Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

A Study of Using Linear Regression Analysis Model to Enhance the Teaching of English Reading Comprehension Skills in Colleges and Universities

Qingling Wang[1,✉], Guangjie Shen[1], Na Zhang[1]

[1] *School of International Cooperation and Exchange, Weifang Engineering Vocational College, Qingzhou, Shandong, 262500, China*

ABSTRACT

This study explores the main influencing factors of college teachers' ability to teach English reading comprehension through quantitative analysis. In this paper, we designed the scale of "Questionnaire on Teaching Ability of College Teachers' English Reading Comprehension" and selected the group of M college teachers as the target of the survey. And on the basis of the collected data, using SPSS software, T-test, correlation analysis and multiple linear regression were carried out. The results showed that there was a significant difference ($P<0.05$) between the teaching effectiveness of teachers in English reading comprehension skills when their education level was below 30 years old or college and below, and that of teachers aged 31 to 40 years old or other highly educated teachers. There is a statistically level difference ($P<0.05$) between different categories of teachers in both logical reasoning and information processing skills. Teachers' teaching ability passed the significance level test ($P < 0.05$) with all four independent variables. Their effects on teaching ability are, in descending order: language comprehension ability, information processing ability, logical reasoning ability and cultural comprehension ability, with corresponding regression coefficients of 0.3076, 0.2867, 0.2484 and 0.1225, respectively. It is possible to enhance the college English reading comprehension teaching.

Keywords: Quantitative Analysis, English Reading Comprehension, Multiple Linear Regression Analysis, Least Squares, Pearson's Correlation Coefficient

1. Introduction

As one of the important courses in colleges and universities, English has an important impact on students' future career development. At present, in the English teaching curriculum of colleges and universities, the proportion of knowledge points related to college students' English reading comprehension is also relatively high, and it can be said that English reading ability is an indispensable

✉ Corresponding author.
E-mail address: 16652859234@163.com (Q. Wang).
Received 12 February 2024; Revised 15 April 2024; Accepted 25 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-242

part of college students' comprehensive English literacy [1-2]. Generally speaking, reading is a mutual and flexible language activity, in which readers can consciously read the material, obtain key information, and then construct the meaning of the text. Therefore, reading comprehension is not just a simple translation of words, but a process in which multiple factors communicate with each other, requiring in-depth understanding and decoding of the complex information read [3-4]. However, in order to truly decode the complex information and deeply understand the meaning of the text, it is necessary to understand the deep structure behind the semantics [5]. No matter what kind of English reading materials, the creators of the materials will more or less provide students with some keywords or clues to the topic sentences, and students can gradually obtain the main information under the guidance of these clues. However, some students who are relatively poor in English reading ability have difficulty in finding the direction in the process of reading the material, either simply understanding the literal information from the first and last sentences, or obtaining the information through a certain key word, and do not have a deep understanding of the text. Especially once encountering words with multiple meanings, students with insufficient vocabulary reserves may be limited to a single layer of meaning, thus directly misunderstanding the whole reading material. In order to effectively improve the effectiveness of English reading teaching, teachers should innovate English reading teaching strategies, integrate statistical analysis techniques into English reading teaching, and boost the overall improvement of college students' English reading ability.

Lustyantie, N. et al. examined the effects of reading interest and achievement motivation on students' English reading comprehension ability, and the quantitative study verified the facilitating effect of reading interest and achievement motivation, so actively cultivating students' interest and stimulating their motivation during the learning process is an important measure to improve the comprehension ability of English texts [6]. Syafitri, N. conducted a study on the correlation between students' reading habits and reading comprehension skills during English language learning and the results of linear regression analysis showed that there was a significant positive correlation between the independent variable reading habits and the dependent variable reading comprehension skills [7]. Xu, H. et al. analyzed the differences in the prediction of text comprehension (TBC) and higher-order comprehension (HOC) abilities in English reading comprehension by motivational factors, and found that intrinsic motivation was a stronger predictor of HOC while extrinsic motivation was a stronger predictor of TBC, and discussed the implications thereof for English language teaching and learning [8]. Syafitri, N. showed that students' reading habits and teachers' classroom instruction are important components of the process of teaching reading comprehension, explored the correlation between the two on students' reading comprehension, and found that students' reading habits were positively correlated with reading comprehension, while teachers' styles were negatively correlated with reading comprehension [9]. Yuliawati, L. argued that students' grammatical and vocabulary mastery contributed to their English reading comprehension by performing multiple linear regression on the collected questionnaire data, and the results showed that both of the proposed independent variables significantly enhanced students' English reading comprehension [10]. Makebo, T. H. et al. explored the relationship between students' reading fluency and comprehension by designing two parameters of reading fluency, namely reading accuracy and reading rate, and the study showed that there was a significant positive correlation between the two parameters of the independent variables and that the two independent variables together contributed to the improvement of students' reading comprehension [11]. Rosyidin, I. et al. compared the effect of PBL and TBL teaching modes on students' English reading comprehension, and the experiment found that the effect of PBL teaching mode in enhancing students' English reading text comprehension was at a significant level and much higher than that of students in TBL teaching mode [12]. Koosha, M. et al. studied the role of self-esteem and autonomy of English language learners on the improvement of their own reading comprehension ability by administering questionnaires to students from different schools with different levels of English language proficiency, and found that there was a significant correlation between the indicators

of self-esteem and autonomy and reading comprehension, and that there was a significant difference between the two indicators in terms of their prediction of reading comprehension [13]. Aini, N. et al. assessed the effect of TELLS strategies on students' reading comprehension using a quasi-experimental, non-equivalent control group design, which was analyzed by collecting students' post-instruction pre-test and post-test scores, indicating that TELLS instructional strategies were more effective in improving students' reading comprehension [14].

In this study, multiple linear regression analysis was used to collect data on college teachers' teaching ability in English reading comprehension through a questionnaire survey and establish a multiple linear regression analysis model. Then the data from the questionnaire survey were processed using correlation analysis, t-test and multiple regression analysis. Finally, the differences of "education, age and educational nature" in teachers' English reading comprehension teaching ability were analyzed, and the correlation between teachers' teaching ability and four factors of "language comprehension ability, cultural comprehension ability, logical reasoning ability and information processing ability" was further explored. According to the survey results, effective teaching improvement strategies were proposed in four aspects: teaching philosophy, teaching concept, teaching method and professional development.

2. Construction of Linear Regression Analysis Model Based on English Reading Comprehension Ability

2.1. Study design

2.1.1. Data sources. In this study, the survey of “Questionnaire on Teaching Ability of College Teachers' English Reading Comprehension Skills” was conducted in September-November 2023 in X province with the group of teachers in M colleges and universities as the target group. Teachers from five colleges and universities were selected as samples for this survey. The questionnaires were distributed 350 in total and 329 valid questionnaires were recovered, with an effective rate of 94%. Using the information recovered from the pre-survey, the preliminary questionnaire reliability and validity test was done, and the results showed that the overall reliability of the questionnaire was 0.8947, which was higher than 0.8, indicating that the questionnaire had good reliability and stability. The overall validity of the questionnaire is 0.9012, which is higher than 0.8, indicating that the questionnaire has good accuracy and operability.

2.1.2. Questionnaire design. This study explores the main influencing factors of college teachers' English reading teaching ability through quantitative analysis. The questionnaire covers two main parts: basic personal information, English reading teaching competence level, and factors influencing English reading teaching competence. First, the basic personal information will be mainly used as the control variables in this study, which mainly includes, gender, age, and education of college teachers. Secondly, college teachers' English reading teaching ability as the dependent variable of this study is divided into four ability dimensions, on the basis of which, this study further describes and divides the four dimensions, resulting in a total of nine secondary indicators, and college teachers' English reading teaching ability dimensions are shown in Table 1

Table 1. Teacher English reading comprehension ability dimension

Capacity dimension	Secondary indicator	Specific content
A: Language comprehension	A1: Vocabulary knowledge	Identify and understand professional terms and academic vocabulary
	A2: Grammatical	Understand complex sentence patterns and grammatical

	structure	structures
	A3: Discourse analysis	Understand the relationship between the overall structure and paragraph of the article
B: Cultural understanding	B1: Cultural background	Understand the cultural elements and backgrounds in the text
	B2: Cross-cultural communication	Understand the expression and communication habits of different cultural contexts
C: Logical reasoning ability	C1: Critical thinking	Evaluate the rationality of the argument and the validity of the evidence
	C2: Logical analysis	Understand the structure and logical relationship of the article
D: Information ability	D1: Information filtering	Select key information from a large amount of information
	D2: Information integration	Integrate different sources of information into a complete understanding

2.2. Multiple linear regression models

2.2.1. Multiple linear regression modeling. Multiple linear regression model [15] General form:

$$y_i = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \cdots + \alpha_k x_{ki} + \varepsilon_i i = 1,2,\cdots,n \tag{1}$$

In the model y_i is referred to as the dependent variable while $x_{1i}, x_{2i}, \cdots, x_{ki}$ is referred to as the independent variable. At $k = 1$, the above equation is known as a univariate linear regression model and at $k > 2$, it is known as a multiple linear regression model. The dependent variable y_i consists of an error term random variable ε_i and a linear function $\alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \cdots + \alpha_k x_{ki}$ of the k independent variables.

The parameter estimation of the multiple linear regression model is much more complicated than that of the univariate linear regression model, and in order to facilitate the computation and analysis, and to facilitate the generalization of the results from univariate totals to multivariate totals in general, the matrix is introduced as a tool to simplify the computation and analysis.

A set of random samples in the overall set of $(y_i, x_{ii}, x_{2i}, \cdots, x_{ki})$, $i = 1,2,\cdots,n$. Then the above formula can be written as the following matrix representation:

$$\begin{aligned} y_1 &= \alpha_0 + \alpha_1 x_{11} + \alpha_2 x_{21} + \cdots + \alpha_k x_{k1} + \varepsilon_1 \\ y_2 &= \alpha_0 + \alpha_1 x_{12} + \alpha_2 x_{22} + \cdots + \alpha_k x_{k2} + \varepsilon_2 \\ &\vdots \\ y_n &= \alpha_0 + \alpha_1 x_{1n} + \alpha_2 x_{2n} + \cdots + \alpha_k x_{kn} + \varepsilon_n \end{aligned} \tag{2}$$

Where k - number of independent variables; $\alpha_0, \alpha_1, \cdots, \alpha_k$ - parameters to be determined; ε_i - random variables. Using matrix operations is expressed as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{3}$$

Denote $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}$, $\alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix}$, $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$.

This can be expressed in terms of a matrix:

$$y = X\alpha + \varepsilon \tag{4}$$

Noting $\hat{\alpha} = \begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_k \end{bmatrix}$, $e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$, the matrix representation of the sample is:

$$y = X\hat{\alpha} + e \tag{5}$$

where $\hat{\alpha}_0, \hat{\alpha}_1, \cdots, \hat{\alpha}_k$ is the fitted value of $\alpha_0, \alpha_1, \cdots, \alpha_k$ respectively, the regression equation is:

$$\hat{y}_i = \hat{\alpha}_0 + \hat{\alpha}_1 x_{1i} + \hat{\alpha}_2 x_{2i} + \cdots + \hat{\alpha}_k x_{ki} \tag{6}$$

Where $\hat{\alpha}_0$ - constant;
$\hat{\alpha}_1, \cdots, \hat{\alpha}_k$ - partial regression coefficient.

2.2.2. Basic assumptions of the model. The basic assumptions are intended to facilitate the estimation of the parameters of the model, and the following assumptions are made for equation (2):

Assumption 1 The model is assumed to be parametrically linear.

Assumption 2 The independent variables are independent of the random error term, i.e.:

$$\operatorname{cov}\left(x_{ji}, \varepsilon_i\right) = 0, j = 1, 2, \cdots, k \tag{7}$$

Assumption 3 Zero mean assumption, i.e.:

$$E\left(\varepsilon_i\right) = 0, i = 1, 2, \cdots, n \tag{8}$$

Assumption 4 Homoskedasticity assumption, i.e.:

$$\operatorname{var}\left(\varepsilon_i\right) = \sigma^2, i = 1, 2, \cdots, n \tag{9}$$

Assumption 5 No autocorrelation assumption:

$$\operatorname{cov}\left(\varepsilon_i, \varepsilon_j\right) = 0, i \neq j, i = 1, 2, \cdots, n, j = 1, 2, \cdots, n \tag{10}$$

Assumption 6 Normality is assumed, viz:

$$\varepsilon_i \sim N\left(0, \sigma^2\right), i = 1, 2, \cdots, n \tag{11}$$

2.2.3. Estimation of model parameters. After specifying the regression theoretical model, the unknown parameters of the model are estimated. Ordinary least squares [16] is the most commonly used method for estimating the unknown parameters and is also the classical estimation method. When there are regression problems that do not meet the model assumptions, some new methods can be used, which are also based on the ordinary least squares method, such as partial least squares estimation, principal component regression, ridge regression and so on. Due to the fact that the computation is very large when there are many parameter variables, it is necessary to use computer software for the operation.

Based on the principle of least squares, the estimate $\hat{\alpha}_i$ of α_i is to be satisfied:

$$Q = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \to \min \tag{12}$$

Derivation of the above equation gives:

$$\begin{cases} n\hat{\alpha}_0 + \left(\sum_{i=1}^{n} x_{1i}\right)\hat{\alpha}_1 + \cdots + \left(\sum_{i=1}^{n} x_{ki}\right)\hat{\alpha}_k = \sum_{i=1}^{n} y_i \\ \left(\sum_{i=1}^{n} x_{1i}\right)\hat{\alpha}_0 + \left(\sum_{i=1}^{n} x_{1i}^2\right)\hat{\alpha}_1 + \left(\sum_{i=1}^{n} x_{1i}x_{2i}\right)\hat{\alpha}_2 + \cdots + \left(\sum_{i=1}^{n} x_{1k}x_{ki}\right)\hat{\alpha}_k = \sum_{i=1}^{n} x_{1i}y_i \\ \cdots\cdots \\ \left(\sum_{i=1}^{n} x_{ki}\right)\hat{\alpha}_0 + \left(\sum_{i=1}^{n} x_{1i}x_{ki}\right)\hat{\alpha}_1 + \left(\sum_{i=1}^{n} x_{2i}x_{ki}\right)\hat{\alpha}_2 + \cdots + \left(\sum_{i=1}^{n} x_{ki}^2\right)\hat{\alpha}_k = \sum_{i=1}^{n} x_{ki}y_i \end{cases} \tag{13}$$

Is further written in matrix form:

$$Aa = B \tag{14}$$

Assume that the coefficient matrix A is full rank, where:

$$\begin{gathered} A = X^T X, B = X^T Y, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix} \\ Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, a = \begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_k \end{bmatrix} \end{gathered} \tag{15}$$

The least squares estimate is then:

$$a = A^{-1}B = \left(X^T X\right)^{-1} X^T Y \tag{16}$$

2.2.4. Model testing. After the unknown parameters of the model are estimated and the multiple linear regression equation is established, the multiple linear regression equation needs to be tested for significance.

1) Goodness-of-fit test of regression equation the magnitude of the coefficient of determination R^2 can indicate how well the estimated equation fits the sample observations, i.e. $SST_n = SSR + SSE$.

Where $SST = \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$ is the total sum of squared deviations, $SSR = \sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2$ is the regression sum of squares, which is a parameter that reflects the effect of regression, and $SSE = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$ is the residual sum of squares. y_i is the regression value on the ith sample point $\left(x_{1i}, x_{2i}, \cdots, x_{ki}\right)$ and $\overline{y}$ is the sample mean of y.

The correlation coefficient is:

$$r = \frac{\operatorname{cov}(X,Y)}{\sqrt{DX \cdot DY}}, \hat{r} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \hat{x}\right)^2} \cdot \sqrt{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}} \tag{17}$$

R^2 is the square of the correlation coefficient:

$$R^2 = 1 - \frac{SSE}{SST} \tag{18}$$

In one-way linear regression, if the value of R^2 is close to 1, then the regression equation is a good fit to the actual observations, and vice versa, the worse the fit.

In a real problem, if R^2 is increased by adding an independent variable, this would lead one to believe that the model can be made to fit better by adding the independent variable. However, the reality is that an increase in R^2 due to an increase in the number of independent variables does not correlate well with a good or bad fit, and then R^2 needs to be adjusted. This brings us to the adjusted goodness of fit:

$$R_1^2 = 1 - \frac{SSE}{n-k-1} \Big/ \frac{SST}{n-1} \tag{19}$$

Where $n-k-1$ - degrees of freedom (residual sum of squares); $n-1$ - degrees of freedom (overall sum of squares).

In multiple linear regression, the closer R_1^2 is to 1, the better the fit.

2) The test of significance of the regression equation can be illustrated by the F test for significance: $F-$ test of the sum of squares decomposition formula:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \tag{20}$$

Denoted as $L_{yy} = Q + U$, there is:

$$F = \frac{U}{Q/(n-2)} \sim F(1, n-2) \tag{21}$$

Rejecting domain $\chi_0 = \{F > F_{1-\alpha}(1, n-2)\}$ indicates that the regression is good.

2.3. Data analysis method

2.3.1. Correlation analysis. Correlation is a data analysis technique that describes and measures the relationship between two variables. The relationship between two variables includes the direction, form, and strength of the relationship. The direction of the relationship includes positive and negative correlation. The form of the relationship includes linear correlation, non-linear correlation. The strength of the relationship includes irrelevant, mild correlation, high correlation and so on. The result of correlation is often quantitatively described by the correlation coefficient, which is the value between -1 and +1, the sign indicates the direction of correlation, the negative sign indicates negative correlation, the positive sign indicates positive correlation, the number indicates the strength of correlation, 0 means there is no relationship, and 1 means complete correlation. Steps of correlation analysis: (1) Use the data of the two sets of variables to draw a scatter plot, and initially determine the direction of correlation and the form of correlation of the variables according to the scatter plot. (2) Solve the correlation coefficients of the two groups of variables to quantitatively describe the correlation between the two groups of variables.

Compared with Pearson correlation coefficient, Spearman correlation coefficient and Kendall harmony coefficient are rank correlation, i.e., the original data are first converted into ranking ranks, and then rank ranks are correlated, so they can be correlated and analyzed for sequential data and nonlinear data [17]. This paper will use formative assessment parameters and paper grades for correlation analysis, so as to verify the evaluation role of the parameters, the parameters and paper grades are in the form of equiproportional continuous values, so Pearson correlation coefficient is used to carry out the correlation analysis. The steps of correlation analysis using Pearson correlation coefficient are as follows:

Step 1: Calculate the correlation coefficient r. The formula for calculating the Pearson correlation coefficient is shown below:

$$r=\frac{\sum XY-\frac{\sum X\cdot\sum Y}{n}}{\sqrt{\left(\sum X^2-\frac{\left(\sum X\right)^2}{n}\right)\cdot\left(\sum Y^2-\frac{\left(\sum Y\right)^2}{n}\right)}} \tag{22}$$

Where X,Y denotes the data for both variables and n denotes the number sample size. Next the test of significance is performed.

Step 2: Formulation of Hypotheses. Null hypothesis H_0 is that there is no relationship between the two groups and alternative hypothesis H_1 is that there is a relationship between the two groups.

Step 3: Determine one- and two-tailed, significance level α, and degrees of freedom df. In this paper, there is no prediction in the direction of the relationship, so a two-tailed test is used. Significance level α is generally 0.05. Degrees of freedom $df=n-2$.

Step 4: Determine the level of significance and analyze the correlation. According to the one and two-tailed test, significance level α, degree of freedom df to find the critical value of the Pearson correlation, if the calculated correlation coefficient r is greater than the critical value, then the null hypothesis H_0 is rejected, that is, the two variables are significantly correlated. On the basis of significant correlation, r is analyzed and $r>0.7$ is usually considered as highly correlated, $r>0.5$ as moderately correlated and $r>0.2$ as mildly correlated.

The above calculations have been encapsulated into the pearsonr function of scipy (a third-party library for the Python language), so it is straightforward to use this function to calculate the Pearson correlation coefficients.

2.3.2. T-test. T-tests are used to determine whether a sample mean is significantly different from the overall mean.T-tests include the one-overall t-test and the two-overall t-test. The single overall t-test is used to determine whether the difference between the sample mean and the overall mean is significant. The double overall t-test is used to test whether the difference between two overall means is significant. In this paper, double overall t-test is used to verify whether there is a significant difference between different categories of teachers' ability to teach English reading comprehension.

The steps of the double overall t-test for independent samples are as follows:

Step 1: Formulate the hypothesis. Null hypothesis H_0 is that the two overall means are not significantly different and alternative hypothesis H_1 represents that the two overall means are significantly different. The formula representation is shown below:

$$\begin{cases}H_0:\mu_1-\mu_2=0\\H_1:\mu_1-\mu_2\neq 0\end{cases} \tag{23}$$

Step 2: Determine the direction of the test, significant level α, degree of freedom df, and critical t score. The hypotheses made in this paper will make predictions about the level of English reading comprehension of different categories of students, so the hypotheses are directional and one-tailed tests are used. Significant level α is usually taken as 0.05. Degree of freedom df can be obtained by adding the degrees of freedom of the two overall samples, let the two overall samples are n_1 and n_2, respectively, then $df=n_1-1+n_2-1$. Combined with the one-tailed test, the level of significance of α, the degree of freedom of df to check the t table of critical values to obtain the critical t-score.

Step 3: Test for homogeneity of variance using Hartley's maximum F value test. Let the variances of the two overall samples be S_1^2 and S_2^2, then the maximum F value is calculated as shown below:

$$F_{\max}=\frac{Max\left(S_1,S_2\right)^2}{Min\left(S_1,S_2\right)^2} \tag{24}$$

If $F_{\max} < 2$, then the two variances can be considered homogeneous according to the thumb principle. Otherwise, the T test cannot be performed and other statistical methods, such as non-parametric tests, need to be considered.

Step 4: Calculate the actual t score and compare it with the critical value t score. The formula for the process of calculating the actual t score is given below:

$$t_{obs} = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\left(n_1 - 1\right)S_1^2 + \left(n_2 - 1\right)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{25}$$

Where $\bar{X}_1 - \bar{X}_2$ is the difference between the actual sample means, $\mu_1 - \mu_2$ is the difference between the overall means, and $s_{\bar{X}_1 - \bar{X}_2}$ represents the standard error of the difference samples, assuming $\mu_1 - \mu_2 = 0$ that the numerator can therefore be simplified. After calculating the actual t scores are obtained, they are compared with the critical value t scores, and if $t_{obs} > t_{\text{Critical}}$, H_0 is rejected and H_1 is accepted, i.e., the two overall means are significantly different. Otherwise, accept H_0, i.e., the two overall means are not significantly different.

The process of calculating the double overall T scores for independent samples has been encapsulated into the ttest_ind function of the scipy library, and the calculation of the Hartley maximum F value test with 95% confidence intervals was implemented using the numpy library.

2.3.3. Multiple regression analysis. A multiple regression model is a statistical model used in statistics to analyze the relationship between two or more independent variables (explanatory variables) and a dependent variable (response variable). The components of a multiple regression model include:

Model definition (determination of independent variables, dependent variables and regression coefficients and errors, etc.), parameter estimation (the most commonly used method is the method of least squares, which estimates the regression coefficients by minimizing the sum of squares of the error terms), model assumptions, model diagnostics, variable selection, model estimation, non-linear models, interactions and higher-order terms, and treatment of categorical variables.

Multiple regression modeling is an important tool in data analysis and prediction and is widely used in the fields of economics, finance, social sciences and biostatistics.

3. Analysis of Factors Influencing the Teaching of English Reading Comprehension and Strategies for Improving It

3.1. Analysis of Factors Influencing Teachers' English Reading Comprehension Ability

3.1.1. Analysis of differences in educational qualifications. Multiple comparisons of different educational qualifications on the implementation of English reading instruction among college teachers are shown in Table 2. The results show that as far as English reading teaching implementation skills are concerned, the level of significance between teachers with college and below degrees and teachers with bachelor's, master's, and doctoral degrees, respectively, is 0.0359, 0.0487, and 0.0351 in order of significance (sig<0.05), which indicates that there is a significant difference between teachers with college and below degrees and teachers with bachelor's, master's, and doctoral degrees. The same significant difference was reflected between teachers with bachelor's and master's degrees (sig=0.0329<0.05). Whereas, the significance between teachers with Bachelor's and Doctoral and

Master's and Doctoral degrees were all greater than 0.05, there was no body difference significant at the 0.05 confidence level.

Table 2. Different degrees of teachers are teaching multiple comparisons

Dependent variable	Educational background	Educational background	Significance
English reading comprehension teaching implementation ability	Junior college	Undergraduate	0.0359
		Master	0.0487
		Doctor	0.0351
	Undergraduate	Junior college	0.0357
		Master	0.0329
		Doctor	0.1353
	Master	Junior college	0.0486
		Undergraduate	0.0332
		Doctor	0.5347
	Doctor	Junior college	0.0359
		Undergraduate	0.01368
		Master	0.5341

3.1.2. Analysis of differences in age. Multiple comparisons of different age zones in the implementation of English reading teaching ability of college teachers As seen in Table 3, in terms of the implementation of English reading teaching ability, the significance value between teachers under 30 years old and teachers between 31 and 40 years old is 0.0021, which is much less than 0.05, so the difference between the two embodies a statistically significant level of significance. The significance between teachers of other age groups is greater than 0.05 and does not show a significant difference at the 0.05 confidence level.

Table 3. The multiple comparisons of teachers in different ages

Dependent variable	Educational background	Educational background	Significance
English reading comprehension teaching implementation ability	Under 30	31~ 40 year-old	0.0021
		41~ 50 year-old	0.2342
		Over 51	0.2038
	31~ 40 year-old	Under 30	0.0031
		41~ 50 year-old	0.5349
		Over 51	0.9034
	41~ 50 year-old	Under 30	0.2358
		31~ 40 year-old	0.6356
		Over 51	0.8329
	Over 51	Under 30	0.2355
		31~ 40 year-old	0.9348
		41~ 50 year-old	0.8329

3.1.3. Analysis of differences in teaching competencies by nature of education. In this paper, Levene's chi-square test of variance and t-test of mean equation were conducted on the educational data of both teacher educators and non-teacher educators, and the differences in each dimension on teaching competence were analyzed as shown in Table 4. The data of each dimension in the table were analyzed one by one:

1) As far as language comprehension ability is concerned, the significance value of the result of Levene's test on variance equation is 0.212>0.05, the variance shows chi-square, while the result of the t-test of the mean value equation, which assumes equal variances, has sig=0.1012>0.05, so it can be judged that language comprehension ability does not reflect the difference in significant on the two types of education, namely, teacher training and non-teaching.

2) In terms of cultural comprehension ability, the significance value of the Levene's test result about the variance equation is 0.215>0.05, and the variance shows chi-square, while the sig of the t-test result of the mean equation assuming equal variances=0.0719>0.05, so it can be judged that there is no significant difference embodied in the cultural comprehension ability on the two types of education, teacher-training and non-teacher-training.

3) On logical reasoning ability, for Levene variance chi-square test results of the significance value is 0.037, did not reach the level of significance (0.05), the variance performance is not chi-square, so it is necessary to agree with the setting of the variance is not equal. The result of the t-test assuming unequal variances is sig=0.0035, which did not reach the significant difference value of 0.05, so it can be concluded that there is a statistically significant difference at the level of logical reasoning ability on the two types of education, teacher training and non-teacher training.

4) As far as the information processing ability is concerned, the significance value of the results of the test of the chi-square test about Levene's variance is 0.671, which reaches the level of significance (0.05), and the variance shows chi-square, while the significance of the results of the t-test of the equation of the mean, which assumes that the variances are equal, equals to 0.0136, which doesn't reach the level of significance (0.05), and therefore, it can be determined that, on the two educational types, the teacher-training and the non-trainers' type , there is a statistically significant difference at the level of significance as far as information processing skills are concerned.

Table 4. Differences in the teaching ability of each dimension

Capacity dimension	Hypothesis variance	Levene variance test		T test of the mean equation
		F	Sig.	sig. (2)
A: Language comprehension	Let's say the variance is equal	1.602	0.212	0.1012
	Let's say that the variance is not equal			0.0890
B: Cultural understanding	Let's say the variance is equal	1.585	0.215	0.0719
	Let's say that the variance is not equal			0.0609
C: Logical reasoning ability	Let's say the variance is equal	0.446	0.037	0.0035
	Let's say that the variance is not equal			0.0035
D: Information ability	Let's say the variance is equal	0.135	0.671	0.0136
	Let's say that the variance is not equal			0.0146

3.2. Correlation between factors influencing teachers' teaching competence

In this section, Pearson correlation analysis is used to study the independent variable to dependent variable correlation and multiple linear regression is used to study the degree of influence and significance of the independent variable on the dependent variable.

3.2.1. Correlation analysis between teachers' teaching competence and factors. The results of the correlation analysis between teachers' teaching ability and the factors are shown in Table 5. The results of the analysis show that the correlation coefficients between teachers' teaching ability and language comprehension ability, cultural comprehension ability, logical reasoning ability, and information processing ability are 0.698, 0.585, 0.699, and 0.676, respectively, and the two-tailed significant p of Pearson's correlation coefficients are all 0.000<0.05. Therefore, it can be concluded that language comprehension ability, cultural comprehension ability, logical reasoning ability, and information processing ability are all significantly correlated with teachers' ability to teach English reading comprehension.

Table 5. The teaching ability of teachers and the relevant analysis of various factors

Correlation content		A:Language comprehension	B:Cultural understanding	C:Logical reasoning ability	D:Information ability
Teaching ability	Pearson correlation	0.698**	0.585**	0.699**	0.676**
	sig. (2)	0.000	0.000	0.000	0.000

3.2.2. Regression Analysis of Teachers' Teaching Competence with Factors. The multiple linear regression analysis of teachers' teaching ability with the factors is shown by Table 6. It can be learned that the VIF values are below 2, which indicates that there is no covariance correlation between the factors. In this regression model, F=60.482, p=0.000<0.001 indicating that the established multiple linear regression model is significant. The D-W value of the model was 1.444, indicating that the regression model did not have first order autocorrelation. The histogram of regression standardized residuals is shown in Figure 1, the normal P-P plot of standardized residuals is shown in Figure 2, and the scatter plot of standardized residuals is shown in Figure 3. The data satisfy normal distribution and homoscedasticity. Therefore, the conditions for the use of multiple regression are met, and

multiple regression can be used to analyze the factors affecting teachers' teaching ability. Multiple linear regression analysis was conducted with teaching ability as the dependent variable and language comprehension ability, cultural comprehension ability, logical reasoning ability, and information processing ability as the independent variables. According to the results of the analysis, it can be seen that all four independent variables passed the significance level test ($P < 0.05$), indicating that the effect of the independent variables on the dependent variable teachers' ability to teach English reading comprehension is significant. As shown by the size of Beta value, the degree of influence of the four independent variables on teaching ability is in the following order: language comprehension ability (0.3076) > information processing ability (0.2867) > logical reasoning ability (0.2484) > cultural comprehension ability (0.1225). Language comprehension ability has the strongest influence on teachers' ability to teach English reading comprehension, followed by information processing ability, logical reasoning ability and cultural comprehension ability respectively.

Table 6. The teaching ability of teachers and the regression analysis of various factors

Ability dimension	Nonnormalized coefficient		Standardized coefficient beta	t	P	VIF
	B	Standard error				
Constant	0.8009	0.3188	-	4.2009	0.000***	-
A:Language comprehension	0.2987	0.3093	0.3076	5.3021	0.000***	1.4995
B:Cultural understanding	0.0935	0.0961	0.1225	2.2034	0.036**	1.4018
C:Logical reasoning ability	0.1614	0.0043	0.2484	3.8977	0.000***	1.7015
D:Information ability	0.2216	0.2041	0.2867	4.9161	0.000***	1.4986
R2	0.658					
Adjust R2	0.647					
F	F=60.482,P=0.000***					
D-W value	1.444					

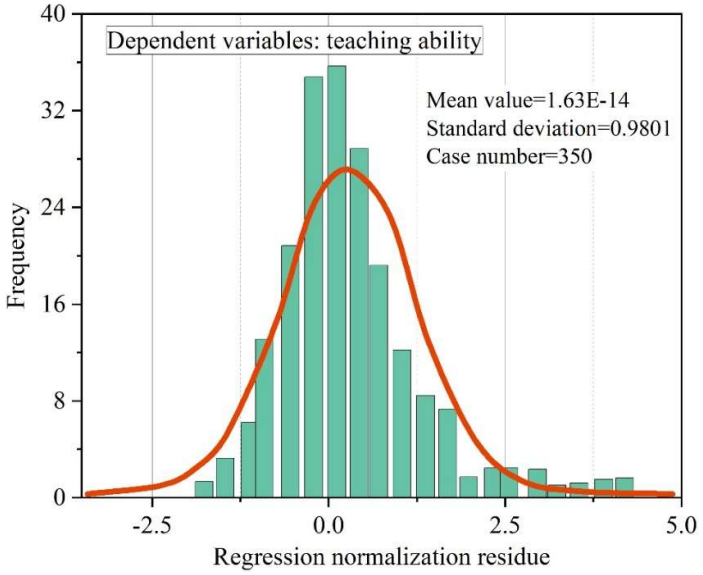

Fig. 1. Regression standardization

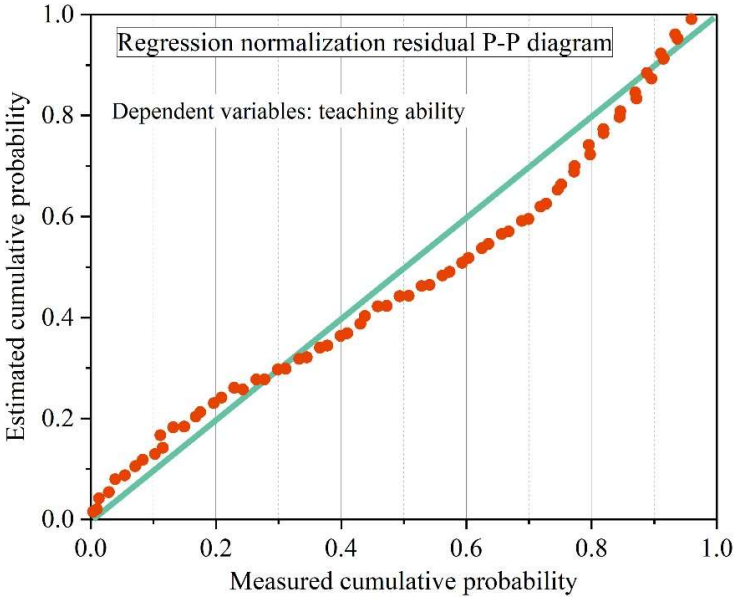

Fig. 2. Standardized residual normal p-p diagram

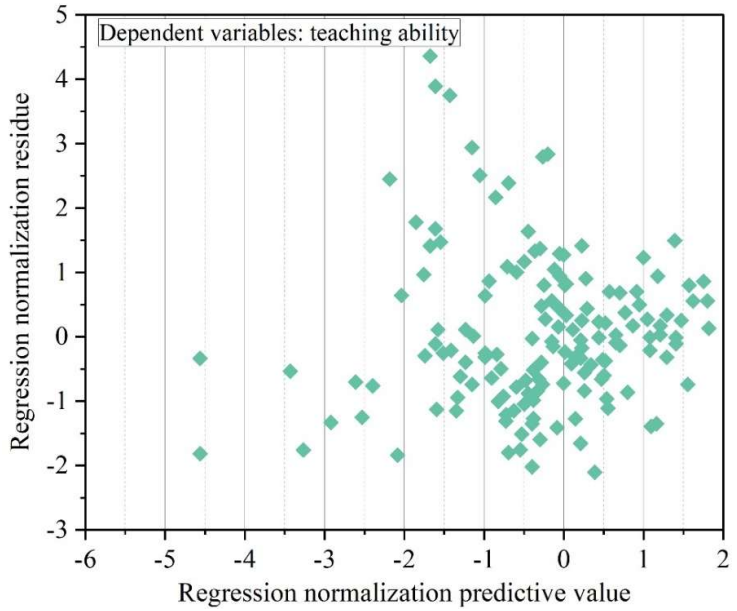

Fig. 3. Standardized residual map

3.3. Strategies for Improving Higher Education Teachers' English Reading Comprehension Skills

3.3.1. Updating teaching concepts. Update the teaching concept and optimize the teaching mode of English reading class. Teachers should abandon the traditional teacher-centered classroom teaching mode, fully consider the physiological and psychological characteristics of students, and analyze the students and the classroom. Both students' learning process and self-development process are inseparable from their own learning initiative. And students' learning initiative will be largely influenced by their own participation in teaching activities and their interest in teaching activities. Learners are the constructors of knowledge, and their knowledge construction activities are the key factors affecting whether the teaching goals can be effectively achieved. Therefore, teachers should emphasize on triggering and promoting students to actively carry out their own knowledge construction activities in reading teaching. Students are the main body of reading learning, the

initiative of learning should be in the hands of students. Teachers should pay attention to guiding students to actively think and practice language, and promote the comprehensive development of students' comprehensive use of English.

3.3.2. Transforming teachers' concepts of teaching reading. Changing teachers' reading teaching concepts and establishing students' correct reading concepts. Teachers should gradually transform themselves into “guides” and “evaluators” in teaching activities, and actively guide students to use reading strategies correctly, so as to lay a solid foundation for the continuous improvement of students' English reading comprehension ability. Teachers should consciously inculcate the concept of reading strategies in students, help students recognize the actual use of various reading strategies, and help students establish a correct view of reading. Teachers guide students to adopt appropriate reading strategies to acquire target information, and students improve their analytical and problem-solving abilities through the learning of reading strategies. Skimming, checking reading, guessing vocabulary, etc. are all very operational and practical reading skills that can effectively help students complete English reading quickly and efficiently.

3.3.3. Choosing appropriate teaching methods for different topics. The focus of English learning should gradually change from learning specific knowledge such as English words and grammar to strengthening the ability to use English in various situations. The new English proficiency assessment standards have made it necessary for teachers to be more focused when teaching reading. The assessment of English reading comprehension is analyzed in depth in five aspects: comprehension of written instructions, descriptions, narratives, directions, and interactions, so teachers need to pay attention to cultivating students' reading comprehension of different types of reading texts in their teaching. When teaching comprehension of written instructions, teachers can adopt the reading teaching strategy of contextual experience, so that students can understand the use of various kinds of indicative words or signs in life and their specific meanings in specific scenes through more realistic contextual experience.

3.3.4. Increased awareness of teachers' professional development. The professionalism of English teachers will indirectly affect students' English reading comprehension. Therefore, in order to keep adapting to the developmental needs of students' abilities, teachers must develop a good habit of lifelong learning and continuously improve their own professionalism. Teachers should not only pay attention to the latest research results and new standards released in some professional fields, but also pay attention to absorbing new concepts and ideas when honing their teaching skills, so that they can constantly update their teaching ideas and optimize their teaching skills on the basis of the original.

4. Conclusion

This study explores the influencing factors of college teachers' English reading comprehension teaching ability through a linear regression analysis model, so as to propose teaching improvement strategies in a targeted way.

1) When teachers' academic qualifications are junior college and below, their teaching effectiveness in English reading comprehension ability will be significantly different from that of teachers with high academic qualifications (bachelor's degree and above) ($P<0.05$). There is a significant difference ($P=0.0021<0.05$) between the teaching effectiveness of teachers when their age is below 30 years old and that of teachers between 31 and 40 years old. There is no significant difference between teacher trainers and non-teacher trainers in language comprehension and cultural comprehension ($P>0.05$); there is a statistically significant difference between their logical reasoning and information processing skills at the statistical level ($P<0.05$).

2) There is a significant difference between teachers' teaching competence and language comprehension, cultural comprehension, logical reasoning, and information processing (p all 0.000<0.05). In addition, all four independent variables passed the significance level test ($p < 0.05$), and their influence on teaching competence was in the following order: language comprehension ability (0.3076) > information processing ability (0.2867) > logical reasoning ability (0.2484) > cultural comprehension ability (0.1225).

3) Based on the results of linear regression analysis, this study suggests that colleges and universities should pay attention to teachers' teaching concepts and change their reading teaching concepts in English teaching. Appropriate teaching methods should be adopted for different teaching topics in order to adapt to the learning needs of different students and improve students' learning participation and English reading comprehension. At the same time, teachers are always reminded that they should enhance their awareness of professional development and adapt to the development needs of students' abilities.

Funding

Research on Professional Development Paths of Foreign Language Teachers in Higher Vocational Colleges under the Background of Artificial Intelligence, Foreign Language Teaching Reform and Research Project under the New Standard of Vocational Education in 2023 of the Advisory Committee of Foreign Language Teaching in Vocational Education, Ministry of Education, P.R. China (Project No. WYJZW-2023SD0016).

References

[1] Septiarini, T., & Rahmat, A. (2018). The Relationship Between Reading Habits and Reading Comprehension of English Department in UNTIRTA. Journal of English Language Studies, 3(2), 178-191.

[2] Sukarni, S. (2019). Reading attitude and its influence on students' reading comprehension. Edukasi: Jurnal Pendidikan dan Pengajaran, 6(1), 193-204.

[3] Davis, M. H., McPartland, J. M., Pryseski, C., & Kim, E. (2018). The effects of coaching on English teachers' reading instruction practices and adolescent students' reading comprehension. Literacy Research and Instruction, 57(3), 255-275.

[4] Dewi, R. S., Fahrurrozi, U. H., & Wahyudi, A. (2020). Reading interest and reading comprehension a correlational study in Syarif Hidayatullah State Islamic University. Jakarta, Talent Development &Excellence, 12(1), 241-250.

[5] Alghonaim, A. S. (2020). Impact of Related Activities on Reading Comprehension of EFL Students. English Language Teaching, 13(4), 15-27.

[6] Lustyantie, N., & Kasan, R. A. (2021). Improving reading comprehension in efl situation: A correlation analysis. Educational Sciences: Theory & Practice, 21(1), 131-139.

[7] Syafitri, N. (2019). THE CORRELATION BETWEEN STUDENTS'READING HABIT AND READING COMPREHENSION IN ENGLISH AS A FOREIGN LANGUAGE. English Education Journal, 91-98.

[8] Xu, H., & Durguno ğ lu, A. Y. (2020). Motivational factors underlying different levels of reading comprehension of English language learners. TESOL Journal, 11(1), e00448.

[9] Syafitri, N. (2018). The correlation between lecturers' teaching styles and students' reading habit towards reading comprehension. Parenting, 58, 25.

[10] Yuliawati, L. (2018). The effects of grammar and vocabulary mastery on students' reading comprehension. Wanastra: Jurnal Bahasa Dan Sastra, 10(1), 1-8.

[11] Makebo, T. H., Bachore, M. M., & Ayele, Z. A. (2022). Investigating the correlation between students' reading fluency and comprehension. Journal of Language Teaching and Research, 13(2), 229-242.

[12] Rosyidin, I., Nurrudin, N., & Dewanti, R. (2022). THE EFFECT OF PROBLEM-BASED LEARNING MODEL TOWARDS STUDENTS'COMPREHENSION OF THE ENGLISH READING TEXT. English Review: Journal of English Education, 10(2), 565-578.

[13] Koosha, M., Abdollahi, A., & Karimi, F. (2016). The relationship among EFL learners' self-esteem, autonomy, and reading comprehension. Theory and Practice in Language Studies, 6(1), 68.

[14] Aini, N., Apriliani, N. S., & Ningrum, A. S. B. (2023). The Effectiveness of Using Tells Strategy on Students' Reading Comprehension. English Education: Journal of English Teaching and Research, 8(2), 133-144.

[15] Zhao Qi. (2022). Modeling and Analysis Method of National Fitness Big Data for Basketball Projects Based on a Multivariate Statistical Model. Security and Communication Networks

[16] Hao Li,Xianchao Dai,Ligang Zhou,Qun Wu,Muhammet Deveci & Dragan Pamucar. (2024). A least-squares framework for developing interval type-2 fuzzy semantics. Applied Soft Computing(PA),112293-112293.

[17] Yaobin Zou,Qingqing Huang & Huikang Qi. (2024). Automatic threshold selection guided by maximizing Pearson correlation. Computers and Electrical Engineering(PC),109815-109815.