The Influence of the Student's Online Learning Behaviors on the Learning Performance

Chunying Li[1], Junjie Yao[1], Zhikang Tang[1(✉)] , Yong Tang[2], and Yanchun Zhang[3,4]

[1] School of Computer Science, Guangdong Polytechnic Normal University,
Guangzhou 510665, China
fzutang@126.com
[2] School of Computer, South China Normal University, Guangzhou 510631, China
[3] Cyberspace Institute of Advanced Technology,
Guangzhou University, Cuangzhou 510006, China
[4] New Cyber Research Department, Peng Cheng Laboratory, Shenzhen 518055, China

Abstract. The emergence of online learning platforms means learners have a variety of learning behavior patterns. Many studies have found that there is a certain correlation between online learning behavior and learning performance. To better optimize the function of an online learning platform in hybrid teaching mode and further improve the quality of teaching and learning, this paper takes the 5y online learning platform as the target scene, and uses the online learning behavior data of 2205 learners and final exam score data as the breakthrough point of learning analytics. Through factor analysis on the behavior data of 13 measurement indicators of learners, this paper uses multiple linear regression model to analyze the correlation between learners' online learning behavior and their final exam scores. The research found that the final examination results of learners are obviously positively correlated with the basic question factors and comprehensive question factors. Therefore, teachers and students who use 5y platform should focus on the use of knowledge point tests and unit tests to improve the quality of teaching and learning within the limited class time.

Keywords: Online learning · Learning analytics · Linear regression

1 Research Background

The Online and Offline hybrid teaching model can stimulate the learning initiative and motivation of the learners. The 7*24 h online learning service in the hybrid teaching model can satisfy the seamless online learning demands of diverse levels and types of learners. Therefore, in-depth analysis of the impact of learner online learning behavior on learner performance can further improve the quality of teaching and learning in the mixed teaching model, as well as provide specific and operable opinions and suggestions for online learning platform developers to improve platform functions. 5y Platform[1] is an online learning platform independently developed by the Teaching and Examination

---

[1] http://5ystudy.gdoa.net/.

Management Center of Guangdong Provincial Institutions of Higher Education. It provides online learning services in various teaching modalities to over 100 institutions in Guangdong, HeBei, Guangxi, and Fujian, and has had a positive social influence. Since 2017, the course of Computer Application Basics as a public fundamental course for non-computer specialty in Guangdong Polytechnic Normal University, has cooperated with the 5y Platform and has achieved good teaching results for mixed teaching, and separation of teaching and examination. However, it has been discovered that a considerable proportion of the course learners did not achieve ideal results in the final examination. Therefore, this paper takes the online learning behavior data and the final exam score data of the course of "Computer Application Foundation" as the starting point for learning analytics, analyzes the impact of test behavior on the end-of-term learning performance, and aims to provide practical opinions and suggestions on the improvement of teaching quality and the system optimization of platform functions in the mixed teaching mode of the Computer Application Foundation course for our school and other universities using the 5y platform.

2 Related Research

2.1 Introduction to Learning Analytics

Learning analytics has gotten a lot of interest in recent years as a study subject for extracting useful information from educational data. The concept of learning analytics was officially introduced in 2011 with the help of "measuring, collecting, analyzing and reporting data on learners and their learning environments to understand and optimize learning and the environment in which it occurs" [1]. Muldner believes that learning analytics is conducive to the self-monitoring of learners' learning status and learning activities, improving their motivation and enabling them to have a positive emotional experience [2]. Han et al. proposed a review framework of learning analytics, including concepts and overview, composition and model, technical system, organization and evaluation [3]. G. Siemens et al. analyzed the value of learning analytics based on large data sets for education, including guiding the reform of higher education and promoting teaching, etc. [4].

2.2 Research on the Relationship Between Online Learning Behavior and Learning Performance

With the rapid development of online education, the factors that affect learners' learning performance and the relationship between learning behavior and their learning performance are scientific issues worth investigating. Song Jia et al. developed a multiple linear regression equation to better understand the impact mechanism of online learning institutions, communication frequency, communication time, and communication mode on in-depth learning [5]. Shen et al. constructed a performance evaluation model of online learning behavior and online learning through stepwise regression analysis of learning behavior data on the school online platform [6], Research by Liu et al. shows that learning analytics and personalized learning resource recommendation set up a personalized

learning path for learners, which helps to increase learners' enthusiasm for participating in learning activities and improve their academic performance [7]. Liu et al. have shown the cognitive input in online learning input. There is a significant positive correlation between emotional input and social input and learning performance [8].

2.3   Other Related Learning Analytics Methods

In recent years, some new methods have also been widely applied in learning analytics in addition to the traditional methods of educational research. For example, multimodal learning analytics is a new direction formed by the intersection of multimodal interaction, learning science, machine learning and other fields, which uses multimodal data to analyze learning behavior in complex environments to optimize the learning experience [9]. Kent et al. used social network analysis to assess the balance between the interactive benefits and the cost of coordination of the learner community [10]. Shen used a variety of intelligent algorithms such as artificial neural network and ant colony probability recommendation to develop a personalized learning path recommendation for users [11]. Karthikeyan et al. used basic information and behavioral performance data of learners' learning to predict academic performance and assess learner performance through the Naive Bayesian and J48 classifiers [12].

3   Platform Functions and Research Samples

The 5y platform provides the supporting video, exercises, tests and exams for the course, as well as the functions of learning notes, learning statistics, learning groups and discussion areas. Besides having the functions that most MOOCs have, the 5y platform's greatest feature is the ability to score subjective and objective topics in the course test, such as Word typesetting, Excel statistical analysis, etc. It greatly frees up the time for teachers to check/correct their homework and reduces the rate of misjudgment in examination corrections. 5y Platform roles mainly include four types: faculty, administrator, teacher and learner. The faculty and administrators are used internally by platform data maintenance and function maintenance personnel. The teacher performs basic classroom teaching functions such as customizing test papers, job publishing, notification publishing, and interactive communication. The learner side includes functions such as video learning, knowledge point testing, unit testing, comprehensive testing and interactive classroom communication. At any time, the learner can check his own learning results and data information such as the rankings of the class score, and then adjust learning strategies in real time according to his own situation. With the help of 5y platform, this paper collects 133297 online learning behavior data and their final exam results data from 2236 students of the Computer Application Basics course in the first semester of Guangdong Polytechnic Normal University from 2020 to 2021. The data set is analyzed by statistical analysis. Predict the final exam results of the learners according to the analysis results.

4   Research Results

4.1   Data Preprocessing

Python 3.7 was used in this study for data preprocessing, which include desensitization, deletion of duplicate and abnormal records, missing value processing, and filling in part of the dimension outliers with the mean or median of the dimension. 2205 learners and 131917 valid data were obtained after pretreatment. The data will be analyzed for learning analytics.

4.2   Learning Analytics

The research mainly includes descriptive statistics of end-term performance, and the impact of online behaviors on learning performance. The impact of learners' online behavior on learning performance is analyzed by factor analysis and multiple linear regression by selecting relevant online behavior indicators. The analysis software is SPSS 23.

4.2.1   Descriptive Statistics

Collect the final exam scores of 2205 learners in this semester. The lowest score is 7, the highest score is 97, the average score is 74.9, and the median score is 80. 318 of them failed, and the failure rate on the exam is 14.4%. There 325 (14.7%) students scored 60 to 69, 449 (20.3%) students scored 70 to 79, 766 (34.7%) students scored 80 to 89, and 347 (15.7%) students scored 90 or more. Most of the students' final examination results are focused on more than 70 points, which indicates that the mixed teaching mode has achieved good teaching quality in general.
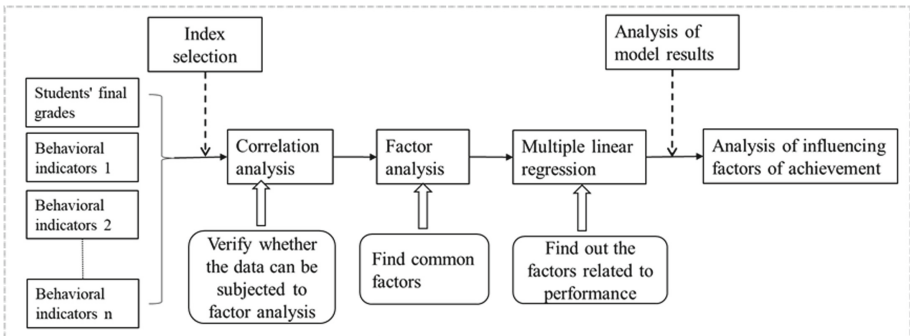

Fig. 1. The process of influencing factors analysis model for achievement

4.2.2   The Effect of Learners' Online Behavior on Learning Performance

The process of influencing factors of learners' online behavior on learning performance is shown in Fig. 1. Firstly, the appropriate online behavior indicators are selected. Then,

the correlation analysis, factor analysis and multiple linear regression are carried out between the selected behavior indicators and the learners' end-of-term performance in turn. Finally, the results of the multiple linear regression model are analyzed to find the main factors affecting the performance.

13 representative online behaviors are selected and named based on the characteristics of the 5y platform and the valid data generated by the platform's learners. Among them, the number of platform tests represents the number of tests performed on the platform, as detailed in Table 1.

Table 1. Variables and their meanings

Variable	Variable meaning	Variable	Variable meaning
A1	Number of unit tests	A8	Average score on comprehensive test
A2	Unit test average score	A9	Average video progress
A3	Number of intensive trainings	A10	Number of video learning
A4	Average score of intensive training	A11	Number of platform tests
A5	Number of knowledge point tests	A12	Number of learning notes
A6	Average score of knowledge point test	A13	Number of comment statements
A7	Number of comprehensive tests		

4.2.2.1 Correlation Test

Correlation analysis is the examination of two or more variable elements for correlation in order to determine the degree of correlation between two variables. Through Pearson correlation analysis of the variables, the correlation coefficient between the unit test average score and the number of unit tests is 0.681, the correlation coefficient between the number of videos and the average progress of video is 0.651, and the correlation coefficient between the number of platform tests and the number of knowledge points learned is 0.885. These correlation coefficients are more significant. This indicates that there is a strong correlation between these variables, and cannot be used directly for multiple linear regression. Therefore, consider the factor analysis of these data first.

4.2.2.2 KMO and Bartley Test

Factor analysis is to extract variables with some correlation into fewer factors, use these factors to represent the original variables, and also classify the variables according to the factors. Its greatest advantage is that the new factors can be named and interpreted so that they can be interpreted. Before factor analysis, *KMO* and *Bartlett* tests are performed on the selected variables to determine whether the selected independent variables are suitable for factor analysis. The calculation formulas for *KMO* statistics are as follows:

$$KMO = \frac{\sum\sum_{i\neq j} r_{ij}^2}{\sum\sum_{i\neq j} r_{ij}^2 + \sum\sum_{i\neq j} \beta_{ij}^2} \tag{1}$$

In the Eq. (1), *R* is the correlation coefficient. *β* For the partial correlation coefficient. The *KMO* is between 0 and 1, the closer to 1, the stronger the correlation between variables, the weaker the partial correlation, and the better the effect of factor analysis. As shown in Table 2, *KMO* statistic 0.667, *KMO* above 0.6 can be used for factor analysis [13], The *Bartlett* test significance level is less than 0.01, indicating that the selected sample data meet the requirements of factor analysis.

Table 2. KMO and Bartlett test

Number of *KMO* Sampling Appropriateness Quantities		.668
Bartlett sphericity test	Approximate Chi Square	10841.457
	Freedom	78
	Significance	.000

4.2.2.3 Calculate Eigenvalue and Variance Contribution Ratio

The characteristic values of each principal component factor obtained from the online learning behavior indicators selected in this paper are 4, and the cumulative contribution rate of variance of the four factors has reached 64.775%. This shows that the extracted four common factors can better explain most of the 14 selected learning behavior indicators. Therefore, the number of common factors is determined to be 4, and they are named *F1*, *F2*, *F3*, and *F4*. The explanatory rate of factor *F1* is 21.633%, which is higher than other factors. It is the first factor that learners' online behavior affects their performance.

4.2.2.4 Refining Analysis Results

Factor rotation using the maximum variance orthogonal rotation (Varimax) method improves the interpretability of the common factor. After five iterations, the matrix converges after 5 iterations.

The factor load factor of 12 variables in the rotated factor load matrix is greater than 0.5, which makes the analysis better. Horizontally, the number of intensive training sessions A3 does not belong to any dimension, so it is an invalid variable and is deleted. The first common factor has a large load on the number of platform tests, the number of knowledge point tests, the average score of intensive training, and the average score of knowledge point tests, which can be named as the basic question factor. The second factor has a large load on the average score of comprehensive tests, the number of unit tests, the number of comprehensive tests, and the average score of unit tests. It can be named as the comprehensive question factor. The third factor has a large load on the average video progress and the number of video learning, which can be named the

video viewing factor. The fourth factor has a large load on the number of comments and learning notes, which can be named learning activity factor.

4.2.2.5 Calculating Factor Score

The factor score and the final reflection of the factor analysis. By calculating the factor score, we can know the scores of the 13 selected learning behavior variables in the four extracted common factors, and analyze the end-term performance level of each variable in the common factor according to the results, as shown in Eq. (2):

$$
\begin{cases}
F1 = 0.003A1 + 0.120A2 + 0.119A3 + 0.285A4 + 0.290A5 \\
\quad +0.303A6 - 0.099A7 - 0.040A8 - 0.079A9 \\
\quad +0.002A10 + 0.304A11 - 0.003A12 + 0.017A13 \\
F2 = 0.322A1 + 0.215A2 + 0.064A3 - 0.059A4 + 0.021A5 \\
\quad -0.189A6 + 0.328A7 + 0.351A8 + 0.030A9 \\
\quad +0.050A10 + 0.049A11 + 0.022A12 + 0.007A13 \\
F3 = -0.095A1 - 0.200A2 + 0.049A3 - 0.102A4 + 0.042A5 \\
\quad -0.175A6 + 0.212A7 + 0.071A8 + 0.521A9 \\
\quad +0.481A10 - 0.003A11 - 0.022A13 \\
F4 = -0.014A1 - 0.055A2 + 0.068A3 + 0.027A4 - 0.003A5 \\
\quad -0.030A6 + 0.056A7 + 0.020A8 - 0.035A9 \\
\quad +0.001A10 - 0.037A11 + 0.649A12 + 0.649A13
\end{cases}
\tag{2}
$$

4.2.2.6 Multiple Linear Regression

Multivariate linear regression is a method of studying the relationship between a dependent variable and multiple independent variables, and it is used to explain the linear relationship between the dependent variable and other independent variables. This section performs multivariate linear regression with the four principal component factors as independent variables and the results as dependent variables to get the following regression models, as shown in Eq. (3):

$$
score = 74.974 + 5.695F1 + 5.946F2
\tag{3}
$$

The *R-Square* of the model is 0.279, which indicates that the model independent variable can explain 27.9% of the dependent variable change, and the *VIF* value is less than 5. This indicates that there is no multiple collinearity among independent variables, and the data residuals follow the normal distribution, indicating that the model is essentially valid. From this model, we can see that the basic factor *F1* and the comprehensive factor *F2* in the principal component factor have a positive influence on the results.

5 Research Conclusions and Recommendations

This study analyzed the online behavior of four different types of tests as well as 13 representative online behaviors. Based on the analysis of the learners' online learning behavior data on this platform and the construction of a multiple linear regression model,

it is found that the basic and comprehensive problem factors have a positive impact on performance, while the video factors and learning activity factors have no direct impact on performance. In view of this conclusion, from the point of view of improving the final examination results, it is suggested that the learners should spend more time and energy on the test questions, and try to ensure the correct rate of the test, rather than pursuing the number of questions. Teachers should guide learners to complete more knowledge point tests and unit tests based on the learners' actual situation in order to improve the learning effect under the premise of limited hours and time for learners. For 5y platform, there is no significant improvement in learning performance for video viewing factor and learning activity factor. One reason is that the data of video viewing factor and learning activity factor are too sparse and not representative. On the other hand, the course developer should improve the video in the course to attract the learning interest of the learners. Although this paper only performed an in-depth analysis on the data of the students enrolled in the course of Computer Application Foundation of Guangdong Normal University for one semester, the selected data are representative in Guangdong Polytechnic Normal University and other applied for undergraduate colleges and universities. Therefore, the results of the study analysis have sufficient reference and practical significance. It provides a relevant reference for the next stage of 5y platform function improvement and the improvement of teaching quality of "Computer Application Foundation" course under the mixed teaching mode.

References

1. LAK.: Shaping the future of the field. https://lak20.solaresearch.org/ (2020)
2. Muldner, K., Wixon, M., Rai, D., et al.: Exploring the impact of a learning dashboard on student affect. In: Int. Conf. Artificial Intelligence in Education, pp. 307–317. Springer International Publishing (2015). https://doi.org/10.1007/978-3-319-19773-9_31
3. Han, X.B., Huang, Y., Ma, J., et al.: A systematic review of learning analysis: review, identification and prospect. Research On Education Tsinghua University **38**(03), 41–51+124 (2017)
4. Siemens, G., Long, P.: Penetrating the fog: analytics in learning and education. EDUCAUSE Review **46**(5), 30 (2011)
5. Song, J., Feng, J.B., Qu, K.C.: Research on the influence of teacher-student interaction on deep learning in online teaching. China Educ. Technol. **11**, 60–66 (2020)
6. Shen, X.Y., Liu, M.C., Wu, J.W., et al.: Research on MOOC learners' online learning behavior and learning performance evaluation model. Distance Education in China (10), 1–8+76 (2020)
7. Liu, M., Zheng, M.Y.: Learning analysis and personalized resource recommendation in the view of intelligent education. China Educ. Technol. **09**, 38–47 (2019)
8. Liu, F.H., Yi, X.T.: Analysis model construction and application research of online learning input. E-education Research **42**(09), 69–75 (2021)
9. Mou, Z.J.: Multimodal learning analysis: learning analysis and analysis of new growth points. E-education Res. **41**(05), 27–32+51 (2020)
10. Kent, C.,Cukurova, M.: Investigating collaboration as a process with theory- driven learning analytics. **7**(1), 59–71 (2020)
11. Shen, Y.F.: A personalized learning path recommendation model based on multiple intelligent algorithms. China Educ. Technol. **11**, 66–72 (2019)

12. Karthikeyan, V.G., Thangaraj, P., Karthik, S.: Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation. Soft. Comput. **24**(24), 18477–18487 (2020)
13. Wu, M.L.: SPSS statistics application practice: Questionnaire analysis and application statistics. Science Press, Beijing (2003)