 subjects with field-dependent cognitive style, accounting for 38.18%. It can be seen that the number of field-independent learners is on the high side and the number of field-dependent learners is on the low side among the second language learners of English, i.e., the overall preference is for the field-independent cognitive style. Among the 68 field-independent learners, there are 35 and 33 male and female students respectively, accounting for 31.82% and 30% respectively. Among the 42 field-dependent learners, there were 19 and 23 male and female students, accounting for 17.27% and 20.91% respectively. The above results show that there are more boys than girls among field independent learners and more girls than boys among field dependent learners.

Table 1. Descriptive statistical results of the overall type of cognitive style

Categories	Gender	N	Proportion
Field independent	Male	35	31.82%
	Female	33	30%
	Total	68	61.82%
Field dependent	Male	19	17.27%
	Female	23	20.91%
	Total	42	38.18%

In order to determine whether there is a correlation between gender and field cognitive style, this study conducted a chi-square test for field cognitive style and gender, and the chi-square test for field cognitive style and gender is shown in Table 2. It can be seen that the two-tailed concomitant probability between field cognitive style and gender is 0.452 (>0.05), so there is no significant correlation between field cognitive style and gender, although there are more male students in the field-independent style and more female students in the field-dependent style, but according to the analysis results, it can be seen that the gender of the learners of English II does not affect the type of their field cognitive style.

Table 2. Field cognitive style and Gender Chi-Square Test

	Value	Fredoom	Incremental Sig.(2-side)	Accuracy Sig.(2-side)	Accuracy Sig.(2-side)
Pearson card	0.553	1	0.452	-	-
Continuity correction	0.152	1	0.741	-	-
Likelihood ratio	0.563	1	0.433	-	-
The test is accurate	-	-	-	0.706	0.352
Linear correlation	0.532	1	0.452	-	-
Valid case number			110		

(2) Subjects' Acquisition Test Findings

In order to understand the overall timely effect of field cognitive style on vocabulary acquisition, the timely test scores of 110 subjects in this study were recorded into SPSS26.0 and analyzed with descriptive statistics, and the descriptive statistical results of the timely test scores of learners with different field cognitive styles are shown in Figure 2. The mean and standard deviation of the vocabulary acquisition timely test scores of the field-independent learners were 10.94 and 3.63, respectively, and the mean and standard deviation of the vocabulary acquisition timely test scores of the field-dependent learners were 9.72 and 3.15, respectively, and the mean value of the timely test scores of the field-independent learners was higher than that of the field-dependent learners by 1.22, and the overall scores of the field-independent learners had a somewhat greater degree of dispersion.

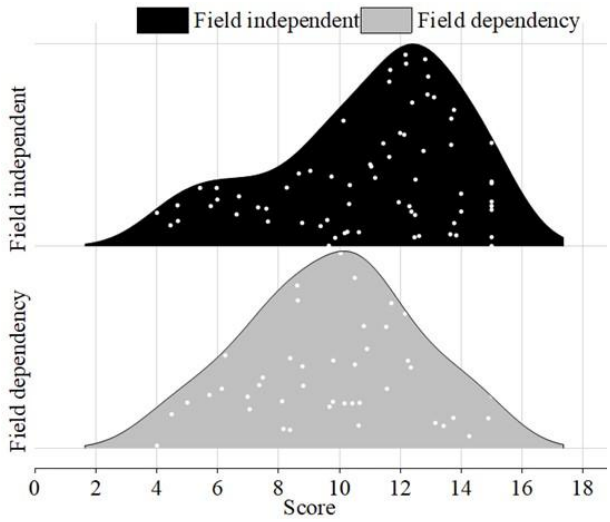

Figure 2. Timely test performance descriptive statistics

In order to understand the overall delayed effect of field cognitive style on vocabulary acquisition, this study conducted a delayed test on 110 English II learners two weeks after the end of the timely test, and the test scores were entered into SPSS26.0 after the end of the test and analyzed with descriptive statistics, and the results of descriptive statistics of delayed test scores of the learners with different field cognitive styles are shown in Figure 3. The mean and standard deviation of the vocabulary acquisition delayed posttest scores of the field-independent learners were 9.05 and 4.43, respectively, and those of the field-dependent learners were 8.41 and 3.55, respectively. The mean of the delayed posttest scores of the field-independent learners was higher than that of the field-dependent learners by 0.64, and the gap between the two means narrowed significantly after two weeks, and both field-independent and field-dependent learners showed a decrease in delayed test scores compared to timely test scores.

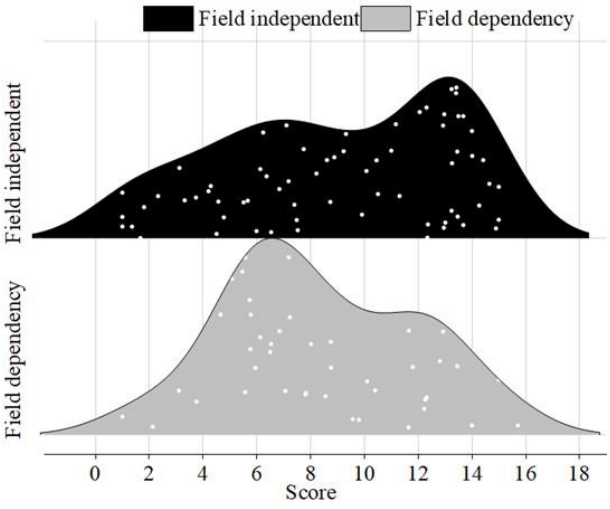

Figure 3. Delayed test performance descriptive statistics

3.2. Correlation between cognitive style and vocabulary acquisition

In order to further explore whether the differences between learners with different field cognitive styles and the timely test and delayed tests are significant or not, the LSD test was conducted in this study, and the LSD multiple post hoc test is shown in Table 3. There is a significant difference between the performance of students with different cognitive styles in the timely and the delayed tests, and the significant level is less than 0.01. That is to say, the difference between field-independent and field-dependent styles has a significant effect on the vocabulary acquisition in English second language.

Table 3. LSD multiple Post Hoc Test

Testing	Test group	Test group	Mean difference	Standard error	Significant level
Timely test	Field independent	Field dependent	1.22	0.355	0.005
Delayed testing	Field independent	Field dependent	0.81	0.237	0.003

3.3. Clustering results and differentiation analysis

The classification of learners' cognitive styles on the pattern of differences in English bilingual vocabulary expressions exists from 2 to 9 categories, so the selection of the best points needs to be realized by SSE. The cleaned data in csv format is first iteratively clustered by *k*-means clustering algorithm. Before clustering to set the value of *k* (number of clusters), in accordance with the common range of k-value in cluster analysis for small sample sizes, *k* was set to 8 numbers from 2 to 9 for clustering, and the results after clustering were saved. Here we choose to run k-means clustering in Python. After that the results of the completed clustering are used to calculate the SSE values corresponding to each k-value using the intra-cluster variance algorithm. Finally the obtained SSE results are plotted as SSE trend graph, due to the large value of SSE, it is now normalized and the final SSE trend image is plotted as shown in Fig. 4. From the figure, we can see that there is an obvious inflection point when $k = 3$. Therefore, $k = 3$ is chosen as the classification number, i.e., there are three types of patterns of learners' cognitive styles on the differences in English bilingual vocabulary expressions. The clustering process strictly followed the four steps described in Section 2.2, including data standardization, initial distance matrix calculation, clustering based on the class averaging method, and optimal k-value determination based on SSE.

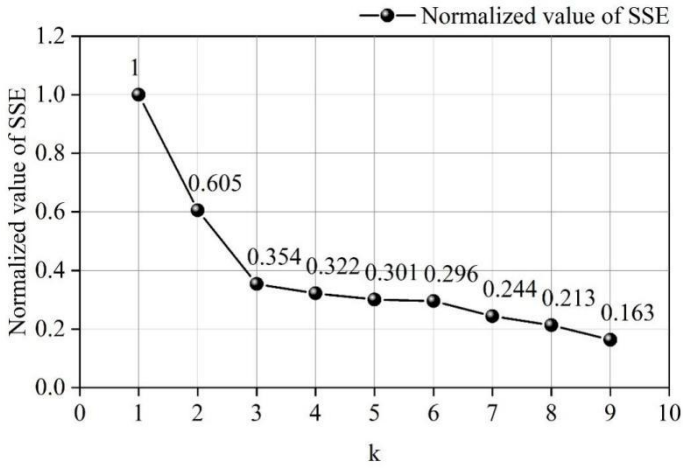

Figure 4. SSE trend diagram

The distribution of the means of the three types of learners on the cognitive style type test and the acquisition test is shown in Figure 5. The mean values of the "field- independent-high retention (Cluster 1)" group on the two tests are 16.52 and 10.63, respectively, and they perform well on the delayed test, which reflects the characteristics of field independent people who are good at out-of-context and in-depth processing, and their memory retention is good. The "Field Dependent-Highly Timely (Cluster 2)" group had mean values of 10.52 and 9.52 on the two tests, respectively, with good timely acquisition, but with a significant drop in performance on the delayed test, reflecting the dependence on specific contexts and the fugitive nature of memory of the field dependents. The "Field Dependent-Low Efficacy (Cluster 3)" group had mean values of 6.65 and 4.98 on the two tests, respectively, and did not perform well at either time point, revealing the ineffective state that can arise when cognitive styles are not matched with learning strategies.

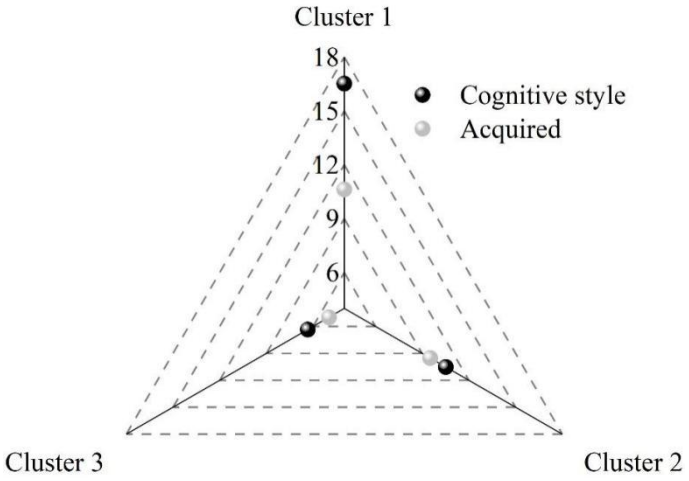

Figure 5. The average distribution of three types of learners on two tests

4. Conclusion

In this paper, data statistics on the differences between learners' cognitive styles and second language vocabulary expression in English were collected, and cluster analysis was conducted based on the statistical results to classify different types of difference patterns. The results of the LSD multiple post-hoc test showed that learners' cognitive styles would produce significant differences in second language vocabulary expression in English, and the levels of significance were all less than 0.01. A total of three categories of learners' cognitive styles were identified through k-means clustering analysis as patterns of differences in second language lexical expressions in English. The mean scores of "field-independent-high retention" learners in the field cognition style test and the acquisition test were 16.52 and 10.63 respectively. The mean scores of "field-dependent-high immediacy" learners in both tests were lower than those of "field-independent-high retention" learners. The mean scores of "field-dependent-low efficiency" learners were the lowest in both tests. In this regard, teachers can adopt differentiated teaching methods for learners of different patterns. For instance, for "field-independent - high retention" learners, timely and clear corrective feedback can be provided. For "field-dependent - high immediacy" learners, encouragement should be given for them to express themselves boldly, with a focus on cultivating fluency and language sense. For "field-dependent-low efficiency" learners, more contextualized vocabulary practice can be offered. Providing targeted teaching for different learners is conducive to improving the efficiency of learners' English learning.

About the Author

Junling Wang, female (born December 1978), Han nationality, native of Shang County, Shaanxi Province. Master's degree, Associate Professor. Research interests: English Language and Literature, English Education.

Suthagar Narasuman, Male (3rd. October, 1966-), Malaysia, Dr., Associate Professor, Teaching English as a second language (TESL).

References

1. Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. Trends in cognitive sciences, 22(2), 154-169.
2. Segalowitz, N. (2014). Individual differences in second language acquisition. In Tutorials in bilingualism (pp. 85-112). Psychology Press.
3. Norbury, C. F. (2019). Individual differences in language acquisition. International handbook of language acquisition, 323-340.
4. Hoang, V. Q. (2021). The differences of individual learners in second language acquisition. International Journal of TESOL & Education, 1(1), 38-46.
5. Xiaoxiao, L., & Sijia, X. (2020). The Effect of Cognitive Style on Non-English Majors' Incidental Vocabulary Acquisition through Reading. INNOVATION AND MANAGEMENT, 804.

6. Ujang, S. (2022). Correlational study of vocabulary mastery and cognitive learning style in vocational school. Journal of English Language Teaching Innovations and Materials (Jeltim), 4(1), 102-126.
7. Aggarwal, C. C. (2018). An introduction to cluster analysis. In Data clustering (pp. 1-28). chapman and hall/crc.
8. Jaeger, A., & Banks, D. (2023). Cluster analysis: A modern statistical review. Wiley Interdisciplinary Reviews: Computational Statistics, 15(3), e1597.
9. Dalmaijer, E. S., Nord, C. L., & Astle, D. E. (2022). Statistical power for cluster analysis. BMC bioinformatics, 23(1), 205.
10. Papi, M., & Teimouri, Y. (2014). Language learner motivational types: A cluster analysis study. Language Learning, 64(3), 493-525.
11. Taheri, H., Sadighi, F., Bagheri, M. S., & Bavali, M. (2019). EFL learners' L2 achievement and its relationship with cognitive intelligence, emotional intelligence, learning styles, and language learning strategies. Cogent Education, 6(1), 1655882.
12. Farooq, M. S. (2015). Cognitive Styles and Quality of Learning: A Case of Language Learners. Journal of Elementary Education, 25(1), 19-37.
13. Wright, J., Ahn, Y. H., & Lee, K. (2018). Investigating Language Learning Strategies in English Conversation Using Non-Hierarchical Cluster Analysis. English Teaching, 73(1), 183-200.