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

Research on Data Analysis and Teaching Strategy Optimization Methods of English Online Classroom Combined with Multiple Regression Modeling

Jie Zhang[1,] ✉

[1] *Xi'an Fanyi University, Xi'an, Shaanxi, 710105, China*

ABSTRACT

Along with the development of the times, online classroom teaching activities have been carried out in different degrees and frequencies in various schools, and the gradual advancement of education informatization has improved the software and hardware environment of online classroom and other forms of teaching. The study designed a 21-item questionnaire related to English online classroom learning and selected all the students who participated in English online classroom teaching in a school for the survey. After collecting the questionnaire data, factor analysis and multiple stepwise regression model were used to conduct multivariate statistical analysis on the English online classroom data. And on this basis, the teaching plan was adjusted according to the actual learning behaviors of the high, medium and low risk level students themselves respectively to achieve personalized teaching. The results show that students' satisfaction with the English online classroom is high, and that pre-course homework analysis, group learning, formative learning evaluation, students' independent learning ability and online learning resources are the key positive factors affecting the learning effect of the English online classroom, with the influence coefficients of 0.036, 0.055, 0.048, 0.044, and 0.062, respectively. At the same time, after the optimization of teaching strategies, the students' logged-in learning behavior, participation rate in interactive test questions and grades were significantly improved, proving the effectiveness of the strategy.

Keywords: factor analysis, multiple stepwise regression model, personalized teaching, English online classroom

1. Introduction

With the continuous development of information technology, the Internet is becoming more and more popular, from the office field into thousands of households. Whether it is people's daily life, or work and study, the presence of the Internet is indispensable [1-2]. In the traditional English teaching,

✉ Corresponding author.
E-mail address: kkship2@163.com (J. Zhang).
Received 25 February 2024; Revised 10 June 2024; Accepted 05 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-217

classroom teaching is the main, most of the English classroom teaching is simple “full of irrigation” type, the students' participation in the classroom is not high, and at the same time can not meet the new era of college students to explore new things [3-4]. At this time, a new way of learning English - online English learning classroom appeared, equipped with the new era of Internet technology, online learning college English classroom, break through the geographical and time constraints, anytime and anywhere can be English teaching [5-7].

At present, online education based on microelectronics, computers and communication networks has made rapid development. There are more and more online English teaching platforms, including Tutor ABC, 51Talk, CCtalk, English Fluent Speaking and so on [8-9]. In the spring 2020 epidemic, online English teaching played a huge role. However, such a new form of education also had some problems during its use in the epidemic, such as individual students skipping classes, limited interaction between teachers and students, and so on [10-11]. The new learning environment has led to dramatic changes in teaching and has put new demands on teachers' teaching. Therefore, under such circumstances, it is necessary to explore how teachers can create online learning environments for students and how to interact with students effectively [12-14].

In the English teaching system of most institutions, teaching is mainly carried out through traditional teaching forms such as classroom lectures, which only stays at the primary stage of consciousness cultivation [15]. Students can only achieve a superficial understanding of English knowledge through teachers' book explanations, which will directly cause students to lack a clear purpose in the learning process, which is not conducive to the cultivation of their learning interests, and then lead to the loss of learning motivation [16-17]. In addition, there is a lack of effective teaching methods and the use of targeted practical teaching resources in personality development and practical exercises [18]. Most of the contemporary college students have the spirit of exploration for new things, and the traditional teaching mode of single class lectures and after-class exercises counseling leads to students just passively accepting knowledge [19]. And online learning college English classroom as a new college English teaching methods, good use of contemporary science and technology new forces, better adapted to the new era of English teaching form, conducive to the cultivation of high-quality, practical talents, and more conducive to the institutions to enhance their own level of school management and governance [20-21].

The emergence of new things will bring new challenges, online learning college English classroom compared with traditional English teaching has many advantages, but because teachers communicate with students across the screen, so they can not grasp the dynamics of students in real time. This will result in a lack of interaction between teachers and students in the online English classroom, and the teaching effect is greatly reduced [22-23]. If we want to effectively cultivate students' ability to use language, it is definitely not enough to simply instill it in class. Practice tells us that designing teaching activities in various forms and with clear purposes based on students' life experiences and interests plays a pivotal role in cultivating students' ability to use language [24-25].

First of all, it is necessary to obtain the factors affecting online English teaching, which is very meaningful to understand the underlying logic of English online teaching. Literature [26] found that the factors affecting teachers' use of English online learning technology include teachers' use of online websites and teachers' perceived usefulness of online technology. Literature [27] quantitatively analyzed through structural equation modeling found that teachers' preferences for teaching English online stemmed from teachers' attitudes toward online teaching, perceived usefulness, and technological accessibility, with technological accessibility being the primary influencing factor. Literature [28] reveals the current situation of English online teaching, in which the lack of English online teaching platforms, the lack of student motivation and the preparation of teachers' online learning materials are the most common problems. Literature [29], based on empirical investigations, clarified the factors affecting the effectiveness of online English teaching, mainly the lack of teachers' online teaching skills, inadequate infrastructure, and teacher-student interaction in the teaching

process as well as detachment from practice. Literature [30] utilized a questionnaire survey method to explore the possibility of online English teaching on the Internet, and the study concluded that English online learning is an effective mode of teaching and can be used as a supplementary channel to face-to-face teaching.

Secondly, focusing on the ways and means of English online teaching, the design of teaching strategies as well as teaching evaluation are conducive to promoting the good development of English online teaching. Literature [31] uses questionnaires, interviews and observation methods to reveal that teaching strategies such as asynchrony can effectively assist teachers in solving the obstacles encountered in the process of English online teaching. Literature [32] empirically analyzed the current status of English online teaching practice in elementary school and found that the English online teaching model played a positive role in facilitating students' nonverbal responses and extended verbal responses, and developing their English interaction skills. Literature [33] conceptualized an evaluation model for assessing the effectiveness of English online education, and confirmed the feasibility of the proposed evaluation model after simulation and simulation tests. Literature [34] simulated and evaluated the online teaching process of Thai language teachers to discover the advantages and disadvantages of online teaching in language course networks, which provides an important reference for the development of online language teaching. Literature [35] reveals that learners' online English learning prefers self-study and recreational teaching modes, and reflects on how online informal English learning can be integrated into confirming English classroom teaching practices within the constraints of the curriculum.

The study elicits the specific process of factor analysis and multiple stepwise regression analysis model for English online classroom on the basis of multivariate statistical analysis. Taking the students participating in English online classroom teaching in a certain school as an example, questionnaires were designed and distributed to collect data related to English online classroom teaching and students' satisfaction, and then the proposed factor analysis model was used to explore the key influencing factors affecting the effectiveness of English online classroom teaching, and the corresponding relationships were mapped by the multiple stepwise regression analysis model. Finally, following the principle of "teaching effect prediction-risk grading-differential adjustment-achievement improvement", we propose a personalized English teaching plan based on students' actual learning behavior, and we start from the three aspects of learning behavior, interaction, test participation rate and achievement. The effect of the program is examined in terms of learning behavior, interactive learning behavior, participation rate in test questions and performance.

2. Data analysis model of English online classroom based on multivariate statistical analysis

2.1. Foundations of multivariate statistical analysis

2.1.1. Concept of multivariate statistical analysis. Definition 1 (Random Vector) Let $X_1, X_2, ..., X_m$ be m random variables, and the vector $X = (X_1, X_2, ..., X_m)^T$ consisting of them is called a random vector.

Definition 2 (Distribution Function of a Random Vector) Let $X = (X_1, X_2, ..., X_m)^T$ be a random vector and R^m a R^m-dimensional real vector space, then its multivariate distribution function is:

$$F(x) = F(x_1, x_2, ..., x_m) = P(X_1 \le x_1, ..., X_m \le x_m) \tag{1}$$

Of these, $x = (x_1, x_2, ..., x_m) \in R^m$.

Definition 3 (Independence of Multiple Variables) Two random vectors X and Y are independent of each other if they satisfy:

$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y) \tag{2}$$

holds for everything x and y.

From the above definitions, if $F(x,y)$ is the joint distribution function of (X,Y), and $G(x)$ and $I(y)$ are the distribution functions of X and Y, respectively, then the following theorem follows:

Theorem 1 Random vectors X and Y are independent if and only if $F(x,y) = G(x)I(y)$ holds.

2.1.2. Numerical characterization of random vectors:

1) Mean of a random vector. Let the random vector $X = (X_1, X_2, ..., X_m)^T$ have m components and the mathematical expectation $E(X_i) = \mu_i (i = 1,2,...,m)$ of each component exists, then the mean value of the random vector X is:

$$E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ ... \\ E(X_m) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ ... \\ \mu_m \end{bmatrix} = \mu \tag{3}$$

Then μ is a m-dimensional vector. When U and V are constant matrices, the mean of the random vector $X = (X_1, X_2, ..., X_m)^T$ has the following properties:

$$E(UX) = UE(X) \tag{4}$$

$$E(UXV) = UE(X)V \tag{5}$$

2) Self-covariance matrix of random vector The self-covariance matrix of random vector $X = (X_1, X_2, ..., X_m)^T$ is given in the following equation:

$$\begin{aligned} Var(X) &= Cov(X,X) = E(X - EX)(X - EX)^T \\ &= \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & ... & Cov(X_1, X_m) \\ Cov(X_2, X_1) & Var(X_2) & ... & Cov(X_2, X_m) \\ ... & ... & ... & ... \\ Cov(X_m, X_1) & Cov(X_m, X_2) & ... & Var(X_m) \end{bmatrix} \end{aligned} \tag{6}$$

3) Covariance matrix and correlation matrix of random vectors. The $Y = (Y_1, Y_2, ..., Y_p)^T$ of random vectors $X = (X_1, X_2, ..., X_q)^T$ are two random vectors respectively, then the covariance matrix between X and Y is a $q \times p$ matrix with elements $Cov(X_i, Y_j)$, i.e.:

$$Cov(X,Y) = (Cov(X_i, Y_j))_{q \times p} \tag{7}$$

If the covariance matrix of a random vector $X = (X_1, X_2, ..., X_q)^T$ exists and the components are known to be greater than zero, then the correlation array of the vector is:

$$R = (r_{ij})_{q \times q} = \left(\frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)}\sqrt{Var(X_j)}}\right)_{q \times q} \tag{8}$$

When A and B are constant matrices, the covariance matrix of the random vector $X = (X_1, X_2, ..., X_q)^T$ has the following properties:

$$Var(UX) = UVar(X)U^T \tag{9}$$

$$Cov(UX,VY)=UCov(X,Y)V^T \tag{10}$$

2.2. Factor analysis model for English online classroom data

2.2.1. Factor analysis model. The factor analytic model decomposes the raw variables of the English online classroom into the form of the sum of two parts, where one part represents a linear combination of a few unobservable implied variables, and the other part is a special factor, which is not related to the common factor, but only to each of the raw variables themselves [36].

There is n sample, each with p observables. These p observations can be represented by the p components of a standardized random vector $X=(X_1,X_2,...,X_p)^T$. And let this random vector x have a mean vector $E(X)=0$, a covariance matrix $cov(X)=\Sigma$ and a correlation matrix equal to X. $F=(F_1,F_2,...,F_m)^T(m<p)$ denote m the common factor with mean vector $E(F)=0$ and covariance matrix $\mathrm{cov}(F)=I_n$, where I_n denotes the unit diagonal array. $\varepsilon=(\varepsilon_1,\varepsilon_2,...,\varepsilon_p)^T$ is called the special factor, which is related only to the components $X_i(i=1,2,...,p)$ of the random vector X and is independent of F and has $E(\varepsilon)=0$. Setting the components of ε to be independent of each other, its covariance matrix Σ_ε is shown below:

$$\mathrm{cov}(\varepsilon)=\Sigma_\varepsilon=\begin{bmatrix}\sigma_{11}^2 & & & 0\\ & \sigma_{22}^2 & & \\ & & \cdots & \\ & & & \sigma_{pp}^2\end{bmatrix} \tag{11}$$

Based on the above description, the following model is defined as a factor model:

$$\begin{cases} X_1=k_{11}F_1+k_{12}F_2+\cdots+k_{1m}F_m+\varepsilon_1\\ X_2=k_{21}F_1+k_{22}F_2+\cdots+k_{2m}F_m+\varepsilon_2\\ \cdots\cdots\\ X_p=k_{p1}F_1+k_{p2}F_2+\cdots+k_{pm}F_m+\varepsilon_p \end{cases} \tag{12}$$

where: $k=(k_{ij})_{p\times m}$ is the matrix of coefficients, often referred to as the factor loading matrix. Where the greater the absolute value of each factor loading k_{ij}, the greater the correlation between X_i and F_j. Typically, there is $|k_{ij}|\le 1$.

2.2.2. Statistical significance of factor loads. According to equation (12), the covariance between variables X_i and F_j is z:

$$cov(X_i,F_j)=cov(\sum_{j=1}^{m}k_{ij}F_j+\varepsilon_i,F_j) \tag{13}$$

Since $F_1,F_2,...,F_m$ is independent of each other and $F_1,F_2,...,F_m$ and $\varepsilon_1,\varepsilon_2,...,\varepsilon_p$ are also independent of each other, it follows:

$$cov(X_i,F_j)=k_{ij} \tag{14}$$

That is, factor loading k_{ij} indicates the degree of correlation between X_i and F_j.

2.2.3. Correlation coefficients of observed variables. According to equation (12), the correlation coefficient between variables X_i and X_j is:

$$r_{ij} = k_{i1}k_{j1} + k_{i2}k_{j2} + \ldots + k_{im}k_{jm} \tag{15}$$

Equation (15) shows that the correlation coefficient between X_i and X_j is also larger when both variables X_i and X_j have larger loadings on a common factor.

2.3. Multiple regression analysis model for English online classroom data

2.3.1. Multiple linear regression. Let the equation be a k-member linear regression equation and involve n independent variables, if the independent variable x has a series of n sets of observations with the dependent variable y as follows:

$$x_{1t}, x_{2t}, \ldots, x_{kt}, y_t \quad t = 1, 2, \ldots, n \quad n > k \tag{16}$$

Determining the k-element linear empirical regression equation can then be reduced to the problem of determining the coefficients $b_0, b_1, \ldots, b_k$ of the parametric regression equation based on the observations [37].

Let the estimate of the dependent variable y_t

$$\hat{y}_t = b_0 + \sum_{i=1}^{k} b_i x_{it} \tag{17}$$

where if $\hat{y}$ is expressed in terms of x_{k+1}, the equation can be found according to the principle of least squares:

$$\begin{bmatrix} s_{11}s_{12}\cdots s_{1k} \\ s_{21}s_{22}\cdots s_{2k} \\ \cdots\cdots\cdots\cdots \\ s_{k1}s_{k2}\cdots s_{kk} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_k \end{bmatrix} = \begin{Bmatrix} s1_{(k+1)} \\ s2_{(k+1)} \\ \cdots \\ s_{k(k+1)} \end{Bmatrix} \tag{18}$$

Equation (18) can be cumulatively abbreviated to $[s_i j]\{b_i\} = \{s_{i(k+1)}\}$.

If $[s_{ij}]$ is invertible, the above equation has a unique solution:

$$\{b_i\} = [s_{ij}]^{-1}\{s_{i(k+1)}\} \text{ and } b_0 = \bar{y} - \sum_{i=1}^{k} b_i \bar{x} \tag{19}$$

This is based on observation $x_{1t}, x_{2t}, \ldots, x_{kt}, y_t \; t = 1, 2, \ldots, n \; n > k$.

The regression equation can be derived:

$$\hat{y} = x_{k+1} = b_0 + \sum_{i=1}^{k} b_i x_i \tag{20}$$

2.3.2. Multiple Stepwise Regression Methods. Stepwise regression analysis is a regression equation that starts with an initial factor and introduces factors in order of significance of the relationship, from largest to smallest. When the introduced independent variable becomes insignificant for various reasons, it is removed from the regression equation. Factors are introduced and removed in stepwise regression by making F-tests to ensure that the factors included in the regression equation are significant until all the significant factors are included in the regression equation before the introduction of new factors [38].

Basic steps in stepwise regression analysis:

1) Calculate the correlation matrix. Stepwise regression analysis solves the system of regular equations by making a significance test and realizing the simple equation with the help of correlation matrix transformation as in equation (18), which consists of a k×k order matrix. In order to calculate, the (k + 1)-order augmentation matrix is obtained, such a matrix is also known as the correlation matrix. In order to improve the accuracy of the calculation results, the primary mean algorithm is no longer used and is replaced by the quadratic mean method, and the original matrix $\left[s_{ij}\right]$ is replaced by the normalized correlation matrix $\left[r_{ij}\right]$ as follows;

$$\begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1k} \\ s_{21} & s_{22} & \cdots & s_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ s_{k1} & s_{k2} & \cdots & s_{kk} \end{bmatrix} \rightarrow \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} & r_{1(k+1)} \\ r_{21} & r_{22} & \cdots & r_{2k} & r_{2(k+1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{k1} & r_{k2} & \cdots & r_{kk} & r_{k(k+1)} \\ r_{(k+1)1} & r_{(k+1)2} & \cdots & r_{(k+1)k} & r_{(k+1)(k+1)} \end{bmatrix} \tag{21}$$

The formula for Eq. r_{ij} is:

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} \tag{22}$$

This matrix is used in the stepwise regression analysis, where the factors introduced in steps 1 and 2 are in different sets of factors $\bar{G}^{(0)}$ and $\bar{G}^{(1)}$, respectively, and the factor with the most significant effect on y is selected to be introduced into the equation. $\bar{G}^{(0)}$ & $\bar{G}^{(1)}$ denote the set of all factors that are not in the regression equation, respectively. Starting from step 3, the first elimination and then introduction, elimination of insignificant factors in the equation, the introduction of factors outside the regression equation that play a significant role in y into the regression equation, of course, whether it is eliminated or introduced, the F-test is an indispensable part of the calculation of regression coefficients, correlation coefficients and residual standard deviation of the regression equation through the calculation of each step. The introduction and exclusion of factors is mainly a transformation of Eq. $\left[r_{ij}\right]$. In case of step n, $\left[r_{ij}^{(n-1)}\right]$ is transformed to $\left[r_{ij}^{(n)}\right]$. i.e:

$$R^{(n)} = \begin{bmatrix} r_{11}^{(n)} & r_{12}^{(n)} & \cdots & r_{1k}^{(n)} & r_{1(k+1)}^{(n)} \\ r_{21}^{(n)} & r_{22}^{(n)} & \cdots & r_{2k}^{(n)} & r_{2(k+1)}^{(n)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{k1}^{(n)} & r_{k2}^{(n)} & \cdots & r_{kk}^{(n)} & r_{k(k+1)}^{(n)} \\ r_{(k+1)1}^{(n)} & r_{(k+1)2}^{(n)} & \cdots & r_{(k+1)k}^{(n)} & r_{(k+1)(k+1)}^{(n)} \end{bmatrix} \tag{23}$$

When expressed in terms of r, the expression of the transformed form is:

$$D_n = \begin{cases} r_{k_1}^{(n)} = r_{k_n}^{(n)} - r_{k_n k_n}^{(n-1)} r_{k_n}^{(n-1)} / r_{k_n k_n}^{(n)} & (k \neq k_n, j \neq k_n) \\ r_{k_n}^{(n)} = -r_{k_n k_n}^{(n-1)} / r_{k_n k_n}^{(n-1)} & k \neq k_n \\ r_{k_n}^{(n)} = r_{k_n}^{(n-1)} / r_{k_n k_n}^{(n-1)} & j \neq k_n \\ r_{k_n k_n}^{(n)} = 1 / r_{k_n k_n}^{(n-1)} & k = k_n, j = k_n \end{cases} \tag{24}$$

2) Factor screening and elimination transformation. The formula for the eigenvalues used for the transformed introduced and eliminated factors is as follows:

(1) Partial regression sum of squares $Q_j^{(m)}$:

$$Q_j^m = \begin{cases} -V_j^m \sigma_{k+1}^2 & (j \in G^{(m)}) \\ V_j^m \sigma_{k+1}^2 & (j \in \bar{G}^{(m)}) \end{cases} \tag{25}$$

where the sum of squares of partial regressions of X_j at the mth step of $Q_j^{(m)}$ is:

$$V_j^m = r_{nj}^{(m)} r_{jn}^{(m)} / r_{jj}^{(m)}, (j \in G^{(m)}, j \in \bar{G}^{(m)}) \tag{26}$$

$r_{nj}^{(m)}, r_{jn}^{(m)}, r_{ij}^{(m)}$ is the element in $R^{(m)}$ at step m.

(2) The remaining sum of squares $Q^{(m)}$ of the regression equation at step m is:

$$Q^{(m)} = \sigma_n^2 r_{nn}^{(m)} \tag{27}$$

(3) Exclusion of factor $x_{km}^{'}$ from the test

In the factors of the m-step regression equation, the partial regression sum of squares $x_{km}^{'}$ (the smallest factor) is selected for elimination, i.e:

$$\left| V_{mk}^{(m)} \right| = \min_{j \in G^{(m)}} \left| V_j^{(m)} \right| \tag{28}$$

Then the F-statistic for $x_{km}^{'}$ is:

$$F_{2,k_m^{'}} = \frac{Q_{km}^{(m)} (N-m-1)}{Q^{(m)}} \text{ or } F_{2,k_m^{'}} = \frac{V_{km}^{(m)} (N-m-1)}{r_{nn}^{(m)}} \tag{29}$$

When $F_{2,k_m^{'}} \le F_2$, $x_{km}^{'}$ is dropped from the regression equation; when $F_{2,k_m^{'}} > F_2$, $x_{km}^{'}$ is dropped.

(4) Introduce the test of factor $x_{k_{m+1}}^{(m)}$

Selecting the factor with the most significant effect among the factors outside the regression equation, i.e., finding the factor $V_{k_{m+1}}^{(m)} = \max_{j \in \bar{G}^{(m)}} V_j^{(m)}$ with the largest partial regression sum of squares, the F-statistic for $x_{k_{m+1}}^{(m)}$ is:

$$F_{1,k_{m+1}} = \frac{V_{k_{m+1}}^{(m)} (N-m-2)}{(r_{nn}^{(m)} - V_{k_{m+1}}^{(m)})} \tag{30}$$

Factor $x_{k_{m+1}}^{(m)}$ is introduced when $F_{1,k_{m+1}} > F_1$, otherwise not.

3) Regression coefficients, correlation coefficients and residual standard deviation for step m. The regression equation for step m is:

$$\hat{x}_{k+1} = \hat{y} = b_0^{(m)} + b_{k_1}^{(m)} x_{k_1} + \cdots + b_{k_m}^{(m)} x_{k_m} \tag{31}$$

Style:

$$b_{k_j}^{(m)} = r_{k_j n}^{(m)} s_{nn} \Big/ s_{k_j} \ (j = 1, 2, \ldots, m) \tag{32}$$

$$b_0^{(m)} = \bar{x}_n - \sum_{j=1}^{m} b_{k_j}^{(m)} x_{k_j}^{(m)} \tag{33}$$

Correlation coefficient:

$$R_y^{(m)} = \sqrt{1 - r_{nn}^{(m)}} \tag{34}$$

Residual standard deviation:

$$S_y^{(m)} = S_{nn}\sqrt{r_{nn}^{(m)} / (N-m-1)} \tag{35}$$

3. Data analysis of the English online classroom

3.1. Data sources

On top of the theoretical research on the evaluation of effective teaching in the English online classroom, the first draft of the assessment items was formed by combining students' feedback on the teaching resources and the design of teaching activities in the English online classroom. Then we communicated with the instructors to categorize and interpret the items, and modified the items with ambiguities or inaccuracies in their presentation. Finally, 25 items containing better overall satisfaction of both sides of teaching (including reverse scoring items), 4 items counting students' satisfaction with English online classroom teaching, and the remaining 21 items for English online classroom related data were formed for factor analysis and multiple regression analysis. For the convenience of data processing, all the tests were conducted in a form that can be converted into a five-point Richter scale.

The test subjects were all the students who participated in the English online classroom teaching, from different engineering majors in the whole university, and four teaching classes were formed after freely selecting the courses through the university's teaching system, which can well represent the overall situation of the students. The test was conducted one week after the end of the course, and 320 valid questionnaires were returned, and all data analysis was done using SPSS software.

3.2. Multivariate statistical analysis of English online classroom data

3.2.1. Satisfaction Analysis of English Online Classroom. Statistics on students' learning satisfaction of English online classroom, Figure 1 shows the results of students' satisfaction evaluation, a~d represent students' satisfaction evaluation of English online classroom, students' logical analyzing ability, mastery of basic knowledge and methods, and students' experience of learning fun, respectively. The results of data analysis show that the mean value of students' overall satisfaction with the English online classroom is 3.41, and the mean values of students' logical analytical ability, mastery of basic knowledge and methods, and students' experience of learning fun are 3.61, 3.80, and 3.32, respectively. Positive self-evaluation of the overall satisfaction and learning gains are presented.

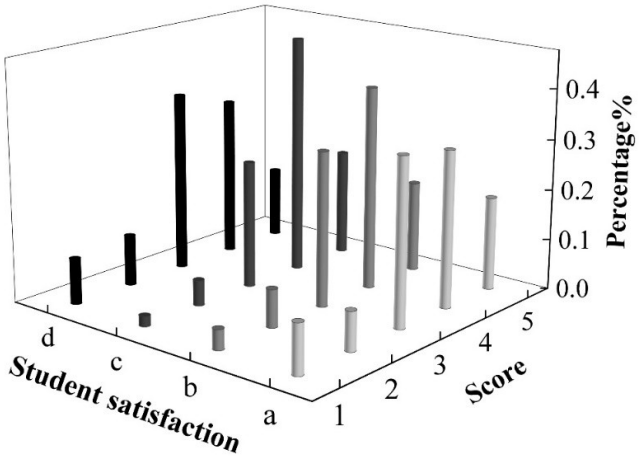


Fig. 1. Student's satisfaction evaluation results

3.2.2. Validity Factor Analysis of English Online Classroom Data. Based on the method of factor analysis, the influencing factors of the effectiveness of English online classroom teaching are investigated. Firstly, KMO and Bartlett's spherical test were performed on all the data of the 21

question items used for factor analysis. The results show that KMO=0.859, Bartlett's spherical test approximate chi-square is: 3030.115, significance is 0.000, indicating that the original variables are suitable for factor analysis. Based on the extraction method of principal component analysis, the standardized factor loadings of all question items ranged from 0.458 to 0.845, with a mean value of: 0.64 (SD=0.085). Next, the data were subjected to standardized orthogonal rotation, which converged after 6 iterations, and 5 common factors were extracted. The detailed report of the results of the public factors extracted after orthogonal rotation is shown in Table 1.The results of the public factors extracted after orthogonal rotation are all greater than 0.5, which can better represent the characteristics of the effectiveness of English online classroom teaching.

Table 1. Common factor results extracted after orthogonal rotation

Project	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
High frequency preschedule	0.739				
The high frequency rate is explained before class	0.699				
Requirements for efficiency of pre-class work	0.689				
The target of pre-class work is clear	0.642				
Teacher key	0.626				
Instant feedback	0.611				
Mind map	0.583				
Revised operation	0.508				
Panel questions		0.871			
Interest in group learning		0.844			
Team learning promotes communication ability		0.805			
Group presentation		0.595			
Panel discussion		0.537			
The input of the classroom presentation			0.722		
Review project diversification			0.627		
Increase normal performance			0.615		
Interest in class presentation			0.558		
Time of completion of homework				0.831	
Watch the time of video				0.785	
Example video					0.759
MOOC video					0.743

The results of factor analysis showed that the cumulative total variance explained by the five male factors was 62.378%. Next, the Cronbach′ s alpha coefficient reliability test was utilized to calibrate the internal consistency reliability of each of the public factors separately, and the results of the calibration of the factor analysis features and the reliability calibration of the scales are shown in Table 2. The reliability coefficients of the internal question items of each public factor are all above 0.61. The overall Cronbach′ s alpha reliability test coefficient for the 21 question items of all the public factors was 0.878, and thus the scale was designed with high structural reliability.

Through data-driven factor analysis, five factors affecting effective teaching in English online classrooms were summarized: pre-class homework analysis, group learning, formative learning assessment, students' self-directed learning ability and online learning resources.

Table 2. Feature verification result and the reliability verification result

Factor serial number	Factor eigenvalues	Variance of the interpretation(%)	Cumulative total variance of (%)	Cronbach α

1	7.112	33.625	33.625	0.851
2	2.078	9.789	43.414	0.873
3	1.625	7.711	51.125	0.728
4	1.333	6.425	57.55	0.631
5	1.021	4.828	62.378	0.612

3.2.3. Multiple Regression Analysis of English Online Classroom Data. Use x_1、x_2、⋯、x_5 to represent the independent variable, i.e., the influencing factors such as pre-class homework analysis, group learning, formative learning assessment, students' self-directed learning ability and online learning resources, and use y to denote the dependent variable, i.e., the learning effect of the English online classroom.

Variables are introduced one by one, and as each variable is introduced, the variables that have been selected for inclusion are tested one by one. When the originally introduced variable becomes no longer significant due to the introduction of later variables, it is eliminated. This process is repeated until significant independent variables are selected into the regression equation and non-significant independent variables are removed from the regression equation.

Step1. Test the suitability of the data for multiple linear regression analysis. Multiple nonlinear regression analysis was used because of the large linear fitting error.

Step2. R denotes the goodness of fit, which is a measure of how well the estimated model fits the observations. The closer the R value is to 1, the better the model is, and generally the adjusted R^2 is more accurate. F-value is the significance test of the regression equation, which indicates whether or not the non-linear relationship between the dependent quantity and all the independent variables in the model is significant in the aggregate to make an inference. sig. is the probability value of the F-test, which is also known as the p-value. df is the degree of freedom of the model. Table 3 shows the fitting results of the multiple regression model with an adjusted R^2 of 0.816, which indicates that the independent variables can explain 81.6% of the variation in the dependent variable, indicating a good fit.

Table 3. The fitting result of the multivariate regression model

Model	R	R^2	Adjust R^2	See.	Change the statistic				
					R^2	F	$df1$	$df2$	$Sig.F$
1	0.644	0.415	0.405	0.061	0.414	59.794	1	320	0.000
2	0.855	0.725	0.599	0.046	0.315	93.809	1	318	0.000
3	0.879	0.778	0.772	0.053	0.056	21.007	1	316	0.000
4	0.899	0.802	0.799	0.039	0.035	14.175	1	314	0.000
5	0.904	0.822	0.816	0.029	0.003	5.815	1	312	0.022

Step3. Multiple stepwise regressions were performed, and Table 4 shows the regression coefficients of the model. Where B is the regression coefficient, indicating the size of the effect of independent variable x on dependent variable y, SDE is the standard error, LL and SL are the lower and upper limits, respectively, t indicates the significance test value of the regression coefficients, and VIF is the variance inflation factor, which is used to determine the degree of multicollinearity of the model.

It can be seen that the tolerance of all the variables is not greater than 1 and none of the VIF exceeds 1.5, there is no multicollinearity in the variables.The Sig. of the t-tests are less than 0.05, which indicates that the independent variables have a significant effect on the dependent variable, and the final multivariate regression model can be derived:

$$y = 0.325 + 0.036x_1 + 0.055x_2 + 0.048x_3 + 0.044x_4 + 0.062x_5 \tag{36}$$

Among them, online learning resources have the greatest impact on the effectiveness of English online learning with a coefficient of 0.062.

Table 4. The regression coefficient of the model

Model	Variable	Non-standardized coefficients		Standard coefficient	*t*	*Sig.*	95%CI		Collinearity statistics	
		B	*SDE*				*LL*	*SL*	Tolerance	*VIF*
1	(Constant)	0.123	0.009		11.76	0.000	0.105	0.141		
	x_1	0.058	0.001	0.652	6.887	0.000	0.049	0.067	0.991	1.313
2	(Constant)	0.201	0.009		19.253	0.000	0.171	0.231		
	x_1	0.045	0.002	0.652	11.895	0.000	0.038	0.052	0.904	1.408
	x_2	0.062	0.002	0.553	-9.844	0.000	0.053	0.071	0.9	1.032
3	(Constant)	0.257	0.006		22.5	0.000	0.218	0.296		
	x_1	0.041	0.001	0.585	9.363	0.000	0.035	0.047	0.847	1.477
	x_2	0.059	0.007	0.472	-8.011	0.000	0.05	0.068	0.822	1.476
	x_3	0.055	0.014	0.259	-2.518	0.000	0.047	0.063	0.862	1.411
4	(Constant)	0.302	0.005		17.435	0.000	0.257	0.347		
	x_1	0.038	0.001	0.555	11.015	0.000	0.032	0.044	0.911	1.003
	x_2	0.057	0.004	0.476	-8.655	0.000	0.048	0.066	0.92	1.391
	x_3	0.052	0.013	0.269	-3.641	0.000	0.044	0.06	0.921	1.171
	x_4	0.049	0.006	0.185	2.671	0.000	0.042	0.056	0.972	1.377
5	(Constant)	0.325	0.032		10.258	0.000	0.276	0.374		
	x_1	0.036	0.004	0.599	11.667	0.000	0.031	0.041	0.996	1.231
	x_2	0.055	0.006	0.468	-8.047	0.000	0.047	0.063	0.88	1.22
	x_3	0.048	0.017	0.269	-4.815	0.000	0.041	0.055	0.98	1.14
	x_4	0.044	0.003	0.166	3.466	0.002	0.037	0.051	0.988	1.333
	x_5	0.062	0.002	0.115	-3.233	0.022	0.053	0.071	0.856	1.047

Step4. Residual test was performed on the regression model and Table 5 shows the results of ANOVA, the regression equation is useful since the Sig. values corresponding to the F values in the table are all 0, which is less than 0.05.

Table 5. Results of ANOVA

Model		Quadratic sum	*df*	Mean square	*F*	*Sig.*
1	Regression	0.255	1	0.255	62.223	0.000
	Residual	0.347	318	0.005		
	Total	0.602	319			
2	Regression	0.427	2	0.221	125.236	0.000
	Residual	0.175	317	0.003		
	Total	0.602	319			
3	Regression	0.465	3	0.158	98.559	0.000
	Residual	0.137	316	0.003		
	Total	0.602	319			
4	Regression	0.488	4	0.122	82.156	0.000
	Residual	0.114	315	0.001		
	Total	0.602	319			
5	Regression	0.497	5	0.098	77.428	0.000
	Residual	0.105	314	0.001		

	Total	0.602	319			

4. Optimization of teaching strategies in the English online classroom

4.1. Adjustment of teaching programs

On the basis of the effectiveness analysis of English online classroom data, the teaching program is adjusted according to the actual learning behaviors of students with high, medium, and low risk levels based on the principle of "teaching effect prediction-risk grading-differential adjustment-improvement of performance". The teaching program is adjusted according to the actual learning behavior of students with high, medium and low risk levels. The personalized teaching route is shown in Figure 2, with specific contents mainly including: pre-course preparation materials, classroom teaching activities, and post-course homework assignments. Similar predictions are carried out on a chapter-by-chapter basis to facilitate timely attention to students' learning needs and adjustment of teachers' teaching programs. The overall principle is: to take the learning needs of the majority group as the guide, and the learning characteristics of the minority group as the supplement, and ultimately to realize differentiated teaching that takes into account individuality, so as to achieve the purpose of comprehensively improving the learning effect of students.

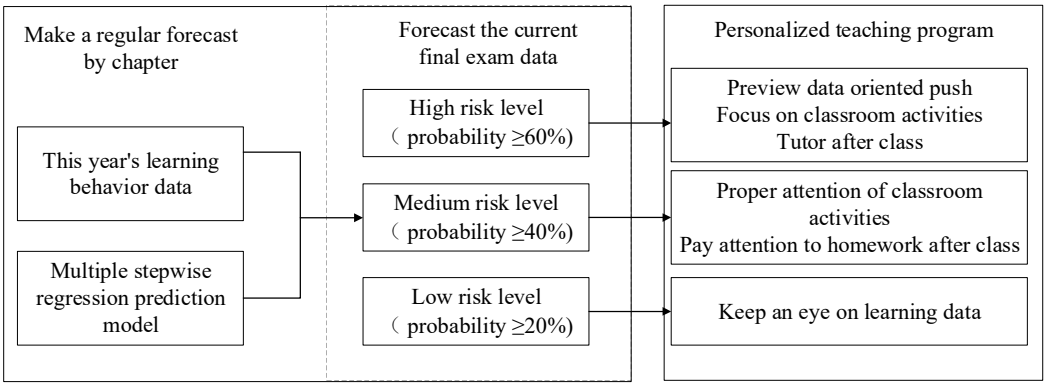


Fig. 2. Personalized teaching route

4.2. The optimization effect of English online classroom teaching

Two classes of 2022 in XX College with 27 undergraduate students each (S1~S54) were selected as experimental subjects. From August to November 2023, English online web courses were selected and tracked so as to obtain data on students' online learning behaviors, with Class 1 as the experimental class, which was taught using adapted teaching strategies, and Class 2 as the control class, which was taught using the traditional English online teaching scheme.

4.2.1. Login Learning Behavior Analysis. Students' behavior of logging on to the platform is an important indicator to ensure that English online learning reflects students' motivation to learn. In this empirical study, the online course started from August 29, 2023, to November 28, 2023, and excluding Saturdays and Sundays during this period, defining Monday to Friday as the study time when students logged into the platform, totaling 71 days, the total number of student logins was obtained from the platform, and the log-in rate was expressed by the ratio of the total log-in time to the total number of study days. In this case, the login rate of all students in class 1 and class 2 is shown in Figure 3.

It can be seen that the average login rate of class 1 is 1.15, 12 students (S3, S4, S11, S13, S14, S15, S18, S19, S20, S23, S26, and S27) have a higher login rate than the average login rate of the class they are in, and the other 15 students have a lower login rate than the average login rate of the class they are in. Among them, student S14 had a login rate of 3.12, which was significantly higher than the other

students in his class, and student S1 had the lowest login rate of 0.30.

Class 2 had an average login rate of 0.82, with 15 students (S28, S30, S31, S32, S33, S35, S36, S39, S40, S41, S43, S49, S50, S51, S54) having a higher login rate than the average login rate of their class, with student S40 having a significantly higher login rate of 1.82 than the rest of the students in his class, and the other 20 students having a lower login rate than the average login rate of their class.

An overall comparison of the login rates of students in the two classes shows that the average login rate in Class 1 is significantly higher than in Class 2, and that the students with the highest login rates in Class 1 are significantly higher than those with the highest login rates in Class 2. In Class 1, the percentage of students with higher than average login rate is higher than that in Class 2. 15 out of 27 students in Class 1 have a login rate higher than 1.0 (logging in to the platform at least once a day), accounting for 55.6% of the total number of students in the class, while 7 out of 27 students in Class 2 have a login rate higher than 1.0, accounting for 25.9% of the total number of students in the class. This shows that the overall online login rate of students in Class 1 is significantly higher than that of Class 2.

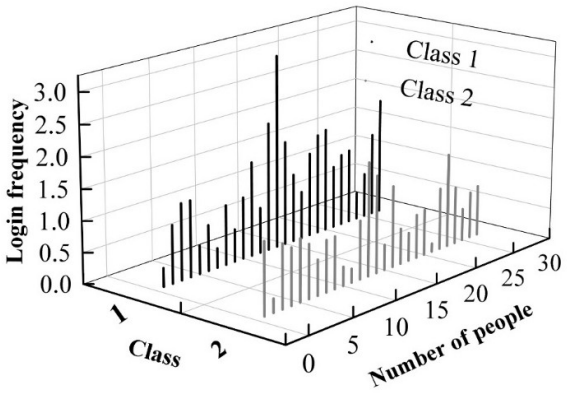


Fig. 3. The login of the student's online learning

4.2.2. Participation Rate Analysis of Interactive Test Questions. Interaction as a necessary means in the process of education and teaching, in which the interaction between learners and learning resources is a necessary process for students to master the content of knowledge. The participation rate of students in online interactive test questions is analyzed as a way to measure the completion of the learning program. The total number of times that the instructor organized the interaction between students in two classes was 18 times in class 1 and 12 times in class 2, and the ratio of the number of times that the students participated in the interaction to the total number of times that the students interacted with the test questions was analyzed as the participation rate. Among them, the participation rate of students in Class 1 and Class 2 in “interactive test questions” is shown in Figure 4.

The average participation rate of all students in Class 1 in the “interactive test” was 0.49, of which 17 students (S21, S2, S23, S7, S11, S18, S12, S14, S16, S20, S17, S4, S8, S24, S9, S26, S13) had higher participation rates than the average participation rate of their classes, and the other 10 students had a participation rate lower than the average participation rate of their class. Among them, student S13 had a participation rate of 0.66, which was significantly higher than the other students in his/her class. Students S9, S24, and S26 all had a participation rate of 0.58, which was also significantly higher than the rest of the students in their class, while student S1 had the lowest participation rate of 0.25.

The average participation rate of the students in class 2 in the “interactive test questions” is 0.25, and the participation rate of 9 students is higher than the average participation rate of the class, among which, the average participation rate of 9 students, S5, S7, S8, S9, S11, S13, S18, S22, S27, is 0.32, significantly higher than the average participation rate of the class, and the participation rate of S28 is

0.08, which is the lowest in the class. Higher than the average participation rate of the class in which they were enrolled, and student S28 had a participation rate of 0.08, which was the lowest participation rate in the class in which they were enrolled.

A comparative analysis of the overall situation in the two classes shows that in terms of the average class participation rate, the students in Class 1 had a significantly higher participation rate in the test questions than in Class 2, but the students with the highest participation rate in the test questions in Class 1 were significantly higher than those with the highest participation rate in Class 2. Of all the students in the two classes, 17 and 9 students, or 62.3% and 33.3% of the total class size, respectively, were above the average participation rate of their classes. Overall, students in class 1 were significantly better than class 2 in terms of participation rate in test questions.

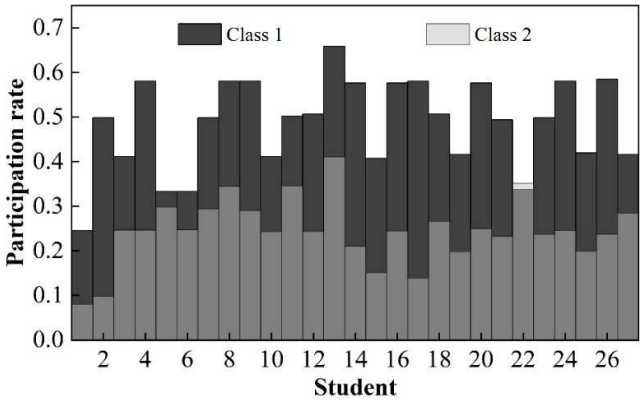
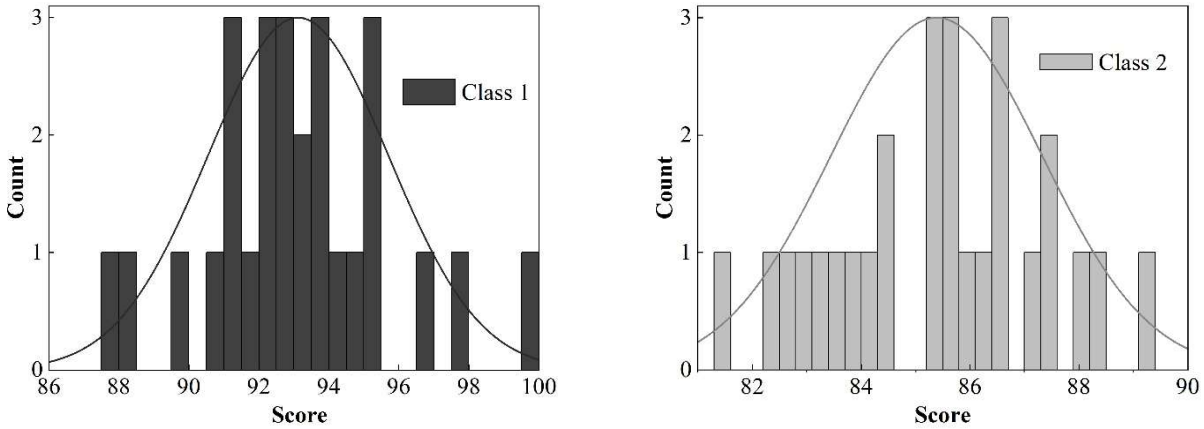


Fig. 4. Class 1 and class 2 students "interactive questions" participation rate

4.2.3. Analysis of English Learning Achievement. The English online classroom scores of the students in the two classes were analyzed, and Figure 5 shows the English scores of the two classes, (a) and (b) for Class 1 and Class 2, respectively. Compared to class 2, there is a significant improvement in the English teaching performance (92.52) of class 1, i.e., the English teaching performance is improved by 7.04 points through the implementation of the adapted English online classroom teaching program.



(a) Class 1 (b) Class 2

Fig. 5. English grades in both classes

5. Conclusion

The study selected all the students who participated in English online classroom teaching in a school,

and a total of 320 valid questionnaires were retrieved, the data were processed using multivariate statistical analysis, and all the data analysis was completed using SPSS software.

The results of the survey show that the mean values of students' overall satisfaction with the English online classroom, logical analysis ability, mastery of basic knowledge and methods, and students' experience of learning enjoyment are 3.41, 3.61, 3.80, and 3.32, respectively, and all of them have achieved a more positive evaluation.

In addition, the results of multivariate statistics showed that:

1) Through factor analysis, pre-class homework analysis, group learning, formative learning assessment, students' independent learning ability and online learning resources are the key factors affecting effective teaching in English online classrooms.

2) Through multiple stepwise regression analysis, the regression equation between the effectiveness of English online classroom and the influencing factors was derived: $y = 0.325 + 0.036x_1 + 0.055x_2 + 0.048x_3 + 0.044x_4 + 0.062x_5$, where x_1、x_2、⋯、x_5 represents the independent variables, i.e., the influencing factors such as pre-course homework analysis, group learning, etc., and online learning resources have the greatest influence on the effectiveness of English online learning (0.062).

Finally, on the basis of the effectiveness analysis of the English online classroom data, the teaching program was adjusted according to the actual learning behaviors of students with high, medium and low risk levels according to the principle of “teaching effect prediction-risk grading-differential adjustment-improvement of performance”, and the feasibility of the optimized teaching strategies was effectively verified according to the experimental results. We adjusted the teaching plan according to the actual learning behaviors of high-, medium- and low-risk students, and verified the feasibility of the optimized teaching strategy according to the experimental results.

References

[1] Fitria, T. N. (2021). Lecturer’s pedagogic competence: Teaching English in online learning during pandemic Covid-19. Journal of English Education, 6(2), 100-108.

[2] Ersin, P., Atay, D., & Mede, E. (2020). Boosting preservice teachers’ competence and online teaching readiness through e-practicum during the COVID-19 outbreak. International Journal of TESOL Studies, 2(2), 112-124.

[3] Sauro, S. (2017). Online fan practices and CALL. Calico journal, 34(2), 131-146.

[4] Wang, J., Zhang, X., & Zhang, L. J. (2022). Effects of teacher engagement on students’ achievement in an online English as a foreign language classroom: The mediating role of autonomous motivation and positive emotions. Frontiers in Psychology, 13, 950652.

[5] Alolaywi, Y. (2021). Teaching online during the COVID-19 pandemic: Teachers' perspectives. Journal of Language and Linguistic Studies, 17(4), 2022-2045.

[6] Krishnan, S. D., Norman, H., & Md Yunus, M. (2021). Online gamified learning to enhance teachers’ competencies using classcraft. Sustainability, 13(19), 10817.

[7] Habıbı, A., Mukmının, A., Rıyanto, Y., Prasojo, L. D., Sulıstıyo, U., Sofwan, M., & Saudagar, F. (2018). Building an online community: Student teachers’ perceptions on the advantages of using social networking services in a teacher education program. Turkish Online Journal of Distance Education, 19(1), 46-61.

[8] Gabriel, B. (2020). Challenges in using ICT Gadgets to cope with effects of COVID-19 on Education: A short survey of online teaching Literature in English. Journal of Humanities and Education Development, 2(4), 299-307.

[9] Han, Y. (2021). Evaluation of English online teaching based on remote supervision algorithms and deep learning. Journal of Intelligent & Fuzzy Systems, 40(4), 7097-7108.

[10] Songbatumis, A. M. (2017). Challenges in teaching English faced by English teachers at MTsN Taliwang, Indonesia. Journal of foreign language teaching and learning, 2(2), 54-67.

[11] Pawlak, M., Derakhshan, A., Mehdizadeh, M., & Kruk, M. (2025). Boredom in online English language classes: Mediating variables and coping strategies. Language Teaching Research, 29(2), 509-534.

[12] Syafiq, A. N., Rahmawati, A., Anwari, A., & Oktaviana, T. (2021). Increasing speaking skill through YouTube video as English learning material during online learning in pandemic covid-19. Elsya: Journal of English Language Studies, 3(1), 50-55.

[13] Atmojo, A. E. P., & Nugroho, A. (2020). EFL classes must go online! Teaching activities and challenges during COVID-19 pandemic in Indonesia. Register Journal, 13(1), 49-76.

[14] Moorhouse, B. L., Li, Y., & Walsh, S. (2023). E-classroom interactional competencies: Mediating and assisting language learning during synchronous online lessons. Relc Journal, 54(1), 114-128.

[15] Martin, F., Sun, T., & Westine, C. D. (2020). A systematic review of research on online teaching and learning from 2009 to 2018. Computers & education, 159, 104009.

[16] Sofi-Karim, M., Bali, A. O., & Rached, K. (2023). Online education via media platforms and applications as an innovative teaching method. Education and Information Technologies, 28(1), 507-523.

[17] Al Abiky, W. B. (2021). Lessons learned for teacher education: Challenges of teaching online classes during COVID-19, what can pre-service teachers tell us. Revista Argentina de Clínica Psicológica, 30(2), 110-118.

[18] Khatoony, S., & Nezhadmehr, M. (2020). EFL teachers' challenges in integration of technology for online classrooms during Coronavirus (COVID-19) pandemic in Iran. AJELP: Asian Journal of English Language and Pedagogy, 8(2), 89-104.

[19] Kusuma, I. P. I. (2022). EFL teachers' online teaching in rural schools during the COVID-19 pandemic: Stories from Indonesia. Studies in english language and education, 9(1), 203-221.

[20] Derakhshan, A., Kruk, M., Mehdizadeh, M., & Pawlak, M. (2021). Boredom in online classes in the Iranian EFL context: Sources and solutions. System, 101, 102556.

[21] Khoshsima, H., Saed, A., & Arbabi, M. A. (2018). Online teachers' attitudes toward using technology in teaching English as a foreign language. Journal of Applied Linguistics and Language Research, 5(2), 134-148.

[22] Sukmawati, S., Sujarwo, S., Soepriadi, D. N., & Amaliah, N. (2022). Online English Language Teaching in the Midst of Covid-19 Pandemic: Non EFL Students' Feedback and Response. Al-Ta lim Journal, 29(1), 62-69.

[23] Mumford, S., & Dikilitaş, K. (2020). Pre-service language teachers reflection development through online interaction in a hybrid learning course. Computers & Education, 144, 103706.

[24] Sepulveda-Escobar, P., & Morrison, A. (2020). Online teaching placement during the COVID-19 pandemic in Chile: challenges and opportunities. European Journal of Teacher Education, 43(4), 587-607.

[25] Rinekso, A. B., & Muslim, A. B. (2020). Synchronous online discussion: teaching English in higher education amidst the covid-19 pandemic. JEES (Journal of English Educators Society), 5(2), 155-162.

[26] Ayu, M., & Pratiwi, Z. F. (2021). The implementation of online learning in English language teaching during pandemic: The teachers voice. Journal of Research on Language Education, 2(2), 93-99.

[27] Huang, F., Teo, T., & Guo, J. (2021). Understanding English teachers' non-volitional use of online teaching: A Chinese study. System, 101, 102574.

[28] Nugroho, A., Ilmiani, D., & Rekha, A. (2020). EFL teachers' challenges and insights of online teaching amidst global pandemic. Metathesis: Journal of English Language, Literature, and Teaching, 4(3), 277-291.

[29] Subekti, A. S. (2020). Covid-19-triggered online learning implementation: Pre-service English teachers' beliefs. Metathesis: Journal of English Language, Literature, and Teaching, 4(3), 232-248.

[30] Rifiyanti, H. (2020). Learners' perceptions of online English learning during COVID-19 pandemic. Scope: Journal of english language teaching, 5(1), 31-35.

[31] Putri, N. R., & Sari, F. M. (2021). Investigating English teaching strategies to reduce online teaching obstacles in the secondary school. Journal of English Language Teaching and Learning, 2(1), 23-31.

[32] Cheung, A. (2021). Synchronous online teaching, a blessing or a curse? Insights from EFL primary students' interaction during online English lessons. System, 100, 102566.

[33] Yu, H., & Li, X. (2021). An evaluation model of English teaching effectiveness based on online education. International Journal of Continuing Engineering Education and Life Long Learning, 31(2), 218-233.

[34] Todd, R. W. (2020). Teachers' perceptions of the shift from the classroom to online teaching. International Journal of TESOL Studies, 2(2), 4-16.

[35] Lamb, M., & Arisandy, F. E. (2020). The impact of online use of English on motivation to learn. Computer Assisted Language Learning, 33(1-2), 85-108.

[36] Ondrej Zavila. (2025). Human factors analysis and classification system - positive experience (HFACS-PE): New approaches to aviation accident and incident investigation. Journal of Loss Prevention in the Process Industries105578-105578.

[37] Jihao You, Dan Tulpan, Cheryl Krziyzek & Jennifer L Ellis.(2025).Prediction of Pellet Durability Index (PDI) in a Commercial Feed Mill Using Multiple Linear Regression with Variable Selection and Dimensionality Reduction. Journal of animal science.

[38] Patricia Gilholm, Paula Lister, Adam Irwin, Amanda Harley, Sainath Raman, Luregn J Schlapbach & Kristen S Gibbons.(2025).Comparison of Random Forest and Stepwise Regression for Variable Selection Using Low Prevalence Predictors: A case Study in Paediatric Sepsis. Maternal and child health journal(prepublish),1-10.