International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 2906-2915

https://doi.org/10.70517/ijhsa463243

A Study on the Application of Multiple Regression Analysis in the Analysis of Learning Data and Optimization of Teaching Strategies in English Online Classrooms

Jie Zhang[1,*]
[1] Xi'an Fanyi University, Xi'an, Shaanxi, 710105, China
Corresponding authors: (e-mail: kkship2@163.com).

Abstract This paper uses multiple regression analysis to construct a prediction model of online learners' academic performance. The English online classroom learning data of 448 college students in college B are taken as the object of analysis. From them, four indicators, namely, the number of study times, daily study time, daily study frequency, and task point completion, were selected as independent variables, and learning achievement was taken as the dependent variable. The regression coefficients of each variable were determined by stepwise multiple regression, and the final regression model of academic performance was determined as $Y=-8.632+0.99x_1+0.231x_2+0.485x_3+0.286x_4$. The results of the constructed prediction model of online learners' academic performance are basically accurate through the discrimination of sample independence, residual normality test, and the discrimination of the absence of multiple covariance in the independent variables, respectively. Teachers can formulate teaching plans and carry out personalized tutoring according to the English online classroom learning data, and verify the learning effect by comparing the before and after data.

Index Terms Multiple regression analysis, residual normality test, multicollinearity, online classroom learning data

I. Introduction

At present, new technologies represented by artificial intelligence, big data, cloud computing, blockchain, Internet of Things, 5G technology, etc., drive human society to rapidly enter the intelligent era based on digital applications. In this context, digitalization has become an important force to generate structural changes in education, and how to empower education and teaching with digitalization has become a key task in the current education field. In the age of informationization, digital technology provides new possibilities and opportunities for English teaching in colleges and universities, so that English teaching is no longer restricted by the traditional classroom teaching mode [1]. The traditional way of teaching English has certain limitations, such as limited teaching resources, single teaching method, etc., while the application of digital technology in English teaching in colleges and universities can effectively solve these problems and improve the quality and efficiency of English teaching [2]-[4].

With the rapid development of information technology and the continuous promotion of education informatization, online education platforms are increasingly becoming an important part of college English teaching [5]. Online education platform is characterized by rich learning resources, flexible learning mode and diverse teaching forms, which provides new opportunities and possibilities for the reform of university English teaching [6], [7]. The development and application of online education platforms help to realize the sharing of high-quality educational resources and personalized learning, which in turn significantly improves student participation and learning achievement in the classroom [8], [9]. However, in the localized application and practical exploration of online platforms in English teaching, there are still problems such as poor integration of resources and insufficient teacher-student interaction [10]. It is necessary to further use digital technology to empower online English education, better realize personalized teaching through data analysis, intelligent recommendation, etc., so as to optimize the platform function and teaching design, deepen the integration of online education and traditional classroom, give full play to the advantages of the online platform, and enhance the teaching effect [11]-[13].

The study analyzed the correlation between learning behavior and learning achievement based on the data of English course learning logs from the online learning platform of university B in city A. In this regard, regression and classification prediction methods are used to predict the academic performance of online learners. Firstly, multiple linear regression methods are used to predict learning performance and determine the regression model of learning performance. Then, the accuracy of regression prediction was tested, and three methods, namely, the discrimination of sample independence, the discrimination of residual normality and the discrimination of multiple covariance, were

selected to evaluate the accuracy of the prediction results. Finally, the optimization method of teaching strategies is proposed according to the analysis results.

II. Experimental data and methodology

II. A.Experimental data

II. A. 1) Data sources

In this study, 448 (112 students in each class) college students from four classes of a major in the College of Business of the class of 2023 in University B of City A were used as data collection objects, among which two students' records had missing information, and the data containing missing information were excluded, and 428 valid data were finally retained. The subject of study was a university English course, and the learning scenario data came from the online course learning log data of the Superstar Learning Platform. The learning performance data comes from the number of check-ins, online test scores and offline written test scores of the Super Star Learning Platform.

II. A. 2) Determination of factors influencing performance

The online learning platform contains 10 attributes such as number of check-ins, resource rationing, and online test scores [14]. In this study, it is proposed to assess and rank the impact of all the individual data attributes on final grades from the raw data and select the set of sub-attributes with the help of the ranking results. The method is to calculate the Pearson's correlation coefficient between all the individual attributes and the final grade categories and rank them according to the magnitude of their correlation coefficients, where the higher the value of the coefficient, the stronger the correlation with the final grade. Some of the attributes contained missing values, irrelevant data, and isolated points, etc. Data cleaning was performed on these attributes, and seven attributes were finally retained, containing the number of times of study (0.5965), daily study time (0.5761), daily study frequency (0.5663), task point completion (0.5132), online test scores (0.3961), resource rationing (0.3162), and course length setting (0.2639). The first 4 items with large coefficients were judged to be the main factors affecting final grades, so they were used as independent variables in the predictive modeling of the online learning platform.

The offline grade mainly contains 2 aspects: usual grade and paper grade. The usual grade consists of 5 attributes, such as quizzes, homework, and experiments. The source of data is the real grades of students' daily quiz assessment. The calculation method of the usual grade is shown in Equation (1):

$$y_1 = \alpha * 0.3 + \beta * 0.4 + \gamma_1 * 0.1 + \gamma_2 * 0.1 + \delta * 0.1 \tag{1}$$

where γ_i represents the usual grade, α represents the sectional exam, β represents the experiment, γ_1 represents the classroom quizl, γ_2 represents the classroom quiz, and δ represents the homework.

The final grade is mainly derived from the usual grade and the paper grade, which is calculated as shown in Equation (2):

$$\gamma_2 = \alpha * 0.6 + \beta * 0.4 \tag{2}$$

Here γ_2 represents the final grade, α represents the paper grade, and β represents the usual grade.

Two parts of grades, online and offline, with 6 attributes and 3426 data were analyzed and studied. While the values are concentrated, in order to avoid the interference of different data attributes on the regression prediction, all the data are subjected to data normalization so that all the data ranges are in the interval of [0, 1].

II. B.Experimental Methods

For the study of a regression model with one dependent variable and two or more independent variables, it is multiple regression. According to the research content of this paper, multiple regression analysis is applied.

Considering the convenience of model calculation and other factors, in practice, the linear model is usually preferred for fitting, and when encountering a nonlinear model, a certain method can also be used to convert it to a linear model. Therefore, in this paper, when applying multiple regression analysis, the linear model is selected for fitting [15]. The application of multiple linear regression model mainly includes the following steps.

II. B. 1) Constructing regression equations

Let y be an observable random variable that receives p non-random factors $x_1, x_2, ..., x_p$ and a random factor ε, and y is linearly related to $x_1, x_2, \cdots, x_p$ with a multiple regression linear equation of the form:

$$y = \beta_0 + \beta_l x_l + ... + \beta_p x_p + \varepsilon \tag{3}$$

where y is the explanatory variable (dependent variable). $x_i(i=l,2,\cdots;p)$ is the explanatory variable (independent variable). $\beta_i(i=0,1,2,\cdots;p)$ is the regression coefficient, reflecting a measure of the linear effect of the i th independent variable x_i on the dependent variable y. The ε represents the error between the regression value and the measured value, usually assumed to be $\varepsilon \sim N(0,\sigma^2)$.

II. B. 2) Applying Least Squares to Estimate Unknown Parameters

With n independent observations on y and $x_i(i=I,2,\cdots;p)$, we obtain n sets of sample data $(x_{i1},x_{i2},\cdots;x_{ip};y_i)\ (i=I,2,\cdots,n)$, we have:

$$\begin{cases} y_l=\beta_0+\beta_1x_{11}+\beta_2x_{l2}+\ldots+\beta_px_{lp}+\varepsilon_l \\ \vdots \\ y_2=\beta_0+\beta_1x_{21}+\beta_2x_{22}+\ldots+\beta_px_{2p}+\varepsilon_2 \\ y_n=\beta_0+\beta_1x_{n1}+\beta_2x_{n2}+\ldots+\beta_px_{np}+\varepsilon_n \end{cases} \tag{4}$$

where $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ are independent of each other and all obey $N(0,\sigma^2)$, then Eq. (4) can be expressed in matrix form:

$$Y=X\beta+\varepsilon \tag{5}$$

$$X=\begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ & & \vdots & & \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix} \tag{6}$$

where $Y=(y_1,y_2,\ldots,y_n)^T$, $\beta=(\varepsilon_0,\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_p)^T$, $\varepsilon=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)^T$, $\varepsilon \sim N_n(O,\sigma^2I_n)$, and I_n is a unit matrix of order n.

Applying the least squares method, the least squares estimates of $b_1,b_2,\ldots,b_p$ are computed to obtain $\beta_0,\beta_1,\beta_2,\ldots,\beta_p$, respectively, and then the multiple linear regression equation (approximate function) is:

$$y=b_0+b_1x_1+\ldots+b_px_p \tag{7}$$

II. B. 3) Hypothesis testing

Multiple linear regression model is only a hypothesis, in the actual problem, to determine whether the model has a good fit with the actual data, how significant the linear relationship of the model, etc., but also need to be tested by combing statistics to decide whether the model is scientifically applicable [16]. The commonly used test methods are as follows:

(1) Regression equation goodness-of-fit test (R test)

The goodness of fit reflects the extent to which the model's estimate $\hat{y}$ can explain the variation in the dependent variable sample γ, which is judged by the decidable coefficient R^2:

$$R=1-\frac{\sum_1^n(Y_i-\hat{Y}_i)^2}{\sum_1^n(Y_i-\bar{Y}_i)^2} \tag{8}$$

The value of R^2 ranges from $0<R^2\le 1$, and the larger R^2 is closer to 1, which indicates that the model fits better.

(2) Significance test of regression equation (F test)

The test of significance of the regression equation aims to make a judgment on whether the linear relationship between y and $x_1,x_2,\cdots,x_p$ in the model holds significantly in the aggregate.

The original hypothesis is $H_0:\beta_1=\beta_2=\cdots=\beta_p=0$ The alternative hypothesis is that $H_1:\beta_j$ is not all 0 If the original hypothesis holds, then there is no significant linear relationship between the dependent and independent variables in the model. , construct the statistic F for the test:

$$F=\frac{\sum_{1}^{n}(y_i-\overline{y}_i)^2/p}{\sum_{1}^{n}(y_i-\hat{y}_i)^2/(n-p-1)}\sim F(p,n-p-1) \tag{9}$$

Given the significance level α , look up the table to get $F_a(p,n-p-l)$, then the test rule is:

If $F\le F_a(p,n-p-l)$, the original hypothesis is accepted and the regression model cannot be used for prediction.

If $F>F_a(p,n-p-l)$, the original hypothesis is rejected, indicating that the regression model is significant and can be used for prediction analysis.

(3) Significance test of regression coefficients (t test)

For the multiple linear regression model, the overall regression equation linear relationship is significant, does not mean that the effect of each independent variable on the dependent variable is significant. Therefore, it is necessary to test the significance of each independent variable to determine whether it is retained in the model or not, and those independent variables that do not have a significant effect on the dependent variable should be excluded from the model. This test can be accomplished through the t test for the independent variables.

The original hypothesis is $H_0:\beta_1=\beta_2=\cdots=\beta_p=0$

Alternative hypothesis is $H_l:\beta_j\neq 0$

Construct the statistic t for the test:

$$t=\frac{\hat{\beta}_j}{\sqrt{c_{jj}\sum\frac{e_i^2}{n-p-1}}}\sim t(n-p-1) \tag{10}$$

Given the significance level α , check the table to get $t_{a/2}(n-p-l)$, then the test rule is:

If $|t|\le t_{a/2}(n-p-1)$, the original hypothesis is accepted, i.e., the effect of the corresponding independent variable is not significant and should be eliminated.

If $|t|>t_{a/2}(n-p-1)$, the original hypothesis is rejected, i.e., the effect of the corresponding independent variable is significant and can be retained.

(4) Other tests

In addition, for the hypothesis test of multiple linear regression equation there are multiple covariance test, residual test, etc., which can be selected according to the practical application needs.

II. B. 4) Determination of optimal multivariate equations

If any of the above tests fails, the method of “stepwise regression” should be used, i.e., eliminate non-significant independent variables and then re-fit the model to analyze the judgment, and repeat until the optimal multiple linear equations are obtained.

III. Analysis of online learning behavior studies

This chapter analyzes six attributes based on 3,426 pieces of data collected to explore the relationship between learning behaviors and academic performance.

III. A. Analysis of the number of studies and course length settings

The number of study times is an important measure of student learning status and an important reference for determining the number of hours and duration of a course. The data shows that the more intensive data on the number of study times in English courses began in November 2023 and peaked in December 2023, with the highest being class 1, where the number of study times reached 4,978 frequencies. In January 2024, the number of times a student takes a course is still informative, proving that in that month students are still in a critical period of course work. Starting from February 2024, the number of times students study begins to decrease, and in March and April 2024, a small number of students still click on the course to start studying, so there is still a certain number of clicks,

which confirms the importance of the online learning platform. For students, they can start learning at any time according to their needs, which is not affected by the teacher's time or the time of the course, as shown in Figure 1. The data show that the length of the English course is the best within 3 months, students are very interested in the knowledge of the stage, they will frequently log in the learning platform to start learning, the length of the course is too short, can not achieve the learning effect of the students, the length of the course is too long, the interest of the students is weakened, which is not conducive to the mastery of the course knowledge.

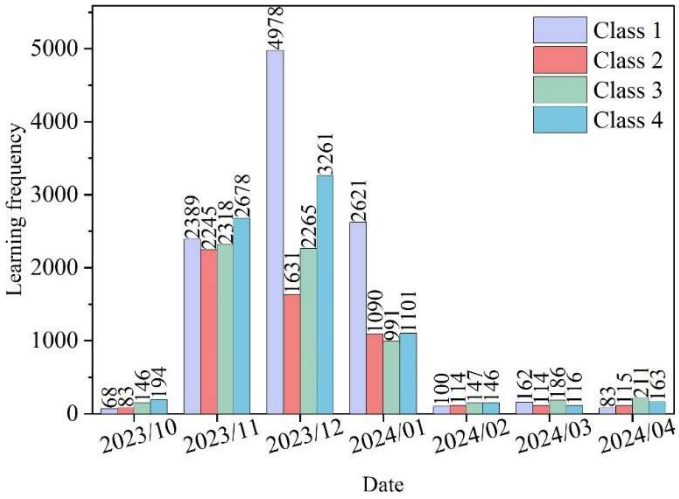

Figure 1: Learning frequency change statistics

III. B. Analysis of daily study time and frequency

By counting the learning clicks of the four classes at different times of the day, we can observe the students' learning habits, so that teachers can choose the appropriate time when posting notices, activities, resources and other contents, as shown in Figure 2. The peak of students' learning frequency is mainly concentrated between 8-12pm, 12-16pm and 16-20pm, 8-12pm and 12-16pm may have the influence of clicks during class, 16 -20 o'clock students have usually finished the day's lesson, and have more independent learning time. The frequency of learning in this time period is as high as 6,413 times, reflecting that students have a certain sense of independent learning and ability.

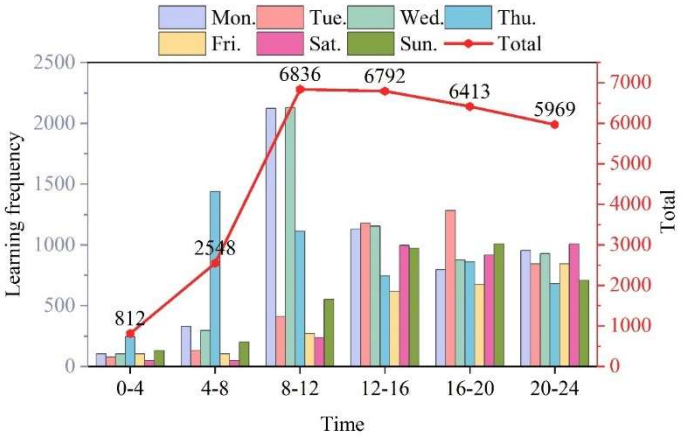

Figure 2: Daily learning frequency change statistics

III. C. Analysis of task point completion versus resource allocation

In the course construction, task points such as course introduction, literature reading, video watching, signing in, homework, etc. are added, and after students complete the task points, the Super Star Learning Platform will make records accordingly. The platform will record in detail the number of times students watch and the length of video watching. By counting the English course task points into three types of resources: documents, videos and chapter tests, analyzing the completion degree of the task points, further observing the students' learning frequency, and

summarizing the students' learning interests in order to select the resources that are acceptable to the students to provide them with, and ultimately achieve a better learning effect. The detailed analysis statistics are shown in Figure 3. Students in Classes 1, 2, 3, and 4 all met the standard or higher for task point completion, with task point completions of 2411, 3086, 2881, and 3142, respectively. And the degree of completion of video resources are higher than other resources, therefore, students are more inclined to learn from video resources, in the process of preparation and setting of resources, the proportion of video resources can be strengthened, and the document resources and test resources can exist as auxiliary resources.

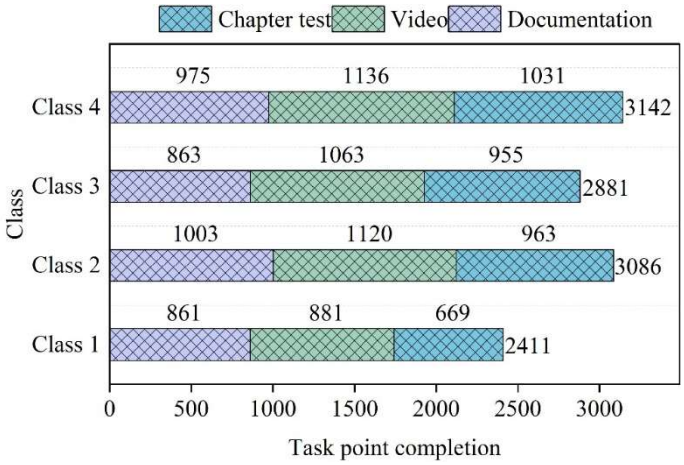

Figure 3: Complete analysis of the mission points

III. D. Analysis of task point completion and learning outcomes

In this paper, the number of students with different score bands, the average score, the highest score and the lowest score of students in Class 1, Class 2, Class 3 and Class 4 in the English final exam were counted respectively, and the specific distribution of the number of students is shown in Fig. 4. In the resource settings of the English course, a total of 80 task points such as video viewing, discussion, chapter quiz, homework, exam, activities, etc. were set up, and the Super Star Learning Platform made detailed records for every task point completed by the students, with different degrees of completion of the task points in each class. By counting the completion of the task points of the four classes and comparing the completion and performance of the classes in the final program assessment, it can be seen whether the task point setting is reasonable and whether it can achieve the preset effect. The task point completion rates of the four classes were 67.63%, 83.69%, 87.96%, and 86.33%, respectively, and the average grades of the four classes were 59.3, 71.2, 80.0, and 72.1, respectively.By comparing the distribution of task point completion rates with the distribution of students' final grades, it can be seen that there is a positive distribution between the task point completion rates and students' grades.

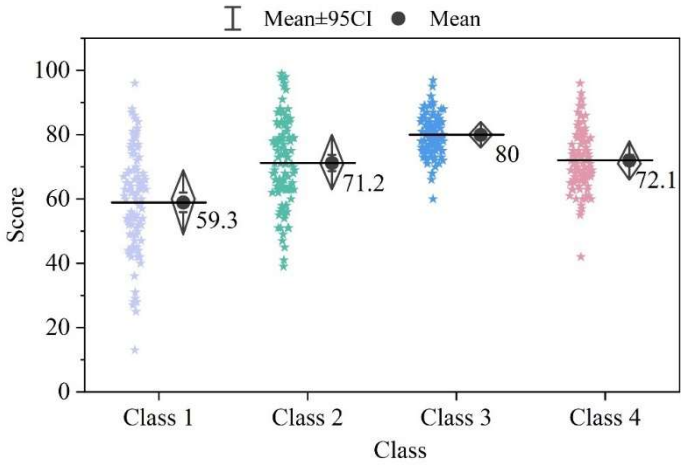

Figure 4: Achievement analysis

IV. Predictive analysis of learning achievement

According to the results of the statistical analysis of online learners' learning behaviors, there is a correlation between the number of study times, daily study time, daily study frequency, task point completion and academic performance, which indicates that the four types of learning behaviors can be used in the study of academic performance prediction. On this basis, these four learning behaviors are taken as independent variables, and learning achievement is taken as dependent variable, and multiple linear regression prediction method is used to predict learning achievement, which comprehensively takes into account the influence of each learning behavior on learning achievement to ensure the accuracy of the regression model of learning achievement prediction.

IV. A. Modeling multiple linear regression

A linear regression model was constructed to fit academic behavior and academic performance with the model equation:

$$\hat{y} = b_0 + b_1 x_1 \tag{11}$$

The least squares method was used to minimize the mean square deviation to calculate the fitted learning behavior and academic performance using the formula:

$$MSE = \frac{\sum (y_j - \hat{y}_j)}{n} \tag{12}$$

Prediction of academic performance can be achieved by fitting individual learning behaviors to academic performance, but individual learning behaviors cannot accurately predict academic performance, so multiple linear regression is used to comprehensively consider the joint influence of multiple learning behavior variables on academic performance, which makes the prediction model of academic performance more accurate, and its general expression is:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_m x_m \tag{13}$$

Meanwhile, there are preconditions that need to be met to perform multiple linear regression: sample independence, i.e., all samples do not interfere with each other. Residuals are normal, i.e. the residuals of the model obey normal distribution. No multicollinearity of independent variables, i.e., there is no extremely strong correlation between independent variables. The prediction model constructed for online learners' academic performance is accurate and reliable only when the above three conditions are met.

IV. B. Experimental procedure and result analysis

SPSS statistical analysis software was used to predict academic performance by multiple linear regression, and it was not possible to directly determine the independent variables of learning behavior that entered the multiple linear regression model in the regression prediction, thus the stepwise regression method was chosen to gradually exclude the independent variables that were not statistically significant with the prediction of academic performance. The independent variables of learning behavior affecting academic performance were identified through the stepwise regression method, and the regression coefficients of the respective variables were obtained as shown in Table 1. According to the table of multiple linear regression correlation coefficients, it can be seen that the significant P-values of the four independent variables are less than 0.05, indicating that these four learning behavior variables are the key variables affecting learners' academic performance, and the coefficients are all greater than zero, showing a positive effect on academic performance. It shows that the number of study times, daily study time, daily study frequency, and task point completion all have a large impact on academic performance, and the predictive model of academic performance can be determined based on the regression coefficients.

Table 1: Regression correlation coefficient

Model	Unnormalized coefficient		Normalization factor	T	Sig.	Common linear statistics	
	B	Standard error	Beta			Tolerance	VIF
Constants	-8.632	1.264	-	-7.512	0.000	-	-
Learning frequency	0.99	0.054	0.187	18.765	0.000	0.889	1.111
Daily learning time	0.231	0.052	0.054	5.085	0.000	0.789	1.246
Daily learning frequency	0.485	-0.008	0.395	31.775	0.000	0.658	1.529
Task point completion	0.286	0.016	0.281	21.373	0.000	0.592	1.698

The finalized regression model for academic performance is:

$$Y = -8.632 + 0.99x_1 + 0.231x_2 + 0.485x_3 + 0.286x_4 \quad (14)$$

x_1, x_2, x_3, x_4 denote the number of study counts, study time per day, study frequency per day, and the number of independent variables of task point completion, respectively.

Through the linear regression model diagnosis, it is verified whether the linear regression model satisfies the following three preconditions, and if it does, it indicates that the results of the regression model are basically accurate.

(1) Discrimination of sample independence

The Durbin-Watson value is used to discriminate the independence of the sample, when the value is near 2, it indicates that the sample is independent. The model summary is shown in Table 2. The Durbin-Watson value in this study is 1.766<2, there is a certain deviation, there is a slight non-independence, but the deviation is not very large, and it will not affect the accuracy of the regression results.R^2=0.459, which means that the independent variables of the number of times of study, time of study per day, frequency of study per day, and completion of the task points can explain 45.9% of the variation of the dependent variable of the academic performance, and in general, an R^2 above 30% means that the regression equation is well fitted.

Table 2: Model summary

Model	R	R^2	ΔR^2	Standard estimation error	Texbin Watson
1	0.632	0.459	0.436	26.5241	1.766

(2) Determination of residual normality

The histogram of regression standardized residuals and the normal P-P plot are used to discriminate whether the residuals of the regression model obey the normal distribution, and the results are shown in Figures 5 and 6. The red curve in the histogram is the curve of normal distribution, while the purple bar indicates the regression standardized residuals, and the contour of the purple bar can be seen as basically aligned with the curve of normal distribution. The purple curve in the normal P-P plot basically coincides with the normal distribution curve, indicating that the residuals of the regression model obey the normal distribution.

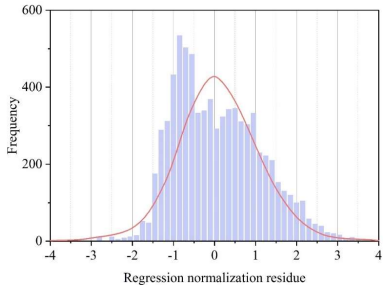

Figure 5: Regression standardized residual histogram

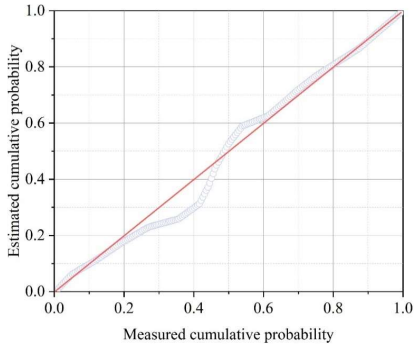

Figure 6: Normal P-P plot of regression standardized residuals

(3) The independent variables do not exist multicollinearity discrimination

From the above Table 1 covariance statistics column items can be seen, the minimum tolerance of 0.592 is much

larger than 0.1, the maximum VIF value of 1.698 is less than 5, when the VIF value is less than 5, it indicates that there is no multicollinearity between the respective variables, and it will not affect the accuracy of the regression results [17]. Comprehensive diagnosis of linear regression above can be seen that the regression model meets the prerequisites of sample independence, normal distribution of residuals and the absence of multicollinearity of independent variables, indicating that the results of the constructed prediction model of online learners' academic performance are basically accurate.

V. Optimization of teaching strategies in the English online classroom

The previous section determined through regression analysis that the number of study sessions, daily study time and frequency, and task point completion all have a large impact on learning achievement. In English online classroom teaching, teachers can utilize the collected data to form teaching strategies based on data analysis.

V. A. Measuring and analyzing the situation and developing teaching programs

The analysis of learning situation before class is really important, is an important factor in the success or failure of lesson preparation and class. Very often teachers start to analyze the learning situation through their own perceptual feelings, such as students' interests and hobbies, understanding the mastery or understanding of relevant knowledge, learning habits and attitudes, etc. [18]. In teaching, teachers should do to understand the cognitive structure of the students, understand the students' weak points, understand the students' personality traits and interests, etc., so that they can carry out targeted teaching, targeted teaching, in order to tailor the teaching to the students, scientific design of the teaching process, the flexible use of teaching methods, so as to carry out effective teaching.

Teachers in teaching can combine the feedback of the pre-test data with the students' passing situation to understand which knowledge students have mastered better, which have not fully mastered, what deviations and problems exist. According to the data to make timely adjustments to the teaching in order to reduce ineffective labor, and accurately find the students who still have problems in understanding the knowledge points in the classroom to focus on breakthroughs.

V. B. The platform monitors learning in real time for personalized tutoring

Compared with traditional teaching, there are the following advantages when utilizing tablet PCs for classroom practice based on interactive technology:

(1) Teachers monitor all students' answers in real time.

(2) Accurately find students who have difficulty in answering questions for personalized tutoring.

(3) The results of the question and answer are generated on-site statistics, which can be explained in a hierarchical manner to improve the efficiency of the lecture.

Students use the tablet to do questions, at this time the teacher's activities are divided into two parts. The first part: through the platform feedback data to find the slow answer students and answer the question of the high error rate of students for individual counseling, to understand their knowledge mastery, personalized and accurate counseling. The second part: after the end of the answer time, count the results of the answer and focus on the topics with a low pass rate.

V. C. Posttest Comparison, Data Validation of Learning Effectiveness

The comparison of data between the rationality of the classroom test and the post-test verifies that the teacher's teaching is working and student learning is changing. The data can provide timely feedback on student learning and enrich the assessment of students. Based on the platform in addition to students can have a summative evaluation, but also provides a process evaluation. Students can also see the process evaluation reports generated, allowing them to see the progress they have made through their learning efforts.

VI. Conclusion

This paper analyzes the correlation between learning behavior and learning achievement based on the learning data of English online courses on Super Star Learning Platform for four classes of a major in School B. After that, a predictive model of online learners' academic performance was constructed by using multiple regression analysis $Y = -8.632 + 0.99x_1 + 0.231x_2 + 0.485x_3 + 0.286x_4$.

The R^2 of the model was 0.459, and the regression equation was well fitted. It shows that the independent variables of number of studies, daily study time, daily study frequency, and task point completion can explain 45.9% of the variance of the dependent variable of academic performance. And the residuals of the regression model obey normal distribution. The maximum VIF value is 1.698, which is less than 5, indicating that there is no multicollinearity

between the respective variables. The above test results indicate that the prediction results of the online learners' academic performance prediction model constructed in this paper are basically accurate.

In order to further improve the teaching effect, it is proposed that the use of rational data can be used to conduct accurate learning analysis, formulate teaching plans, carry out personalized tutoring as well as verify the learning effect.

References

[1] Susanty, L., Hartati, Z., Sholihin, R., Syahid, A., & Liriwati, F. Y. (2021). Why English teaching truth on digital trends as an effort for effective learning and evaluation: opportunities and challenges: analysis of teaching English. Linguistics and Culture Review, 303-316.

[2] Bui, T. H. (2022). English teachers' integration of digital technologies in the classroom. International Journal of Educational Research Open, 3, 100204.

[3] Sim, J. S. E., & Ismail, H. H. (2023). Using digital tools in teaching and learning English: Delving into English language teachers' perspectives. Creative Education, 14(10), 2021-2036.

[4] Salam, U., Wahdini, W., Surmiyati, S., Rezeki, Y. S., Riyanti, D., & Suthathothon, P. (2023). Teachers' challenges and strategies in using digital media in teaching English. Journal of English Language Teaching Innovations and Materials (Jeltim), 5(1), 49.

[5] 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.

[6] Hz, B. I. R., & Daulay, E. (2021). Online learning media: English education department students' perspective. Metathesis: Journal of English Language, Literature, and Teaching, 5(1), 50-64.

[7] Suputra, D. (2021). Teaching English through online learning (A literature review). The art of teaching English as a Foreign Language, 1(2), 65-70.

[8] Putri, N. R., & Sari, F. M. (2021). Investigating English teaching strategies to reduce online teaching obstacles in the secondary school. J. English Lang. Teach. Learn, 2(1), 23-31.

[9] Sari, F. M. (2020). Exploring English learners' engagement and their roles in the online language course. Journal of English Language Teaching and Linguistics, 5(3), 349-361.

[10] Kassymova, G. M., Tulepova, S. B., & Bekturova, M. B. (2023). Perceptions of digital competence in learning and teaching English in the context of online education. Contemporary Educational Technology, 15(1), ep396.

[11] Huang, Q. (2022). Does learning happen? A mixed study of online chat data as an indicator of student participation in an online English course. Education and Information Technologies, 27(6), 7973-7992.

[12] Ayu, M., & Sari, F. M. (2021). Exploring English Teachers' Strategies in Managing Online Learning through Google Classroom. ELT Worldwide: Journal of English Language Teaching, 8(2), 318-330.

[13] Xu, D., & Tsai, S. B. (2021). A Study on the Application of Interactive English-Teaching Mode under Complex Data Analysis. Wireless Communications and Mobile Computing, 2021(1), 2675786.

[14] Zhang Lulu,Tian Yue & Song Shibo. (2021). Research on the influence of college students' self-directed learning behavior based on online learning platform. Journal of Physics: Conference Series,1931(1),

[15] Gholamreza Hesamian,Faezeh Torkian,Arne Johannssen & Nataliya Chukhrova. (2024). A learning system-based soft multiple linear regression model. Intelligent Systems with Applications,22,200378-.

[16] Yu Qiqing & Liu Ruiqi. (2020). A consistent test of independence and goodness-of-fit in linear regression models. Communications in Statistics - Simulation and Computation,51(7),1-20.

[17] Shakeel Ahmad,Abdul Majid & Muhammad Aslam. (2024). On Some Robust Liu Estimators for the Linear Regression Model with Outliers: Theory, Simulation and Application. Journal of Statistical Theory and Practice,18(4),49-49.

[18] Luyan Zheng & Keok Cheong Lee. (2023). Examining the Effects of "Small Private Online Course and Flipped-Classroom"-Based Blended Teaching Strategy on First-Year English-Major Students' Achievements. Sustainability,15(21),