Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 25, Issues 1, pp. 255–270, DOI 10.12694/scpe.v25i1.2292

ENGLISH DISTANCE TEACHING BASED ON SPOC CLASSROOM AND ONLINE MIXED TEACHING MODE

MEIJUAN ZHANG*AND XIAOLI ZHU†

Abstract. At present, the English Small Private Online Course (SPOC) online mixed teaching model has problems in evaluating students' learning and organizing teaching papers. For example, the evaluation is chaotic and unable to meet the key points of organizing the paper. Starting from the thinking chain that accepting learning outcomes can promote learning behavior, a score prediction method and test paper generation algorithm (TPGA) based on a learning evaluation diagnostic model are designed. Among them, the performance prediction algorithm is designed by combining multiple linear regression (MLR) and random forest (RF). The TPGA is based on students' learning status. The research results show that most of the predicted values output by the performance prediction model are not significantly different from the actual values. They are within a reasonable range. Meanwhile, under the influence of TPGA, the number of students in the experimental group is higher in the 70-80 and 80-90 segments, with 27 and 6, respectively. The experimental group has a higher average score rate on each type of question and knowledge point. Both models have high student satisfaction, indicating that the results oriented online mixed learning strategy designed in the study can effectively improve students' learning outcomes.

Key words: SPOC classroom; English online teaching; Teaching system design; Mixed teaching

1. Introduction. With the development of computer and network technology, the limitations of traditional classroom English education are gradually becoming apparent. These limitations are mainly reflected in the time and space, as well as significant limitations on students and classroom forms [1, 2, 3, 4]. However, English distance teaching based on online teaching technology and computer technology can solve the limitations of traditional classroom English education. This educational approach not only solves time and physical limitations, making teaching more flexible, but also makes the teaching process more intelligent, breaking away from the manual handover part of traditional classroom teaching [5, 6]. However, online English teaching also has certain drawbacks. For example, due to the main use of online video connections in education, it is more difficulty to grasp students' classroom learning status. The learning effect is unstable. The hybrid mode of online teaching solves this problem. This method redeploys online learning activities through activity and resource oriented, forming a more stable and flexible teaching system [7]. SPOC (Small Private Online Course) classroom refers to a small-scale restricted online classroom, which means that students participating in classroom education need to use online learning and must be restricted by classroom access conditions [8]. Only students who meet the requirements can be allowed to enter the classroom. The main limitations of SPOC teaching methods include group work, classroom cohesion, and acceptance of learning achievements. Therefore, by combining SPOC teaching with online mixed teaching mode, this study mainly focuses on evaluating the learning outcomes of students in teaching [9]. Then a teaching achievement evaluation system is established to improve the effectiveness of SPOC online mixed teaching system. The main motivation and novelty of this study lies in the focus not only on the advantages of online teaching, but also on the existing problems. It is attempted to solve these problems through the combination of hybrid mode and SPOC teaching. The contribution lies in proposing a new teaching achievement evaluation system to improve the effectiveness of SPOC online blended teaching system. By deeply understanding students' learning processes and outcomes, teaching effectiveness can be truly improved [10]. Therefore, this special method is chosen. The first part of the study introduces research purposes. The second part designs a learning evaluation model and paper formation strategy based on SPOC classroom theory. The third part conducts experimental verification on the learning

*School of Foreign Languages, Jiangsu Open University, Nanjing, 210036, China (Corresponding author: Meijuan_Zhang2023@outlook.com)
†Department of Student Affairs Management, Nanjing Institute of Technology, Nanjing, 211167, China

evaluation model and paper formation strategy research based on SPOC classroom theory. The fourth part draws research conclusions.

2. Related Work. In recent years, the research on SPOC classroom has been gradually enriched. Aiming at the drawbacks of cramming teaching in traditional college English translation courses, Sun H proposed a SPOC teaching method. It is named “Rain Classroom”, which is beneficial for improving classroom learning efficiency. This teaching mode integrates the WeChat platform with other multimedia intelligent teaching tools. The learning process is divided into three steps, namely before, during and after class. The study findings revealed that the method has a positive effect on the reform of English translation teaching [10]. Zhang’s team designed a teaching method based on the concept of conclusion-based interpretation and SPOC flipped classroom. This model follows the flipped classroom journey of teaching objectives-teaching activities-learning evaluation. A two-year experiment is carried out in the teaching of engineering cost specialty. According to the findings, the teaching model can improve learning efficiency [8]. Law et al. designed a SPOC flipped classroom teaching mode for private teaching. This mode fully utilizes online teaching mode, video teaching, layered teaching, interactive discussion and other learning forms. They are integrated into a complete teaching framework. Under the influence of this teaching mode, students’ total grades are positively correlated with their classroom participation [12]. The team established a hierarchical learning method based on the SPOC flipped classroom. The model is divided into three main parts, namely the cognitive layer, the design layer and the application layer. For the teaching, the teaching form centered on activities and resources is adopted. A classification and grading evaluation method is used in evaluating teaching effectiveness. Compared with the traditional teaching mode, students are more satisfied with the improved teaching mode [13]. The Kastrati team proposed an exploratory model for the value of customized machine learning SPOC education models. Students’ cognition and attitudes towards the SPOC teaching model are analyzed. Study findings demonstrate that the SPOC teaching mode more directly affects the students’ knowledge and skills [14].

Computer technology is being used more wisely in classroom planning. The Gitinabard team designed a corresponding student performance prediction model for the hybrid online-offline teaching approach. The method combined the performance of students in offline classrooms with the evaluation results of online classrooms to form early predictions. The research results show that the method achieves stable cross-category data prediction [15] . Liu Q’s team proposed an intelligent education evaluation model for the performance prediction of students in the physical education. The model is mainly based on the LSTM model. It tracks the students’ exercise activities and extracts sufficient information from the exercise activity to predict performance. The results show that the model is effective [16]. Chen’s team proposed an automatic TPGA based on genetic algorithm. It adjusts the difficulty factor according to the online learning data. The research results show that the model can effectively control the difficulty of the test paper [17]. Computers are more inclined towards intelligent improvement of traditional teaching procedures in teaching. Therefore, the research also conducts a model for teaching evaluation and test-taking strategies. A one-stop automated online teaching system is designed, providing a new realization path for SPOC online blended teaching.

The Soufiane O team mainly explored the three basic concepts of SPOC, blended learning, and flipped classroom. A working method for developing online training in the form of SPOC is proposed. This study mainly focused on issues such as student evaluation and teaching paper organization in the online blended teaching mode. A score prediction method and test paper generation algorithm based on a learning evaluation diagnostic model are designed. Therefore, this study has innovation in practical application and design of practical methods [18]. Zhou Y focused on the implementation of SPOC blended learning mode in the choir command teaching system. Although all research is focused on the SPOC blended learning model, this study is to address student learning evaluation and organization of teaching papers. Reference 2 focuses on how to improve the effectiveness of practical teaching in choir conducting. The characteristic of this study is to focus on the student learning evaluation and generate solutions through machine learning prediction [19]. The Ram í rez Monoso team studied a gamified mobile collaboration tool to improve the utilization of online learning resources. It is based on a learning evaluation diagnostic model, which uses complex machine learning algorithms to predict students’ learning outcomes and organize teaching papers. Both are efforts to improve students’ learning outcomes. But the exploration of evaluation systems and machine learning applications in this study is more in-depth and innovative [20].

Table 2.1: Literature Content

Author Name	Research Contents	Findings
Sun H [10]	Proposed a SPOC teaching method called ”Rain Classroom” to improve classroom learning efficiency	The designed methods have positive impacts on the reform of English translation teaching
Zhang et al [8]	Designed a teaching method based on conclusion interpretation concepts and SPOC flipped classroom	The designed teaching model can improve students’ learning efficiency
Law et al. [12]	Designed a SPOC flipped classroom teaching model for private teaching	Under the influence of this teaching model, students’ total grades are positively correlated with their classroom participation
An X [13]	Established a layered learning method based on SPOC flipped classroom	Compared to traditional teaching models, students are more satisfied with the improved teaching model
Kastrati et al [14]	Proposed an exploratory model for SPOC education model for applying machine learning	The SPOC teaching model directly affects students’ mastery of knowledge and skills
Gitinabard et al [15]	Designed a corresponding student performance prediction model for hybrid online offline teaching methods	This method achieves stable cross-category data prediction
Liu Q et al [16]	Proposed an intelligent education evaluation model for predicting the performance of sports students	The designed model is effective
Chen et al [7]	Proposed an automatic TPGA based on genetic algorithm to adjust difficulty factors based on students’ online learning data	This model can effectively control the difficulty of the test paper
Soufiane O et al [18]	Studied the basic concepts of SPOC, blended learning, and flipped classroom. Proposed a working method for conducting online training in the form of SPOC. Mainly focused on the student evaluation and organization of teaching papers in online blended teaching mode, and designed a score prediction method and paper generation algorithm based on a learning evaluation diagnostic mode	Innovative in practical applications and design practices
Zhou Y [19]	Studied the implementation of SPOC blended learning mode in the choir command teaching system.	Focusing on solving the student learning evaluation and teaching paper organization, and proposed solutions to student learning evaluation problems through machine learning prediction and algorithm generation
Ramirez Monoso et al [20]	Researched a gamified mobile collaboration tool to enhance utilization of online learning resources.	Based on a learning evaluation diagnostic model, complex machine learning algorithms are used to predict learning outcomes and organize teaching papers

This study conducts in-depth research on the student learning evaluation and teaching paper organization. A new scoring prediction method and test paper generation algorithm are designed, which have obvious innovation in specific problem-solving methods, machine learning applications, and practical applications. Literature contents are shown in Table 2.1.

In recent years, research on SPOC classrooms has gradually enriched. Many research teams have designed and implemented various SPOC based teaching models. The results demonstrate their positive impact on improving students’ learning efficiency and engagement. However, although these studies have made some important discoveries, there are still many unexplored areas in how to apply computer technology more intelligently to classroom planning. For example, how to design models that can predict student performance. How to adjust the difficulty of exam papers based on students’ online learning data. They are currently important

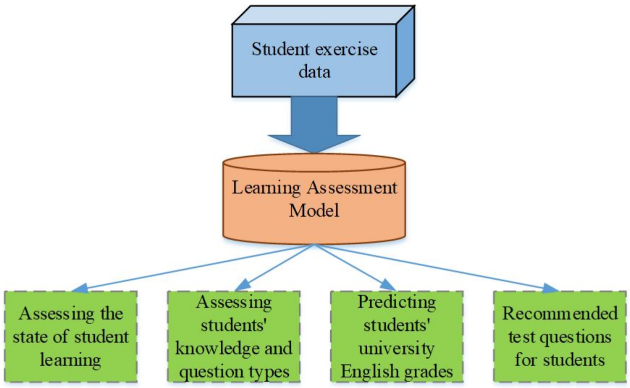


Fig. 3.1: Framework Diagram of Learning Evaluation Model Based on SPOC Classroom and Online Mixed teaching Mode

research directions. Therefore, this study will design a model for teaching evaluation and exam strategies. A one-stop automated online teaching system of teaching evaluation product prediction automatic exam is implemented, providing a new implementation path for SPOC online blended teaching. This will be the novelty and contribution of this study.

When designing scoring prediction methods and test paper generation algorithms (TPGA), relying on learning evaluation diagnostic models, some new and not yet widely applied technologies such as deep learning and neural networks are also introduced. This will increase the innovation of the research. The research results not only demonstrate the accuracy of the prediction model, but also reveal the positive impact of TPGA on students' academic performance. This discovery has not been fully explored in previous studies. Therefore, the research provides a new perspective for the online blended learning strategies. The research results indicate that results-based online blended learning strategies can effectively improve students' learning outcomes. This conclusion not only validates the research hypothesis, but also provides a new practical strategy for the online blended teaching model, which has a certain degree of innovation.

3. Diagnostic model and verification of test paper formation strategy.

3.1. Construction of Learning Evaluation Model Based on SPOC Classroom Theory. The goal of this research is to create a teaching system that can analyze students' learning status, knowledge point relevance and predict college English performance based on SPOC classroom theory, providing personalized learning suggestions for students in English distance teaching. Figure 3.1 is the essential framework diagram of the learning evaluation model based on the SPOC classroom and online mixed teaching mode proposed in this study.

In Figure 3.1, firstly, a learning evaluation model is built. Then the student learning data is preprocessed to build a sub-model from the three dimensions, namely learning status, knowledge point correlation, and college English grade test scores. Finally, the three sub-models are combined. A learning evaluation model according to the SPOC theory is obtained. Student problem chart (S-P table) analysis method is used to analyze the learning situation of students. Among them, the student's attention coefficient is an important parameter in the analysis. The question type and knowledge point (QTKP) score are extracted from the database to construct the SP table. For values higher than or equal to the average, they are recorded as 1. Those that are lower than the average value are recorded as 0. The S-P table analysis method is used to analyze students' familiarity with QTKP, and calculate students' attention coefficient. An example of an S-P table is shown in Figure 3.2

The calculation method of student's attention coefficient is shown in equation 3.1.

$$AF_a = \frac{SumSL_0 - SumSR_1}{SumSL_R - TS_a \cdot SAS} \tag{3.1}$$

	Title 7	Title 1	Title 10	Title 6	Title 3	Title 2	Title 5	Title 9	Title 8	Title 4	Total score	%
Student 2	1	1	1	1	1	1	1	1	1	1	10	100
Student 5	1	1	1	1	1	1	1	0	1	1	9	90
Student 9	1	0	1	1	1	0	1	1	1	0	7	70
Student 4	1	1	1	1	0	0	0	1	1	1	7	70
Student 8	1	1	1	1	1	1	0	0	0	0	6	60
Student 3	1	0	1	0	0	1	1	1	1	0	6	60
Student 1	1	1	1	0	1	0	1	0	0	1	6	60
Student 6	1	1	1	1	0	1	1	0	0	0	6	60
Student 10	1	1	0	1	1	0	0	0	0	1	5	50
Student 7	0	1	0	0	0	1	0	1	0	0	3	30
Total score	9	8	8	7	6	6	6	5	5	5		

Fig. 3.2: Schematic Diagram of S-P table

In equation 3.1, $SUMSL_0$ represents the student's quantity who answered wrongly on the left side of the S curve. $SumSR_1$ is the student's quantity who answered correctly on the right side of the S curve. TS is the students quantity who answered correctly on the left side of the S curve. is the total score of the students a .SAS is the average score of all students. The larger the value of the attention coefficient is, the more unstable the learning state of the students. Teachers need to pay more attention. In normal conditions,AF_a the value is less than 0.5. If the value of AF_a is higher than 0.5, teachers should pay attention to the state of students. If it is higher than 0.7, teachers need to pay more attention. According to the AF_a value and the score of each knowledge point, the learning status is evaluated and analyzed. The specific formula for students' familiarity with each QTKP is shown in equation 3.2.

$$MS = \frac{KPQ_{GE}\text{ No.}}{KPQ\text{No.}} \times 100\% \tag{3.2}$$

The method used in the study is based on students' score rates in various knowledge points and different question types. By calculating familiarity and attention coefficient, students' college English grades are predicted. The main variables include students' score rates on various knowledge points and different question types, test scores, and the total number of correct answers in the knowledge point question types. The effective parameters for the main variable include the average score of knowledge points and question types, as well as the percentage of students' average score. Firstly, the student's score rates for each knowledge point and different question types are obtained. They are stored in a matrix format. Then, familiarity with each question type and knowledge point, and attention coefficients are calculated. Finally, the students are determined based on their familiarity and attention coefficient. The percentage of students' scores in knowledge point question types and attention coefficients are output.

The learning state evaluation algorithm (LSEA) is used to calculate students' familiarity with each QTKP and the student's attention coefficient, so as to provide personalized teaching guidance to students according to the learning style. The following is the specific calculation procedure. The initial step is to obtain the scores of students in each knowledge point and different question types (QTs). They are stored in a matrix form. There are A students in the database. Each student's scoring rate is B. The specific calculation method of the matrix is shown in equation 3.3.

$$y_{ab} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ y_{o1} & y_{o1} & \cdots & y_{op} \end{bmatrix} \tag{3.3}$$

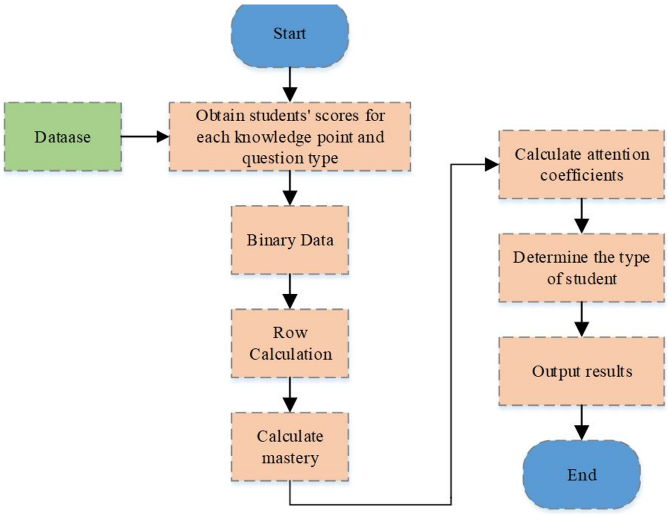


Fig. 3.3: Algorithm Steps for Assessing Students' Learning Status

The LSEA can only deal with binary data, but the scoring rates of each knowledge point and different QTs are numbers distributed within the range [0,1]. Therefore, it must be processed uniformly. The average of the QTKP score is assumed to be the threshold. It is also used as a criterion for determining familiarity. Those greater than or equal to the threshold are recorded as familiar and marked as 1, while unfamiliar ones are marked as 0. The specific formula is shown in equation 3.4.

$$Line(y_{ab}) = \begin{cases} 0, & \text{if } y < \frac{\sum_{a=1}^{o} y_{ab}}{o} \\ 1, & \text{if } x \geq \frac{\sum_{a=1}^{o} y_{ab}}{o} \end{cases} \tag{3.4}$$

In equation 3.4, $y_a b$ is the knowledge score percentage of the a student on the QTb . The score of student a is y_a. The number of students with correct answers of the QTb is y_b. The specific expression is shown in equation 3.5.

$$\begin{cases} y_a = \sum_{b=1}^{p} y_{ab}, \\ y_b = \sum_{a=1}^{o} y_{ab}, \end{cases} \tag{3.5}$$

Then the familiarity with each QTKP, and attention coefficient are calculated. $y_a b$ is the student's score percentage in the QTKP. y_a is the student a's test score.y_b is the total number of students who answered correctly in QTb. η is the student's average score percentage. The expression of the student's attention coefficient $AF_a$is shown in equation 3.6.

$$AF_a = 1 - \frac{\sum_{b=1}^{o} (y_{ab})(y_b) - (y_a)(\eta)}{\sum_{b=1}^{y_a} y_b - (y_b)(\eta)} \tag{3.6}$$

Then the student type is determined. The students' familiarity and attention coefficient obtained in the above calculation are judged according to the student classification standard. Finally, the student's score percentage in the QTKP, the student's attention coefficient and the student type are output. The working method of LSEA is described above. Figure 3.3 displays the specific flowchart.

In Figure 3.3, firstly, it is necessary to extract all data related to the user's usage process from the database, including the number of questions done, exam score rate, knowledge point score rate, question type score rate, and CET-4 score. Next, the collected data is processed, which includes steps such as cleaning the data and processing invalid data. In the feature selection stage, the information related to the predicted target is used

Table 3.1: Examples of Student Scores for Each Question Type

Student number	Essay writing	Short news	Long conversation	Listening chapters	Vocabulary comprehension	Long read	Read carefully	Chinese to English
00100	F	F	F	F	P	F	F	F
00101	P	P	F	P	F	P	T	P
00102	F	F	P	P	P	P	F	P
00103	F	P	F	P	P	P	F	P
00104	P	F	P	F	F	P	P	F
00105	P	P	F	P	P	F	P	P
00106	P	F	P	F	P	P	F	F

as a feature, which should be related to English practice. At the same time, by observing the data, hidden features are discovered to increase feature dimensions. In the construction step of the prediction model, the dataset is divided into the training set and the validation set. Then, a random forest model and a multiple linear regression model are used to construct prediction models for the training set, respectively. The model fusion is carried out through voting method. Finally, the prediction model is validated through a validation set. Once verified, the model can be launched for users to use. This is the entire process of constructing a prediction model for college English Test Band 4 scores.

At present, the student's question score is a continuous value distributed between [0,1]. The Apriori algorithm is a Bool-type attribute association analysis algorithm. Therefore, the score should be replaced by a binary Boolean data P and F . The difficulty of each QT varies. Based on the average score of each QT, the data is performed binary-boolean operations. If the student's score on a QT is higher than or equivalent to the average, then it is defined as P . If the student's score on a QT is less than the average, then it is defined as F .$CR_a b$ is the percentage of student 's score on QT b. The specific expression is shown in equation 3.7.

$$CR_{ab} = \begin{cases} F, & \text{if } CR_{ab} < \overline{CR_{ab}} \\ P, & \text{if } CR_{ab} \geq \overline{CR_{ab}} \end{cases} \tag{3.7}$$

The above cleaned data in Table 3.1 are used for hierarchical and symmetric analysis, i.e. analyzing the association between P and F , high and low scoring questions. Then they are combined to produce an association rule table.

Data are analyzed using the Apriori algorithm in SPSS. SPSS converts data into executable data for Apriori algorithm through data flow. The specific operation process is shown in Figure 3.4.

According to Figure 3.4, the data is first filtered, followed by binary conversion. Finally, the Apeior algorithm is introduced to obtain the analysis results.

3.2. Construction of the Prediction Model for College English Scores Based on SPOC and Online Mixed Teaching Mode. Predicting CET-4 scores is not a feature of the current college English practice diagnosis system. Many students who have just entered college are unfamiliar with the QTKP of CET-4. They can't accurately estimate their ability. Therefore, they are unsure if they can pass level 4. In this study, when using the original system for field research, 58% of students believe that even though they complete some CET-4 exercises, they cannot fully confirm their English skills. They are not sure whether they can pass the college English proficiency test. The flowchart of the algorithm is shown in Figure 3.5.

In Fig. 3.5, this study collects the data during the initial use of the system and the scores left by students for college English college exams. Following data processing and feature screening, the data are trained with MLR and RF respectively. The vote method is used for model fusion. After verification, a more accurate predictive model of college English grades is finally obtained. The above data alone cannot accurately predict CET scores, and hidden features must also be extracted. To improve the prediction accuracy, the authors uses the following hidden features. The student's test score is the basis of evaluation, so this score should be added

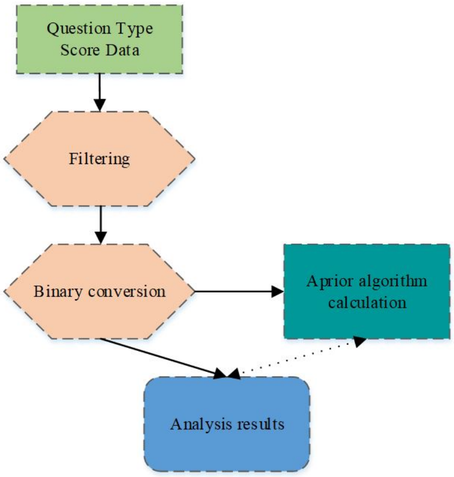


Fig. 3.4: Flow Chart of SPSS Data Flow Operation

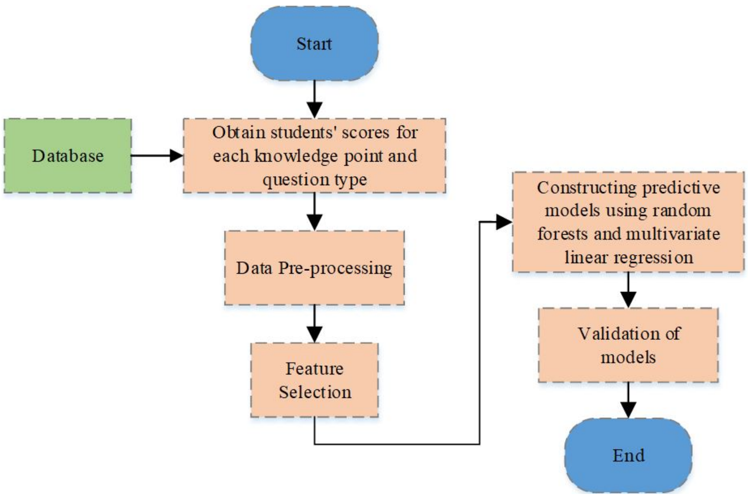


Fig. 3.5: Flow chart of college English grade score prediction

to the feature set. The specific calculation formula is shown in equation 3.8.

$$SR = \frac{AC}{T} \tag{3.8}$$

In equation 3.8,AC represents the number of questions answered correctly. T represents the question quantity in the test paper. After adding hidden features to the feature set, interactive variables need to be built at the same time. The purpose of adding interaction variables is to explore the internal relationship between features. The prediction model can be created once the feature selection and data processing are finished. RF and MLR are used in this study. In addition, the root mean square error (RMSE) and root mean square error are used to verify the model.

$$\begin{cases} RMSE = \sqrt{\frac{\sum_{a=1}^{o}(Y_{PV,a}-Y_{RV,a})^2}{o}}, \\ RMSLE = \sqrt{\frac{\sum_{a=1}^{o}(\log(Y_{PV,a}+1)-\log(Y_{RV,a}+1))^2}{o}} \end{cases} \tag{3.9}$$

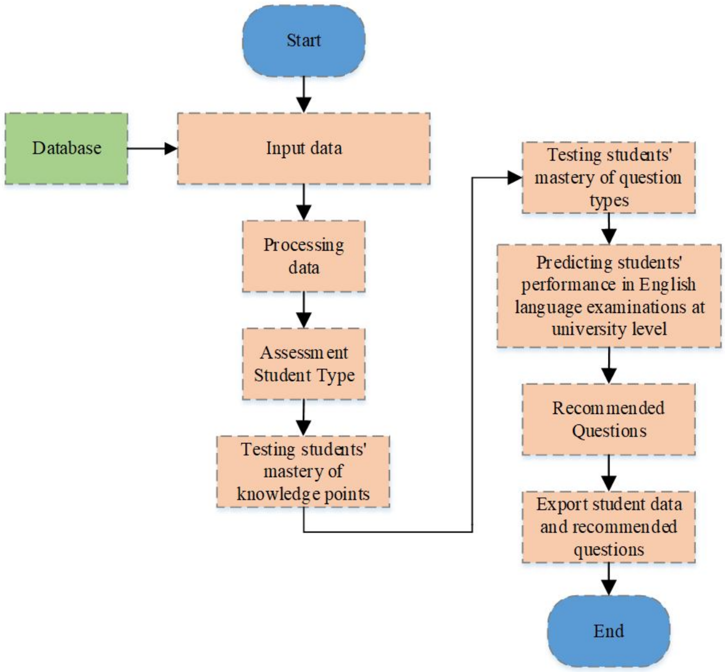


Fig. 3.6: Flow chart of student learning diagnostic evaluation model

In equation 3.9, $Y_{PV,a}$ represents the predicted value and $Y_{RV,a}$ represents the real value. In summary, the above is an overview of the operational steps involved in the comprehensive diagnostic evaluation model. The flow diagram of the overall model is shown in Figure 3.6.

In Figure 3.6, when extracting data, the user data to be evaluated is obtained from the database. When processing data, to meet the requirements of the model, user data needs to be cleaned and transformed. When evaluating learner types, a learning state evaluation model should be used to calculate the user's mastery of knowledge point types and attention coefficients to determine the learner type. When diagnosing the mastery of user knowledge points, the knowledge point association rule table is adopted. Based on the user's knowledge point scoring rate, the user's knowledge strengths and weaknesses are determined. When diagnosing the mastery of user question types, the question type association rule table is used to determine the strong or weak item question types based on the user's question type scoring rate. When predicting user scores, a score prediction model is used to estimate user scores. When recommending test questions, targeted and personalized test question recommendations should be provided to users.

3.3. The Design Method of the TPGA for the Learning Evaluation Model Based on the SPOC Classroom Theory. At present, there are six algorithms widely used in the college English diagnosis system, namely practice question set, random question set, QT set, knowledge point set, knowledge point maximization set and knowledge point association rule set. Except for knowledge point maximization and knowledge point association, all these algorithms are relatively simple. Most of them require students to set parameters by themselves. According to the results of the statistical study presented in the prior section, the research has established a more accurate model for evaluating learning status. Therefore, this part proposes a test recommendation algorithm based on students' learning status. The algorithm considers the student's abilities and learning stability, reasonably allocates the number of test questions with different difficulties. Then it is implemented in the system, so that students can use it when updating the system. According to the goal of the document clustering algorithm based on students' learning status, the design idea of the algorithm is as follows.

1. Correspond the student type calculated by the student status assessment model with the difficulty of the work.
2. Read the attention factors calculated through the assessment model. The total number of test questions is shown in equation 3.10.

$$AT = AT_{four} \cdot (1 + AF) \tag{3.10}$$

   To prevent decimals from appearing in the calculation of the questions, this study adjusts them to integers in the algorithm.
3. Students' knowledge scores are distributed in 5 ranges from the highest score to the lowest score to match the 5 ranges in the distribution coefficient of the test difficulty. That is, knowledge points with higher scores are assigned to more difficult topics. Knowledge points with lower scores are assigned to easier topics.
4. The number of sub-tests assigned to each knowledge item is shown in equation 3.11.

$$AKP_{eachD} = \frac{AT}{AKP} \tag{3.11}$$

In the previous study, this study collects data from multiple colleges and universities in China. It is compared with the college English proficiency test questions in this study. It is concluded that the English level of some students is at the middle level. There is a little gap with the ability level of college students. Therefore, the student data of the school is used for research. The traditional educational measurement method adopts the scoring rate to judge the difficulty of test questions. The difficulty coefficient is applied to explain the difficulty. The specific calculation formula of the difficulty coefficient is shown in equation 3.12.

$$L_a = \frac{R_a}{E_a} \tag{3.12}$$

In equation 3.12,R_a represents the student quantity who answered the question correctly. E_a represents student quantity answering the question. Therefore, the difficulty coefficient P for the entire test paper is defined as shown in equation 3.13.

$$P = \frac{\overline{A}}{H} \tag{3.13}$$

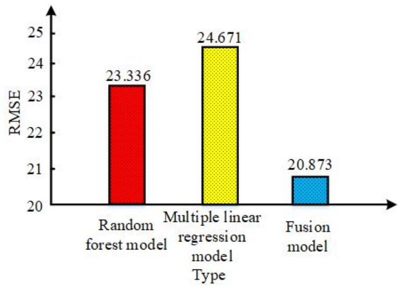
In equation 3.13, A is the average score of the test papers,$\overline{A}$ and is the full score of test papers.

From the score model of the college English proficiency test, the college English proficiency test adopts a normal distribution model to allocate the number of test questions with different difficulties. The normal probability density function formula is shown in equation 3.14.

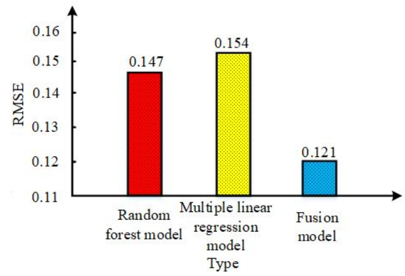
$$f(x) = \frac{1}{\sqrt{2\phi\gamma}} e^{-\frac{(x-\phi)^2}{2\gamma^2}} \tag{3.14}$$

ϕ is the expected value and γ is the variance of the normal distribution. The four-level score interval [0, 710] is divided into 5 main intervals, corresponding to the difficulty of the 5-level test questions. Specifically, they are [0, 0.2], [0.2, 0.4], [0.4, 0.6], [0.6, 0.8], and [0.8, 1.0]. The critical values are 0.2, 0.4, 0.6, 0.8, and 1.0 respectively. The corresponding four-level score values are 332, 364, 391, and 421 respectively. According to the obtained four-level score values, the score interval [0, 710] is divided into five score intervals. Each interval corresponds to a different probability interval of test difficulty, as shown in equation 3.15.

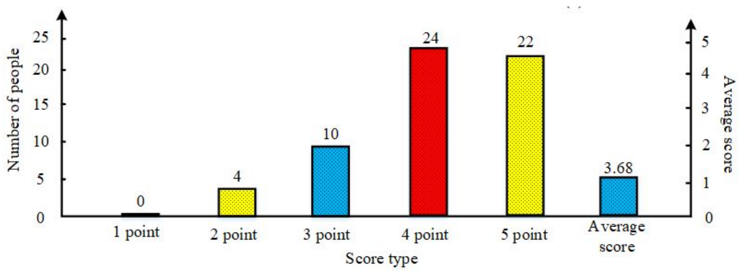
$$\begin{cases} L_1 = \int_0^{332} \sigma_{\phi,\gamma}\, dx \\ L_2 = \int_{332}^{364} \sigma_{\phi,\gamma}\, dx \\ L_3 = \int_{364}^{391} \sigma_{\phi,\gamma}\, dx \\ L_4 = \int_{391}^{421} \sigma_{\phi,\gamma}\, dx \\ L_5 = \int_{421}^{710} \sigma_{\phi,\gamma}\, dx \\ L = \sum_{a=1}^{5} L_a = L_1 + L_2 + L_3 + L_4 + L_5 \end{cases} \tag{3.15}$$



(a) RMSE



(b) RMSLE



(c) Evaluation Score

Fig. 4.1: Comparative Analysis Results and Satisfaction Results

4. The effect analysis of the diagnostic model and test-taking strategy.

4.1. Effect Analysis of Diagnostic Model. In terms of parameter configuration for random forest models, grid search and cross validation methods are used to determine the optimal parameter combination. The main parameters to be adjusted include the number of trees (n_estimators), selected from [100, 200, 300, 400, 500], max depth, selected from [10, 20, 30, None], and maximum number of features (max_features), selected from ['auto ',' sqrt ']. For the multiple linear regression model, the regularization parameter (alpha) is mainly adjusted and selected from [0.1, 0.01, 0.001, 0.0001]. In the diagnostic effect analysis, the RMSE and root mean square logarithmic error (RMSLE) are used for comparative analysis of different models. A functional evaluation questionnaire is designed to investigate the satisfaction of students who use the evaluation model with the model. The comparative analysis results and satisfaction results are shown in Figure 4.1.

From Figure 4.1, in the comparison of RMSE indicators, the RMSE value of the RF model is 23.336, and the RMSE value of the MLR model is 245.671. The RMSE value of the model incorporating voting methods is 20.873. The RMSE value of the fusion model designed by the research is the lowest. In the RMLSE index, the RMSE value of the RF model is 0.147. The RMLSE value of the MLR model is 0.154. After fusing the two models through voting, the RMLSE is 0.121. The designed fusion model has the lowest RMSE value, which is more accurate. In terms of satisfaction rating, the satisfaction score of 1 is 0 people. The satisfaction score of 2 is 4 people. The satisfaction score of 3 is 10 people. The satisfaction score of 4 is 24 people. The satisfaction score of 5 is 22 people. Most students are satisfied with the model effect.

In the comparison of RMSE indicators, the RMSE value of the RF model is 23.336, and the RMSE value of the MLR is 245.671. However, the RMSE value of the model designed in the study that fused these two models through voting method is 20.873. This means that the fusion model performs better in prediction error. As the RMSE value decreases, the prediction error decreases. The prediction accuracy of the model is higher. This is very important for practical applications. The model predicts the results as accurately as possible to make correct decisions based on the predicted results. After comparing the prediction models, the comparative results of the predicted value and the actual value obtained from the verification set data are shown in Figure 4.2.

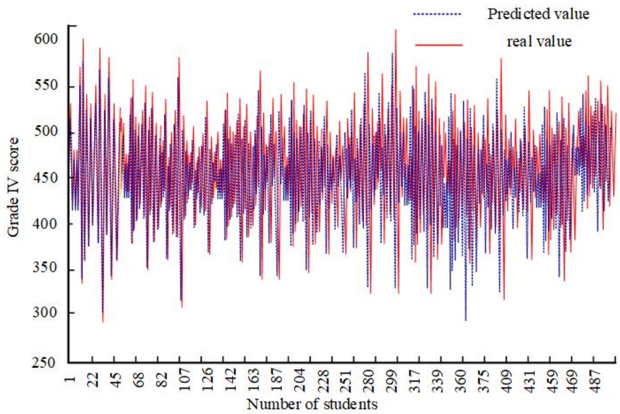


Fig. 4.2: Comparison chart of predicted value and actual value

In Figure 4.2, the abscissa indicates the total number of students. The ordinate represents the student's score. The majority of predicted values are relatively close to the actual values, which are all within a reasonable range. This shows that the model designed in the study is accurate and stable. It can still guarantee accurate performance prediction under large-scale data.

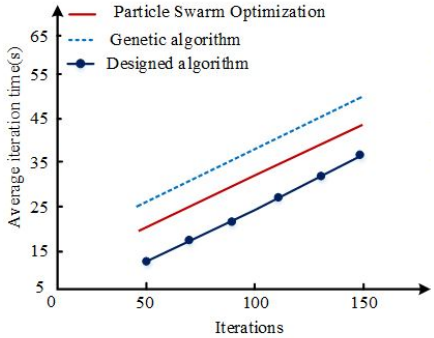
4.2. Experimental Verification of the Algorithm for TPGA. In terms of parameter configuration for the test paper generation algorithm, the following parameters are mainly adjusted, including population size, selected from [50, 100, 150, 200], and iterations, selected from [100, 200, 300, 400, 500]. The fitness function is set based on the difficulty, differentiation, and covered knowledge points of the test paper. Through multiple experiments and comparisons, the optimal parameter combination is ultimately determined. In verifying the TPGA, the running time and application effect of the algorithm in different question bank sizes are compared. The scale of the question bank mainly chooses four scales of 500, 1000, 15000 and 2000. The TPGA designed in the research is compared with the particle swarm TPGA and the genetic TPGA under different question bank scales. The running time comparison is shown in Figure 4.3.

From Figure 4.3, under the four scales of question bank scales of 500, 1000, 15000 and 2000, with an increase in iteration steps, the running time of the TPGA has a positive correlation with the number of iteration steps. Compared with the particle swarm and the genetic algorithm, the designed algorithm has the lowest dashed line position. It demonstrates that the duration of the TPGA is the shortest.

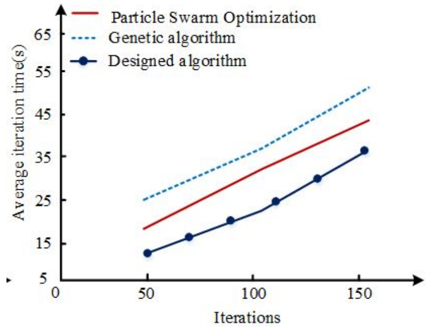
Under four scales of question bank size 500, 1000, 15000, and 2000, as the iteration steps increase, the variation in the running time formed by the designed test paper generation algorithm shows a positive correlation with the iteration steps. This indicates that the test paper generation algorithm designed in the study has higher operational efficiency when processing large-scale data. This is very important for practical applications, as it usually need to process a large amount of data. If the algorithm runs for too long, the efficiency in practical applications will be greatly reduced. The application effect is shown in Table 4.1.

From Table 4.1, the difficulty coefficient of the designed TPGA is 0.593, which is similar to the other two comparison algorithms. The error rate formed is only 0.837%, which is much smaller than other algorithms. The knowledge points cover rate is higher, reaching 98.31%. In terms of the TPGA, the scores of the algorithm designed in the research are 7.98, 27.65, 29.57, and 30.73 in memorization ability, comprehension ability, simple application ability and comprehensive application ability respectively, all of which are higher than other algorithms. The score of innovation ability is 14.28, which is close to 14.29 of genetic test paper. However, the TPGA designed by the research has higher scores in memorization ability, comprehension ability, simple application ability, and comprehensive application ability while maintaining the same innovation ability as other models. The effect is better. The model needs to run in a server environment. The running efficiency in the server is also analyzed, as shown in Figure 4.4.

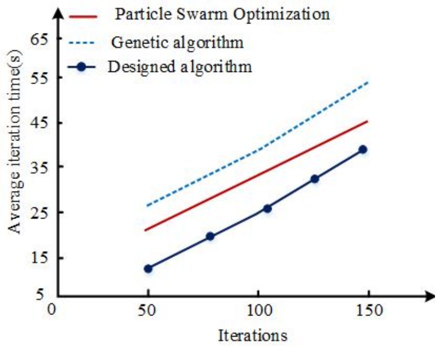
From Figure 4.4, the time for the test model in the identical machine testing circumstances is 1.49s, while



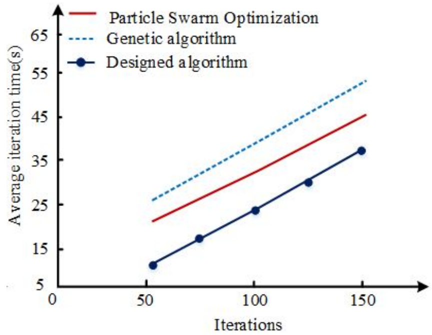
(a) Scale of Question Bank 500



(b) Scale of Question Bank 1000



(c) Scale of Question Bank 1500



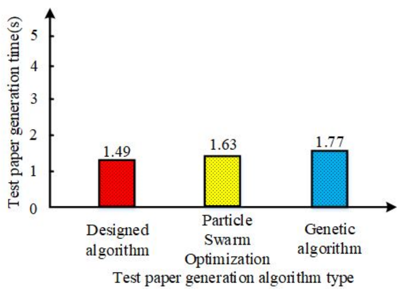
(d) Scale of Question Bank 2000

Fig. 4.3: Runtime Comparison

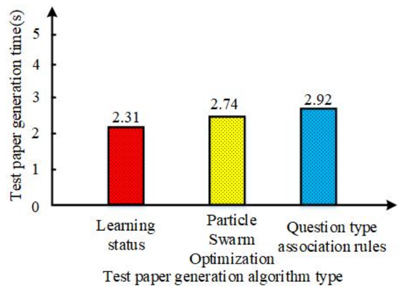
Table 4.1: Algorithm Application Effect Comparison

Algorithm		Genetic algorithm	Particle swarm algorithm	Research design algorithm
Degree of difficulty		0.586	0.626	0.593
Error rate (%)		2.168	4.002	0.837
Coverage of knowledge points		96.48	87.52	98.31
Ability Level Rating	Memorization ability	6.37	5.47	7.98
	Comprehension	24.93	26.75	27.65
	Simple Application Capability	27.63	23.48	29.57
	Comprehensive ability	26.85	27.57	30.73
	Creativity	14.29	11.82	14.28

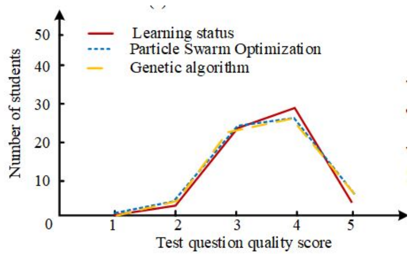
the particle swarm test model and the genetic test model in the local test environment are 1.63s and 1.77s. The designed model takes less time. It has higher calculation efficiency under the identical machine testing circumstances. In the server testing circumstances, the time for the designed model is 2.31s, while the particle swarm test model and the genetic test model are 2.74s and 2.92s respectively. The learning state of the TPGA designed by the research also takes less time to operate, which has higher operation efficiency under the same server test environment. In addition, in terms of the quality and quantity scores of the test papers, the number of people who scored between the quality scores of 1 and 5 in the TPGA designed by the study are 0, 2, 24,



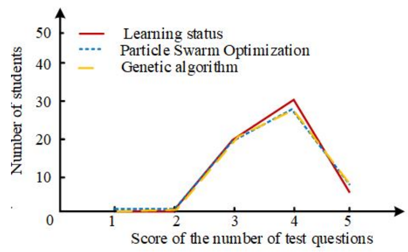
(a) Native Test Environment



(b) Server Test Environment



(c) Change of Test Question Quality Score



(d) Change of The Number of Test Question

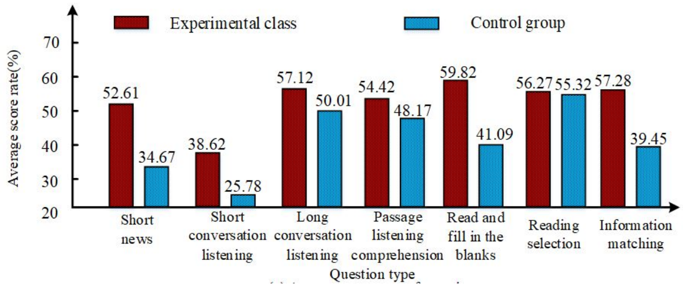
Fig. 4.4: Efficiency and quality analysis results of test papers

30 and 4 respectively. The number of people scoring between 1 point and 5 points are 0, 0, 24, 32 and 4 respectively. Compared with other models, more 4-point ratings and 5-point ratings are obtained. Student satisfaction is relatively higher. Finally, how the TPGA improves students' academic performance is examined. The knowledge points and question scoring rates are used to measure it. The specific results are shown in Figure 4.5.

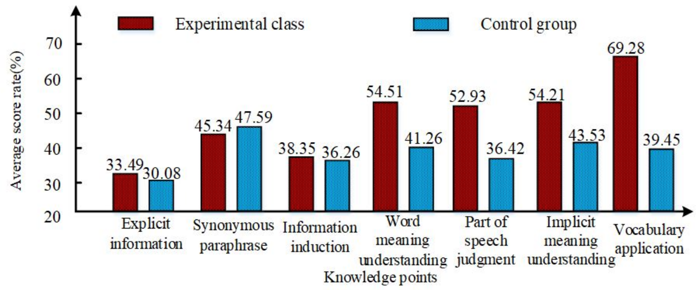
From Figure 4.5, with the help of the designed learning status TPGA, the students in the experimental group performs seven tasks in short news, brief conversation listening, long conversation listening, short listening comprehension, reading comprehension, reading selection and information matching. The average scoring rates in the QTs are 52.61%, 38.62%, 57.12%, 54.42%, 59.82%, 56.27% and 57.28%. In the control group, the average score who adopts the traditional learning system in the seven QTs are 34.67%, 25.78%, 50.01%, 48.17%, 41.09%, 55.32% and 39.45%. In the experimental group, the average score rates in the information disclosure, synonymous reporting, information induction, word meaning understanding, part of speech judgment, implicit meaning understanding and vocabulary application are 33.49%, 45.34%, 38.35%, 54.51%, 52.93%, 54.21% and 69.28% respectively. In the control group, the average score rates in the seven knowledge points are 30.08%, 47.59%, 36.26%, 41.26%, 36.42%, 43.53%, and 39.45%. In each QTKP, the average score rate of the experimental group students is relatively high. The learning status of the TPGA designed by the research can help students improve their grades.

With the help of the learning state test paper model designed in the research, the average score rate of the experimental group students in each question type and knowledge point is higher. This indicates that the learning state test paper generation model designed in the research can effectively improve their academic performance. This is very important for practical teaching, because the goal is to improve their academic performance. If a model can effectively improve their academic performance, then it is a successful model.

5. Conclusion. The research mainly starts from the achievement promotion behavior in the English SPOC online mixed teaching mode. A joint teaching strategy of learning diagnostic assessment + study performance



(a) Average Score rate of Question Type



(b) Average Score of Question Points

Fig. 4.5: Knowledge points and score rate of question types

prediction + test paper teaching model is designed. The MLR technique and the RF algorithm are combined to design a performance prediction model. The learning status is taken as the main consideration when designing the TPGA. According to the performance prediction model, the research results show that 24 people have a satisfaction rating of 4 points for the prediction model. 22 people have a satisfaction rating of 5 points. These results indicate that most people express high satisfaction with the accuracy and effectiveness of the prediction model. In terms of TPGA, the designed TPGA receives more 4 and 5 points, and student satisfaction is relatively high. This indicates that the TPGA has been widely recognized by students. It plays a positive role in improving their academic performance. In the experimental group, there are more students in sections 70-80 and 80-90, with 27 and 6 students respectively. This result further proves the effectiveness of the teaching strategy and predictive model. The designed performance prediction model and TPGA are effective, which can encourage students to improve their grades. This conclusion not only proves the effectiveness of the research method, but also provides new ideas and directions for future teaching models. While discussing the research results, the methods and models of this study are mainly based on the SPOC online blended teaching mode. Therefore, the applicability may be influenced by the teaching mode and student learning status. Regarding the limitations of case studies, the research mainly focuses on the student population with scores ranging from 70 to 90, which may overlook the learning situation of students in other grades. Therefore, future research should consider a wider student population to improve the universality of the model and teaching strategies. For future work, it is recommended to further optimize and adjust the learning diagnosis assessment, learning performance prediction, and exam teaching models to adapt to more teaching modes and student types. More research is expected to validate and improve the models and strategies of this study.

Funding. The research is supported by: ”The Design and Practice of Teaching English Courses at Open Universities ”, 2023 Teaching Reform and Research Project of Jiangsu Open University(23-YB-08); “Exploration and Practice on Improving the Effectiveness of Continuing Education Courses from the Perspective of Educational Psychology”, 2022 Higher Education Scientific Research and Planning Project, the Chinese Association of Higher Education (22JX0413).

REFERENCES

[1] Chakraborty, P., Mittal, P., Gupta, M., Yadav, S. & Arora, A. Opinion of students on online education during the COVID-19 pandemic. *Human Behavior And Emerging Technologies.* **3**, 357-365 (2021)
[2] Dhawan S. Online learning:, A. panacea in the time of COVID-19 crisis. *Journal Of Educational Technology Systems.* **49**, 5-22 (2020)
[3] Sekeroglu, B., Dimililer, K. & Tuncal, K. . *Student Performance Prediction And Classification Using Machine Learning Algorithms[C]//Proceedings Of The 2019 8th International Conference On Educational And Information Technology.* pp. 7-11 (2019)
[4] Mahmood, S. Instructional strategies for online teaching in COVID-19 pandemic. *Human Behavior And Emerging Technologies.* **3**, 199-203 (2021)
[5] Nambiar, D. The impact of online learning during COVID-19: students' and teachers' perspective. *The International Journal Of Indian Psychology.* **8**, 783-793 (2020)
[6] Fauzi, I. Khusuma I H S. *Teachers' Elementary School In Online Learning Of COVID-.* **5**, 58-70 (2020)
[7] Chen, J. SPOC-based flipped learning model applied in interpreting teaching. *International Journal Of Emerging Technologies In Learning (iJET).* **15**, 4-13 (2020)
[8] Zhang, L., Xuan, Y. & And, Z. and application of spoc-based flipped classroom teaching mode in installation engineering cost curriculum based on obe concept. *Computer Applications In Engineering Education.* **28**, 1503-1519 (2020)
[9] Chen, J. SPOC-based flipped learning model applied in interpreting teaching. *International Journal Of Emerging Technologies In Learning (iJET).* **15**, 4-13 (2020)
[10] Sun, H. A SPOC teaching mode of college english translation based on” Rain Classroom”. *International Journal Of Emerging Technologies In Learning (iJET).* **14**, 182-193 (2019)
[11] Zhang, L., Xuan, Y. & And, Z. and application of spoc-based flipped classroom teaching mode in installation engineering cost curriculum based on obe concept. *Computer Applications In Engineering Education.* **28**, 1503-1519 (2020)
[12] Law, L., Hafiz, M., Kwong, T. & Eva Wong, .. Enhancing SPOC-flipped classroom learning by using student-centred mobile learning tools. *Emerging Technologies And Pedagogies In The Curriculum.* pp. 315-333 (2020)
[13] An, X. & Qu, C. hierarchical learning model based on deep learning and its application in a SPOC and flipped classroom. *International Journal Of Emerging Technologies In Learning (iJET).* **16**, 76-93 (2021)
[14] Kastrati, Z., Kurti, A. & Hagelb”ack, J. . *The Effect Of A Flipped Classroom In A SPOC: Students' Perceptions And Attitudes[C]//Proceedings Of The 2019 11th International Conference On Education Technology And Computers.* pp. 246-249 (2019)
[15] Gitinabard, N., Xu, Y., Heckman, S. & Lynch, C. How widely can prediction models be generalized? Performance prediction in blended courses. *IEEE Transactions On Learning Technologies.* **12**, 184-197 (2019)
[16] Liu, Q. & Huang, Z. Yin Y,Hui Xiong,Yu Su, Guoping Hu. *Ekt: Exercise-aware Knowledge Tracing For Student Performance Prediction.* **33**, 100-115 (2019)
[17] Chen, X., Zhong, D., Liu, Y., Li, Y., Liu, S., Deng, N. & Others Auto-generating examination paper based on genetic algorithms[C]//Advances on P2P. *Parallel.* pp. 887-893 (2020)
[18] Soufiane, O., Class, M. & Learning[j], B. Higher Education. (2021)
[19] Zhou, Y. Exploration of the Implementation of SPOC Blended Learning Model in the Teaching System of Choral Conducting[J]. *Curriculum And Teaching Methodology.* **6**, 51-57 (2023)
[20] Ramírez-Donoso, L., Pérez-Sanagustín, M., Neyem, A. & Others Fostering the use of online learning resources: results of using a mobile collaboration tool based on gamification in a blended course[J]. *Interactive Learning Environments.* **31**, 1564-1578 (2023)

Edited by: Mudasir Mohd
Special issue on: Scalable Computing in Online and Blended Learning Environments: Challenges and Solutions
Received: May 16, 2023
Accepted: Oct 7, 2023