

A Quantitative Study of Multiple Regression Modeling in the Improvement of College English Vocabulary Learning Efficiency

Jipei Zhang^{1, ✉}

¹ School of Foreign Languages, Henan University of Urban Construction, Pingdingshan, Henan, 467036, China

ABSTRACT

The expansion of English vocabulary is the foundation of college students' English learning and the key to improve English learning. This project centers on the quantitative analysis of college English vocabulary learning efficiency improvement, through the questionnaire survey to understand the use of English vocabulary learning strategies of students. The influencing factors of English vocabulary learning efficiency improvement are selected, correlation analysis is carried out, and then multiple regression model is used to explore the role of each variable on the improvement of English vocabulary learning efficiency. The results show that students most often use the metacognitive strategy of pre-planning (3.674), and that students who are good at learning are more inclined to adopt the metacognitive strategy to control vocabulary learning from a macro perspective. Multiple self-elements and environmental elements together positively affect the improvement of English vocabulary learning efficiency ($p < 0.01$), with the most significant effects of learning strategies (0.482), teaching methods (0.457) and learning strategies (0.416). It is recommended to promote the efficiency of English vocabulary learning through innovative teaching methods, combining word class memorization, expanding the scope of reading, and vocabulary association learning.

Keywords: Multiple regression model, correlation analysis, quantitative analysis, questionnaire survey, English vocabulary learning

1. Introduction

In the process of learning English at university, students encounter many problems in English learning. The problems encountered in listening comprehension, reading comprehension and writing practice are often caused by the lack of vocabulary mastery. It can be seen that to learn college English well, the first problem to be solved is the problem of students' vocabulary [1-4]. It directly affects students'

✉ Corresponding author.

E-mail address: hncjzjp@126.com (J. Zhang).

Received 05 March 2024; Revised 25 May 2024; Accepted 05 December 2024; Published Online 15 April 2025.

DOI: [10.61091/jcmcc127a-268](https://doi.org/10.61091/jcmcc127a-268)

© 2025 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

interest in English learning, their mastery and utilization of English-related knowledge and skills, and ultimately affects the effect of teaching and learning. The improvement of English vocabulary learning efficiency is affected by many factors [5-8].

First of all, it is the emotional factor. Emotional state directly affects the formation and retrieval of memory. Positive emotions are often accompanied by better learning results. When students learn in a relaxed and happy atmosphere, they are more likely to absorb new knowledge and memorize new vocabulary. Therefore, creating a favorable learning environment is crucial [9-12]. Secondly, there are social and cultural factors. Social culture is a major factor affecting vocabulary learning. When learning English vocabulary, college students often use social media, online courses and other platforms to observe the habits and expressions of native English speakers [13-16]. This kind of online learning not only broadens their vocabulary, but also improves their language application skills [17]. In addition, motivation plays an indispensable role in vocabulary learning. Intrinsic motivation and extrinsic motivation together influence learners' behaviors and attitudes [18-19]. For college students, the purpose of learning English vocabulary varies, some want to pass the IELTS exam, while others want to be able to communicate fluently. No matter where the motivation comes from, a strong sense of purpose enables students to maintain endurance and resilience in the face of difficult learning tasks [20-23]. In addition, the factors affecting the improvement of English vocabulary learning efficiency include study time, study method, study frequency, etc. Multiple regression model is one of the commonly used tools in educational research, which can help teachers and students to analyze and explain the causal relationship between different variables, which is of great significance to improve English learning efficiency [24-27].

Literature [28] designed an intelligent recommendation system for English vocabulary learning using an intelligent recommendation model, which is based on machine learning algorithms to predict students' abilities and learning styles, which is conducive to mobilizing students' interest in learning and improving teaching effectiveness. Literature [29] used quasi-experimental and quantitative methods to explore the effects of multiple associative vocabulary teaching strategies and traditional strategies on students' vocabulary development. It was emphasized that associative vocabulary teaching strategies were more effective than traditional vocabulary teaching strategies. Literature [30] conducted a meta-analysis of the results of vocabulary learning using mobile devices and vocabulary learning using traditional learning methods. It revealed that learning vocabulary using mobile devices leads to greater achievement compared to traditional methods. Literature [31] examined the application of adaptive learning model of English vocabulary in the field of English teaching and learning. It revealed its effectiveness in improving learners' vocabulary acquisition, memorization and overall language proficiency. Literature [32] explored the effectiveness of adaptive algorithms in English vocabulary learning and emphasized that the algorithms are effective in enhancing students' English vocabulary learning. Literature [33] explored the connotation of AI technology and its principles of natural language understanding, and analyzed the advantages of its application in the field of education, proposing practical measures for adopting AI tools in the input-output stage of English language learning. Literature [34] and literature [35] explored the influence of songs on English vocabulary learning. It reveals that song-based learning not only enhances students' English vocabulary learning efficiency, but also is well received by students. Literature [36] considered the strategies that students prioritize when learning English vocabulary. Guessing the meaning of words based on context, looking up in the dictionary and memorizing vocabulary words were emphasized as the most common vocabulary learning strategies used by students. Strategies such as note-taking and asking for help were rarely used.

The above studies explored the improvement of English vocabulary learning efficiency from methods such as song learning method and mobile device learning method, which not only showed the importance of English vocabulary to the improvement of English learning and the quality of English teaching, but also reflected the importance that academics have attached to the improvement of

English vocabulary. However, at present, the research on the application of multiple regression modeling in English vocabulary learning is still relatively rare.

In this paper, we design a scale of English vocabulary learning strategies and select the influencing variables of learning efficiency improvement from the factors of own environment and external environment, and use questionnaire survey method to investigate the first grade students in a school. The collected data were statistically analyzed to reveal the use of English vocabulary learning strategies, and the differences in the use of learning strategies among students with different English proficiency levels and students of different genders were discussed. Subsequently, correlation analysis of each English vocabulary learning efficiency improvement influencing variable was conducted and regression hypotheses were proposed. Then the multiple regression analysis model is applied to analyze the relationship between each variable and English vocabulary learning efficiency improvement through model testing and regression analysis. Finally, the results of the analysis are synthesized to explore the strategies for the improvement of college English vocabulary learning efficiency.

2. Research survey design

In order to better enhance the efficiency of college English vocabulary learning, the study first conducts a questionnaire survey on the research subjects in order to understand the current status of the use of English vocabulary learning strategies of these research subjects, and then utilizes the multiple regression equations to excavate the influencing factors of the enhancement of English vocabulary learning efficiency. The survey design of this paper is as follows.

2.1. Subjects of study

The questionnaire of this study was sent to some first-year students of a university in M. The questionnaire was distributed in the middle of June 2024, and 352 questionnaires were distributed and 329 were retrieved, with a recovery rate of 93.47%. Removing 13 invalid questionnaires, a total of 316 valid questionnaires were obtained, and the effective recovery rate of the questionnaires was 96.05%.

2.2. Questionnaires

2.2.1. Learning Strategies Questionnaire. The English Vocabulary Learning Strategies Questionnaire is divided into two parts: the first part is the respondents' personal data, and the second part is the vocabulary learning strategies, which consists of 15 questionnaires covering three metacognitive strategies (including pre-planning, self-assessment, and self-supervision) and seven cognitive strategies (including repetition, categorization, contextualization, looking up in the dictionary, doing exercises, making connections, and note-taking). On a five-point scale, students choose the option that most closely matches their learning strategy from "1=this sentence is not suitable for me", "2=this sentence does not apply to me at all", "3=sometimes it applies to me, sometimes it does not apply to me", "4=this sentence applies to me basically", and "5=this sentence applies very well to me".

2.2.2. Influencing factors questionnaire

1) Selection of Influencing Factors. In this paper, students' own environmental factors (English knowledge X1, learning attitude X2, learning strategy X3, learning motivation X4) and external environmental factors (learning tools X5, teaching philosophy X6, teaching style X7, teaching attitude X8) are selected as the influencing factors for the improvement of college English vocabulary learning efficiency.

2) Questionnaire Design. According to the above selection of influencing factors, the questionnaire is divided into two parts and eight dimensions, namely, own environment factors and external

environment factors, and the questionnaire questions are designed based on these eight dimensions, with a total of 30 question items. Likert five-point measure was used.

2.3. Research methodology

For the collected data, this study used statistical methods such as mean, standard deviation, correlation analysis and multiple regression model to process them. Taking English knowledge, learning attitude, learning strategy, learning motivation, learning tools, teaching philosophy, teaching style, and teaching attitude as independent variables, and English vocabulary learning efficiency improvement as dependent variables, the study utilizes multiple linear regression modeling to explore how these factors affect the improvement of English vocabulary learning efficiency in universities, and to find out the key factors affecting the improvement of English vocabulary learning efficiency.

2.3.1. Correlation analysis methods. Correlation analysis is a test of the correlation between the independent variables and the dependent variable and is a precursor to linear regression analysis. If correlation analysis reveals that some independent variables are not correlated with the dependent variable, then regression analysis is not necessary. If correlation exists, the exact relationship between them is further verified through regression analysis. In addition, correlation analysis can also determine the covariance of the respective variables, if some of the independent variables have high correlation, then there may be covariance.

In statistics, correlation analysis is usually described by Pearson's correlation coefficient correlation coefficient. It is expressed by r and the formula is shown in (1):

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

where n is the sample size, x_i, y_i are the i point observations corresponding to variables x, y , $\bar{x}$ is the x sample mean, and $\bar{y}$ is the y sample mean.

From the formula, we can see that the range of values of r is $[-1, 1]$ and if $|r| < 1$, then the closer $|r|$ to 1, the higher the correlation between x and y . If $r = 0$, then x and y are not linearly correlated. If $r = -1$, then x and y have a perfectly negative linear correlation. If $r = 1$, then x and y have a perfect positive linear correlation.

2.3.2. Multiple regression models. The general process of multiple linear regression analysis consists of three main steps: first, establishing dependent quantitative relationships, i.e., estimating regression statistics and regression equations. Second, estimating the off-regression error and testing statistical hypotheses on the regression equation and regression statistics. Third, further utilization of the regression equation.

1) Establish a quantitative relationship between Y dependent X . Assuming that the dependent variable Y has a linear relationship with multiple independent variables $X_1, X_2, X_3, \dots, X_m$ and an error term e_i , a classical model of m -element linear regression is obtained:

$$Y_i = a + b_1 X_{i1} + b_2 X_{i2} + \dots + b_j X_{ij} + b_k X_{im} + e_i = \hat{Y}_i + e_i \quad (2)$$

where: $i = 1, 2, \dots, n$ denotes the i nd set of observations, $j = 1, 2, \dots, m$ denotes the j th independent variable, b_j denotes the partial regression coefficient, and e_i denotes the random error. It is represented by the matrix as:

$$Y = Y' + E = XB + E \quad (3)$$

Where Y is a vector of dependent variable observations, Y' is a vector of regression values, B is a vector of partial regression coefficients of the independent variables, X array is the structure matrix or so called design matrix, and E is the error vector, which are denoted as follows:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ L \\ Y_i \\ L \\ Y_n \end{pmatrix}; X = \begin{pmatrix} 1 & X_{11} & X_{12} & L & X_1 & L & X_{1m} \\ 1 & X_{21} & X_{22} & L & X_{21} & L & X_{2m} \\ L & L & L & L & L & L & L \\ 1 & X_{i1} & X_{i2} & L & X_{ij} & L & X_{im} \\ L & L & L & L & L & L & L \\ 1 & X_{n1} & X_{n2} & L & X_{nj} & L & X_{nm} \end{pmatrix} \quad (4)$$

$$B = \begin{pmatrix} a \\ b_1 \\ L \\ b_j \\ L \\ b_m \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \\ L \\ B_j \\ L \\ B_{m+1} \end{pmatrix}; E = \begin{pmatrix} e_1 \\ e_2 \\ L \\ e_i \\ L \\ e_n \end{pmatrix}$$

2) Estimation of off-regression errors and statistical hypothesis testing of regression equations and regression statistics. The most commonly used parameter estimation method in classical models is ordinary least squares estimation (LSM), while maximum contingency estimation or moment estimation is mostly used in non-classical models. In this paper, ordinary least squares estimation is used.

For n set of sample observations (Y_i, X_{ij}) , $i = 1, 2, \dots, n$, and $j = 0, 1, \dots, m$ of the dependent and independent variables, after obtaining the parameter estimates of the sample function, then the regression statistic of Y can be obtained Y' . According to the definition of regression, the sum of squares from the regression must be minimized, i.e., the regression statistic of the dependent variable to must be satisfied:

$$Q = RSS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n \left(Y_i - a - \sum_{j=1}^m b_j X_{ij} \right)^2 = \min \quad (5)$$

The matrix is of the form $Q = RSS = E'E = Y'Y - B'X'Y = Y'Y - B'K$.

Take the first order partial derivatives for a and b respectively and make the partial derivatives zero to obtain the system of regular equations:

$$\begin{cases} \sum (a + b_1 X_{1i} + b_2 X_{2i} + L + b_{mi} X_{mi}) = \sum Y_i \\ \sum (a + b_1 X_{1i} + b_2 X_{2i} + L + b_{mi} X_{mi}) X_{1i} = \sum Y_i X_{1i} \\ \sum (a + b_1 X_{1i} + b_2 X_{2i} + L + b_{mi} X_{mi}) X_{2i} = \sum Y_i X_{2i} \\ M \\ \sum (a + b_1 X_{1i} + b_2 X_{2i} + L + b_{mi} X_{mi}) X_{mi} = \sum Y_i X_{mi} \end{cases} \quad (6)$$

expressed in terms of a matrix:

$$\begin{pmatrix} n & \sum X_{li} & L & \sum X_{mi} \\ \sum X_{li} & \sum X_{li}^2 & L & \sum X_{li}X_{mi} \\ L & L & L & L \\ \sum X_{mi} & \sum X_{mi}X_{li} & L & \sum X_{li}^2 \end{pmatrix} \begin{pmatrix} a \\ b_1 \\ b_2 \\ L \\ b_m \end{pmatrix} \\
= \begin{pmatrix} 1 & 1 & 1 & 1 \\ X_{11} & X_{12} & L & X_{1n} \\ L & L & L & L \\ X_{m1} & X_{m2} & L & X_{mn} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ L \\ Y_n \end{pmatrix} \quad (7)$$

or expressed as: $\frac{\partial Q}{\partial B} = \frac{\partial Y'Y - 2B'X'Y + B'X'XB}{\partial B} = -2X'Y + 2X'XB = 0$.

Solving this system of regular equations yields the following result:

$$B = (X'X)^{-1} X'Y = A^{-1}K = CK \quad (8)$$

Where, A is the information array, K is the constant array and C is the inverse of the information array.

After obtaining the regression statistics, a significance test for linear regression can be made by decomposing the sum of squares and degrees of freedom i.e. F test or by calculating the sampling error of the regression parameters i.e. t test.

The procedure for the F test is as follows:

$$F_j = \frac{U_{pj}}{MS_Q} = \frac{b_j^2 / c_{jj}}{Q / (n - m - 1)} \quad (9)$$

$$U = SS_y - Q, U_{pj} = \frac{b_j^2}{c_{jj}}, MS_U = U / m \quad (10)$$

$$MS_Q = \frac{Q}{n - m - 1}, j = 1, 2, Lm$$

where MS_U is the linear regression mean square, MS_Q is the off-regression mean square, U_{pj} is the partial regression sum of squares, Q is the off-regression sum of squares, n is the number of groups of observations, m is the total number of independent variables, and SS_y is the sum of squares of the off-means y of the Y .

The procedure for the t test is as follows:

$$t_j = \frac{b_j - \beta_j}{s_{b_j}} = \frac{b_j}{s_{b_j}} \quad (11)$$

$$s_{b_j} = s_{Y/X} \sqrt{c_{jj}}, s_{Y/X} = \sqrt{\frac{Q}{(n - m - 1)}} \quad (12)$$

where $s_{Y/X}$ is the standard deviation from the regression and s_{bj} is the standard error of the partial regression coefficient for the j rd variable. Theoretically the F test and the t test are equivalent and they are squared, i.e. $F_j = t_j^2$.

3) Further utilization of regression equation. The least squares regression and the significance test can be used to obtain the effect of the respective variable on the dependent variable, as well as the

magnitude and manner of the effect, i.e., whether it is positive or negative, and the main influencing factors can be identified.

After modeling, the regression equation needs to be tested for significance to verify the correctness of the regression model.

A common method to test the significance of the regression equation is the F-test.

F-test: to test the overall significance of the regression equation, ANOVA is used to test the significance of the linear relationship between all independent and dependent variables.

Construct the following hypothesis testing problem:

$$\begin{aligned} H_0 : \beta = \beta_2 = \dots = \beta_p = 0 \\ H_1 : \exists i \in \{1, 2, \dots, p\}, \beta_i \neq 0 \end{aligned} \quad (13)$$

If the original hypothesis H_0 holds, i.e. $\beta_i = 0$, construct the F statistic:

$$F = \frac{ESS / p}{RSS / N - p - 1} \sim F(p, N - p - 1) \quad (14)$$

The F statistic should be satisfied given the confidence level $1 - \alpha$:

$$P\{F > F_\alpha(p, N - p - 1)\} < \alpha \quad (15)$$

According to the principle of “small probability events will not occur in a test”: under the premise that H_0 is correct, if $F > F_\alpha(p, N - p - 1)$, there are enough reasons to think that the original hypothesis is incorrect, so as to reject the original hypothesis.

Since the F-value is equivalent to the F-statistic, the p-value of the F-statistic is used in this paper to test the significance of the regression model.

The p-value is a value used to determine the result of a hypothesis test and is the probability that the overall population deviates more from the original hypothesis than the resulting sample when the original hypothesis proposition is valid. If the P-value is small, the original hypothesis proposition is to be doubted based on the theory that small probability events are unlikely to occur in a single trial or observation, and the original hypothesis proposition can be ruled not to be true in problems that require a decision to be taken. Generally the test level α is given in advance (in this paper we take $\alpha = 0.05$), and when $P \leq \alpha$, the original hypothesis is rejected, otherwise the original hypothesis is not rejected.

3. Findings and quantitative analysis

3.1. Analysis of English Vocabulary Learning Strategies

3.1.1. Learning strategy use. English vocabulary learning strategies are categorized into metacognitive and cognitive strategies. The use of university English vocabulary learning strategies is shown in Figure 1. Among the three metacognitive strategies, pre-planning is more frequently used with the highest mean value of 3.674, and self-assessment and self-monitoring are relatively less frequently used. Among the cognitive strategies, three strategies such as repetition, contextualization and note-taking were used more frequently, with mean values of 3.058, 3.092 and 3.544, respectively, while four strategies such as categorization, dictionary checking, doing exercises and making connections were used less frequently, with the mean results all below 3.

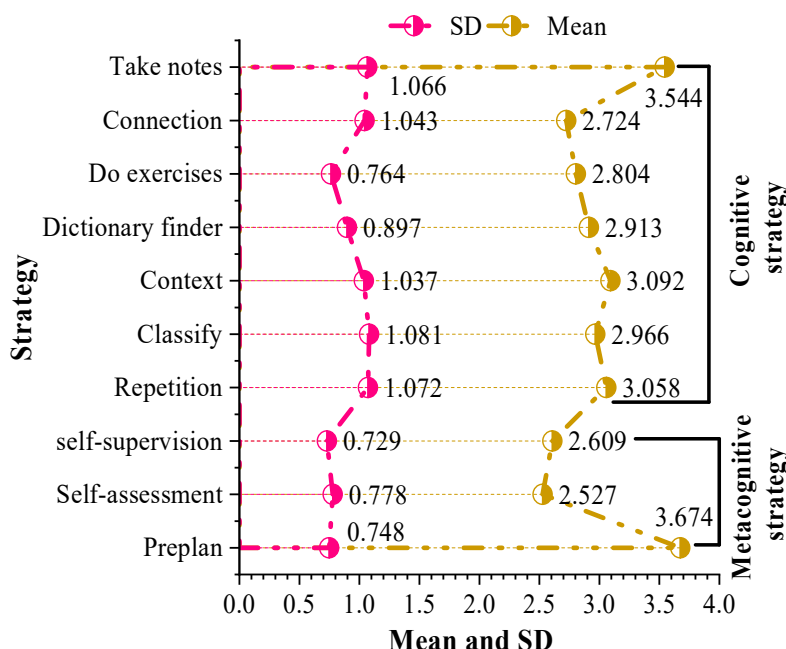


Fig. 1. The use of English lexical learning strategies

3.1.2. Analysis of differences in English proficiency. In order to compare the use of strategies among students with different levels of English achievement, 80 students who participated in the questionnaire survey were selected, and the first 15 were selected as the group of academically gifted students and the last 15 were selected as the group of academically disadvantaged students according to the English single-subject scores of the college entrance examination in descending order, totaling 30 students. The mean and standard deviation data of the two groups are shown in Table 1.

The mean values of metacognitive strategies used by the academic excellence group are higher than those of the struggling students group, with the mean values of the two groups being 3.186 and 2.610 respectively, which means that students with excellent academic performance are better able to regulate English vocabulary learning on a macro level, better able to plan in advance, self-evaluate and self-supervise, and more conscious of their learning.

In terms of the use of cognitive strategies, both groups of students can use certain vocabulary learning strategies. In terms of repetition, contextualization and doing exercises, the mean value of the struggling group is slightly higher than that of the gifted group. For several other strategies, including categorization, dictionary searching, linking and note-taking, the means of the gifted group are higher than those of the struggling group, and the biggest difference is in the linking strategy, with the means of the gifted and struggling groups being 3.297 and 2.306 respectively.

Table 1. The average and standard deviation data of the two groups

Categories	Strategy	Mean		Standard deviation	
		Excellent learning	Poor learning	Excellent learning	Poor learning
Metacognitive strategy	Preplan	3.662	2.771	0.966	0.952
	Self-assessment	3.036	2.478	0.956	0.986
	self-supervision	2.859	2.581	0.963	1.069
Cognitive strategy	Repetition	3.139	3.163	1.002	0.931
	Classify	3.115	2.716	0.989	0.954
	Context	3.038	3.129	1.145	0.954
	Dictionary finder	3.275	2.701	0.997	1.052
	Do exercises	3.021	3.182	0.963	0.968
	Connection	3.297	2.306	1.009	0.953
	Take notes	3.797	3.474	1.271	1.191

3.1.3. Analysis of gender differences among students. The English vocabulary learning strategy use of students of different genders is shown in Table 2. Both male and female students can use various strategies in vocabulary learning, but female students use learning strategies more frequently than male students, and the overall English vocabulary learning strategy use of female and male students is 3.240 and 2.790 respectively. Among them, the greater differences are in the three strategies of self-assessment, self-monitoring, dictionary checking, and note taking, with the difference values of 0.929, 0.670, 0.633, and 1.116.

Table 2. The use of English lexical strategies for different gender students

Categories	Strategy	Mean		Standard deviation	
		Female	Male	Female	Male
Metacognitive strategy	Preplan	3.812	3.446	0.964	0.998
	Self-assessment	3.134	2.205	0.982	0.978
	self-supervision	3.095	2.425	1.018	0.995
Cognitive strategy	Repetition	3.011	2.826	0.936	1.001
	Classify	2.939	2.761	0.973	0.958
	Context	3.029	3.167	0.956	0.973
	Dictionary finder	3.254	2.621	0.936	0.998
	Do exercises	2.878	2.758	0.981	0.995
	Connection	3.046	2.613	1.015	1.082
	Take notes	4.197	3.081	0.981	1.006

3.2. Correlation analysis and regression hypothesis

3.2.1. Correlation analysis. The degree of correlation between the independent and dependent variables is measured using Pearson's product-difference correlation. The Pearson correlation coefficient varies between -1 and +1, when $r > 0$ indicates that the two variables are positively correlated, $r < 0$ indicates that the two variables are negatively correlated, the larger the absolute value of r , the stronger the correlation between the two variables. It is generally accepted that: $|r| \geq 0.8$, it can be considered that the two variables are highly correlated. $0.6 \leq |r| < 0.8$, it can be considered that the two variables are highly correlated. $0.4 \leq |r| < 0.6$, it can be considered that the two variables are moderately correlated. $0.2 \leq |r| < 0.4$, it can be considered that the two variables are poorly correlated. $|r| < 0.2$, it can be considered that the two variables are basically uncorrelated.

The correlation analysis of the factors influencing the improvement of English vocabulary learning efficiency (IEIIE) is shown in Fig. 2, **significantly correlated at the 0.01 level (bilaterally) and *significantly correlated at the 0.05 level (bilaterally). Overall, the correlation coefficients between English language proficiency of high school students and each of the influencing factors are in the range of 0.402-0.563, and there is a moderate correlation between the improvement of English vocabulary learning efficiency at university and each of the influencing factors.

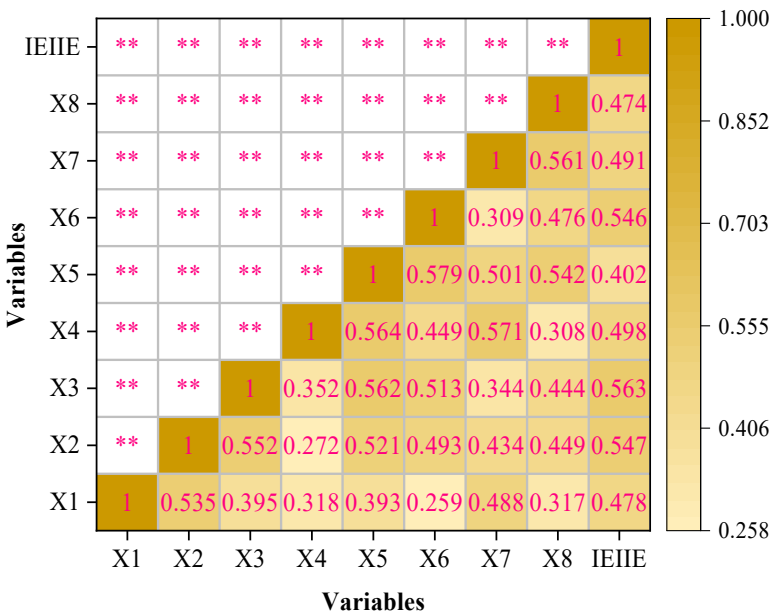


Fig. 2. The correlation between the influence factors of the efficiency of English lexical learning

3.2.2. Regression assumptions. Usually, Pearson correlation coefficient mainly describes the degree of correlation between variables, while the specific description of interdependent quantitative relationships between different variables, such as the relationship between the influence of variables, the degree of influence and the direction of the influence of the variables, etc., is realized through regression analysis. Based on the Pearson correlation analysis in the previous section, the following hypotheses are made in this study for the convenience of research and drawing valid conclusions:

H1: English knowledge has a positive influence on the improvement of English vocabulary learning efficiency.

H2: Learning attitude has a positive influence on the improvement of English vocabulary learning efficiency.

H3: Learning strategies have a positive effect on English vocabulary learning efficiency.

H4: Learning motivation has a positive influence on the improvement of English vocabulary learning efficiency.

H5: Learning tools have a positive influence on the improvement of English vocabulary learning efficiency.

H6: Teaching philosophy has a positive influence on the improvement of English vocabulary learning efficiency.

H7: Teaching methods have a positive influence on the improvement of English vocabulary learning efficiency.

H8: Teaching attitude has a positive influence on the improvement of English vocabulary learning efficiency.

3.3. Basic test and regression process

3.3.1. Basic tests

1) Diagnosis of multicollinearity. Multicollinearity refers to the phenomenon of model estimation distortion or difficulty in accurately estimating the relationship between variables due to precise correlation or high correlation between the dependent variables in a linear regression model. The presence of multicollinearity in multiple linear regression is mainly tested by eigenvalues, conditional indices, tolerance, and variance inflation factor (VIF). The study shows that the eigenvalues of all independent variables are greater than 0.01, the conditional indices are less than 30, the tolerances are greater than 0.01, and the VIF values are less than 10, which indicates that there is no multicollinearity among the independent variables. Regression analysis of multiple covariance diagnosis is shown in Table 3, the results of multiple covariance test in this study are less than 10, which meets the standard requirements, so it is suitable for regression analysis.

Table 3. Multiple conlinear diagnosis of regression analysis

Model	Common linear statistics			
	Eigenvalue	Conditional index	Allowance	VIF
(Constant)	7.659	1.000		
X1	0.058	12.596	0.655	1.401
X2	0.045	13.273	0.708	1.445
X3	0.039	11.419	0.611	1.761
X4	0.034	14.781	0.694	1.715
X5	0.037	12.261	0.688	1.754
X6	0.023	11.333	0.746	1.642
X7	0.014	13.438	0.795	1.718
X8	0.015	15.328	0.673	1.345

2) Normal distribution of residuals and variance chi-square test. The histogram of regression-standardized residuals and the scatter plot of regression-standardized residuals are used to determine whether the residuals are normally distributed and whether there is heteroscedasticity. Figure 3 shows the histogram of regression standardized residuals, and the residuals conform to the normal distribution, indicating that the fitting effect is good. Regression standardized residuals scatter plot shown in Figure 4, in a specific range of predicted values no matter how to change, the corresponding residuals are always in the vicinity of the 0 level line, the amplitude remains basically stable, there is no obvious sign of heteroskedasticity. It can be determined that there is no significant difference between the standardized residuals and the standardized normal distribution, which indicates that the conditions are met and the regression model is appropriate and feasible.

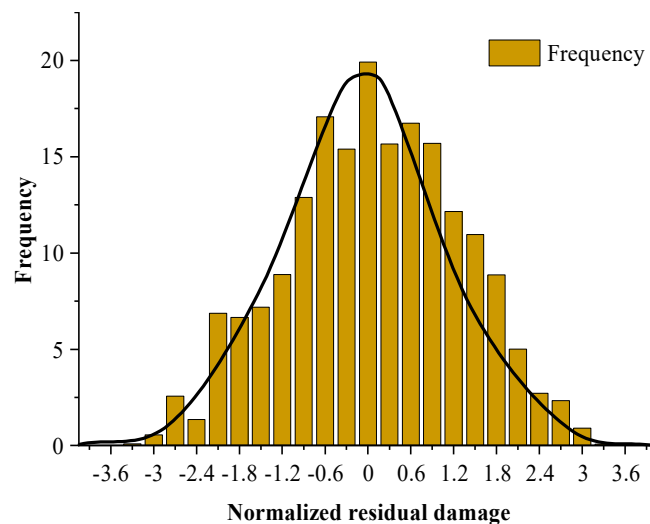


Fig. 3. Residual histogram of regression normalization

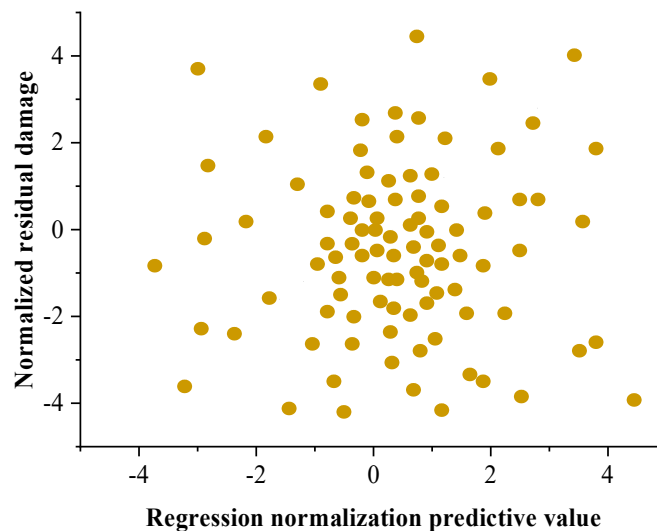


Fig. 4. Normalized residual dispersion point

3.3.2. Regression analysis. The eight independent variables were sequentially entered into the regression model, and the regression coefficients and significance tests of the equations are shown in Table 4. The standardized multiple linear regression equation is obtained as:

$$IEIIE = 0.363 \cdot X_1 + 0.416 \cdot X_2 + 0.482 \cdot X_3 + 0.201 \cdot X_4 + 0.276 \cdot X_5 + 0.358 \cdot X_6 + 0.457 \cdot X_7 + 0.322 \cdot X_8$$

From the results of the coefficients of the standardized multiple linear regression equations, each factor has a positive effect on the improvement of English vocabulary learning efficiency ($p < 0.01$), i.e., English knowledge X_1 , learning attitude X_2 , learning strategy X_3 , learning motivation X_4 , learning tools X_5 , teaching philosophy X_6 , teaching style X_7 , and teaching attitudes X_8 have a significant positive effect on the improvement of English vocabulary learning efficiency, supporting the Hypotheses $H_1 \sim H_8$. Among them, learning strategy X_3 , teaching style X_7 and learning strategy X_3 have the greatest influence, with the values of standardized regression coefficients of 0.482, 0.457 and 0.416, respectively.

Table 4. Regression coefficient and significance test of equation

Model	Nonnormalized coefficient		Normalized coefficient		
	B	SD	Beta	T	Sig.
(Constant)	0.595	0.075		10.610	0.027
X1	0.358	0.039	0.363	1.443	0.001
X2	0.424	0.051	0.416	1.418	0.003
X3	0.489	0.065	0.482	2.794	0.000
X4	0.208	0.074	0.201	1.696	0.009
X5	0.248	0.084	0.276	2.269	0.001
X6	0.371	0.082	0.358	3.621	0.007
X7	0.458	0.065	0.457	1.485	0.002
X8	0.317	0.053	0.322	1.671	0.006

4. Strategies for Improving the Efficiency of English Vocabulary Learning in Universities

Vocabulary is the foundation of language learning. Both mother tongue and second language learning require learners to master a large amount of vocabulary. Based on the previous quantitative analysis of the efficiency improvement of college English vocabulary learning, this chapter discusses the strategies for improving the efficiency of college English vocabulary learning.

4.1. Innovative teaching methods

University teachers can study a variety of teaching methods to continuously improve students' interest in learning. For example, when teaching words, pictures can be used to strengthen students' memory, and students can also be asked to make sentences using the words they have just learned, integrating teaching strategies with small games to stimulate students' interest in learning and improve their motivation to learn vocabulary.

4.2. Memorization in combination with word classes

Students strengthen their understanding of vocabulary with the help of the connection between similar words, and whenever they see a word they will be able to enumerate its different forms and its near-synonyms and antonyms to expand from one word to many words, thus forming a chain vocabulary learning approach that can improve the efficiency of vocabulary learning of students.

4.3. Expanded reading

The learning of English is usually divided into four parts: listening, speaking, reading and writing, and reading is a better way to improve vocabulary in the process of students' English learning. By expanding the scope of reading and strengthening the "interaction" with vocabulary, the vocabulary reading capacity can be improved.

4.4. Vocabulary association learning

The learning of English vocabulary is also inseparable from daily life. Nowadays, many signs or places in public places are labeled in English, and students should pay more attention to these aspects in their lives. Mastering vocabulary through their own observations can deepen students' memory and stimulate their thinking at the same time. When memorizing vocabulary from textbooks, students can also develop an associative approach, linking abstract words with real objects in daily life and creating situations to deepen their understanding.

5. Conclusion

The importance of English vocabulary learning as the foundation of English learning is self-evident. This study investigates the improvement of college English vocabulary learning efficiency, analyzes the use of English vocabulary learning strategies, and mines the influencing factors of the improvement of English vocabulary learning efficiency based on the multiple regression model, and then proposes the strategies for the improvement of college English vocabulary learning efficiency.

1) Pre-planning (3.674) is the most frequently used metacognitive strategy, while repetition, contextualization and note-taking are the more frequently used cognitive strategies, with statistical means greater than 3. Students with good academic performance use metacognitive strategies more, and students with poor academic performance use more strategies such as repetition, contextualization and doing exercises. Female students (3.240) used various vocabulary learning strategies more frequently than male students (2.790).

2) The positive effects of eight own environment and external environment factors on the improvement of English vocabulary learning efficiency were verified by multiple regression analysis, and all the hypotheses made in the study were verified. Among them, learning strategy, teaching style and learning strategy have the most significant effect on English vocabulary learning, with standardized coefficients of 0.482, 0.457 and 0.416 respectively.

College English vocabulary learning needs to guide students to adopt effective strategies, and students can improve the efficiency of vocabulary learning through word class memorization, expanding reading and vocabulary association learning. At the same time, teachers should innovate teaching methods to stimulate students' interest in vocabulary learning and cultivate their ability to learn vocabulary independently. Only by memorizing vocabulary scientifically can students really improve the efficiency of vocabulary mastery and use vocabulary appropriately in language learning and communication.

References

- [1] Alsalihi, H. D. (2020). Main difficulties faced by EFL students in language learning. *Journal of the College of Education for Women*, 31(2), 19-34.
- [2] Misbah, N. H., Mohamad, M., Yunus, M. M., & Ya'acob, A. (2017). Identifying the factors contributing to students' difficulties in the English language learning. *Creative Education*, 8(13), 1999-2008.
- [3] Hossain, M. A. (2018). Difficulties of learning English Language at the secondary level: A case study of Thakurgaon district. *Journal of Education and Training*, 5(2), 165-181.
- [4] Chaikovska, O., & Zbaravska, L. (2020). The efficiency of Quizlet-based EFL vocabulary learning in preparing undergraduates for state English exam. *Advanced education*, 84-90.
- [5] Schuetze, U. (2017). Efficiency in second language vocabulary learning. *Die Unterrichtspraxis/Teaching German*, 50(1), 22-31.
- [6] Tsai, C. C. (2018). A comparison of EFL elementary school learners' vocabulary efficiency by using flashcards and augmented reality in Taiwan. *The New Educational Review*, 51, 53-65.
- [7] Bai, Z. (2018). An analysis of English vocabulary learning strategies. *Journal of language Teaching and Research*, 9(4), 849-855.
- [8] Wu, Q., Zhang, J., & Wang, C. (2020). The effect of English vocabulary learning with digital games and its influencing factors based on the meta-analysis of 2,160 test samples. *International Journal of Emerging Technologies in Learning (iJET)*, 15(17), 85-100.
- [9] Tyng, C. M., Amin, H. U., Saad, M. N., & Malik, A. S. (2017). The influences of emotion on learning and

memory. *Frontiers in psychology*, 8, 235933.

- [10] Tan, J., Mao, J., Jiang, Y., & Gao, M. (2021). The influence of academic emotions on learning effects: a systematic review. *International journal of environmental research and public health*, 18(18), 9678.
- [11] Wang, X. (2022). Influences of learning emotion on learning outcome in online teaching mode. *International Journal of Emerging Technologies in Learning (iJET)*, 17(8), 126-139.
- [12] Lv, J., & Yang, J. (2023). Prediction of College Students' Classroom Learning Effect Considering Positive Learning Emotion. *International Journal of Emerging Technologies in Learning (Online)*, 18(5), 161.
- [13] Li, V. (2017). Social media in English language teaching and learning. *International Journal of Learning and Teaching*, 3(2), 148-153.
- [14] Al Arif, T. Z. Z. (2019). The use of social media for English language learning: An exploratory study of EFL university students. *Metathesis: Journal of English Language, Literature, and Teaching*, 3(2), 224-233.
- [15] Erarslan, A., & Topkaya, E. Z. (2017). EFL students attitudes towards e-learning and effect of an online course on students success in English. *The Literacy Trek*, 3(2), 80-101
- [16] Jelena, J., Bojana, K., & Rakić, D. (2022). THE EFFECTS OF ONLINE ENGLISH CLASSES ON STUDENTS'MOTIVATION TO LEARN ENGLISH LANGUAGE. *Journal of Language & Literary Studies/Folia Linguistica & Litteraria*, (39).
- [17] Amin, B., Rafiq, R., & Mehmood, N. (2020). The impact of social media in English language learning. *Journal of Critical Reviews*, 7(10), 3126-3135.
- [18] Lee, J. H., Ahn, J. J., & Lee, H. (2022). The Role of Motivation and Vocabulary Learning Strategies in L2 Vocabulary Knowledge: A Structural Equation Modeling Analysis. *Studies in Second Language Learning and Teaching*, 12(3), 435-458.
- [19] Isakovna, T. N., & Kosimov, A. (2023, April). THE IMPORTANCE OF MOTIVATION TO IMPROVE STUDENTS'VOCABULARY SKILLS IN THE EFL CLASSROOM. In *Integration Conference on Integration of Pragmalinguistics, Functional Translation Studies and Language Teaching Processes* (pp. 129-133).
- [20] Zhang, Y., Lin, C. H., Zhang, D., & Choi, Y. (2017). Motivation, strategy, and English as a foreign language vocabulary learning: A structural equation modelling study. *British Journal of Educational Psychology*, 87(1), 57-74.
- [21] Tanaka, M. (2017). Examining EFL vocabulary learning motivation in a demotivating learning environment. *System*, 65, 130-138.
- [22] ALBODAKH, M. K. K., & Cinkara, E. (2017). The Relationship between Learner Motivation and Vocabulary Size: The Case of Iraqi EFL Classrooms. *Arab world English journal*, 8(2).
- [23] Susanto, A., Halim, F. A., & Thasimmim, S. N. (2019). Vocabulary learning strategies, vocabulary skills, and integrative motivation levels among university students. *Vocabulary Learning Strategies, Vocabulary Skills, and Integrative Motivation Levels among University Students*, 8(5C), 323-334.
- [24] Ghalebi, R., Sadighi, F., & Bagheri, M. S. (2020). Vocabulary learning strategies: A comparative study of EFL learners. *Cogent Psychology*, 7(1), 1824306.
- [25] Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, 40(3), 651-675.
- [26] Orogun, A. O., Ariyo, O. E., Ajayi, O. O., & Onyekwelu, B. (2021). A Multiple Linear Regression Model for Analyzing and Predicting Educational Development in Nigeria. *NIPES-Journal of Science and Technology Research*, 3(1).
- [27] Radhy, Z. H. (2019). Application of Multiply Regression Linear Model and New Technology Method in

- Estimating Learning and Education of Students. *International Electronic Journal of Mathematics Education*, 14(1), 87-90.
- [28] Fang, H., & Caili, L. (2021, September). The Application of Big Data Analysis in College English Classroom Vocabulary Memory and Learning Efficiency Teaching. In *2021 4th International Conference on Information Systems and Computer Aided Education* (pp. 370-375).
- [29] Aljburi, M. A., & Khaghaninejad, M. S. (2024). Exploring the efficiency of associative vocabulary teaching strategies to foreign language learners. *International Journal of Instruction*, 17(3), 291-314.
- [30] Mahdi, H. S. (2018). Effectiveness of mobile devices on vocabulary learning: A meta-analysis. *Journal of educational computing research*, 56(1), 134-154.
- [31] Li, M. (2024). Exploration of the Application of English Vocabulary Adaptive Learning Model in English Teaching and Learning. *Journal of Electrical Systems*, 20(9s), 1333-1338.
- [32] Chang, L. (2024, June). Improvement of Vocational English Vocabulary Learning Based on Adaptive Algorithms. In *Proceedings of the 2024 International Conference on Intelligent Education and Computer Technology* (p. 1).
- [33] Hou, Z. (2021, February). Research on adopting artificial intelligence technology to improve effectiveness of vocational college English learning. In *Journal of Physics: Conference Series* (Vol. 1744, No. 4, p. 042122). IOP Publishing.
- [34] Triwardani, H. R. (2022). Building Vocabulary through song as effective method in learning English. *Lingua*, 18(1), 98-111.
- [35] Phisutthangkoon, K., & Panich, M. (2016). Effectiveness of English song activities on vocabulary learning and retention. In *The European Conference on Language Learning* (pp. 1-12).
- [36] Bahanshal, D. (2015). The effectiveness of vocabulary learning strategies on English language acquisition of the Saudi learners. *International Journal of Humanities and Social Sciences*, 1(1), 47-56.