Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Media Learning Behavior Analysis Research Report Study on Learning Duration, Course Completion Rate, and Grade Distribution

Yunyi Wang

2941308283@qq.com

Fuzhou Melbourne Polytechnic

Article

Keywords: Online learning behavior, learning duration, course completion rate, score distribution, learning analysis, educational data mining

Posted Date: August 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10202959/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Media Learning Behavior Analysis Research Report Study on Learning Duration, Course Completion Rate, and Grade Distribution

Yunyi Wang [1] *

[1] Department of International Business, Fuzhou Melbourne Polytechnic, Fuzhou, 350108, Fujian, China

***Corresponding author, Email: [2941308283@qq.com]**

June 2026

ABSTRACT

With the in-depth promotion of educational informatization, online learning has become an important form of education, but it faces the problems of low pass rate and high dropout rate. This study focuses on the analysis of online learning behavior, examines the learning time, course completion rate, score distribution, and explores the optimization direction of platform effectiveness. Adopt the combination of literature review, empirical analysis and case study to build a theoretical framework, establish a behavioral characteristic engineering that includes learning time, course completion rate, performance distribution and interaction characteristics, and explore the construction of prediction models such as classification, integrated learning, deep learning and clustering. The study found that there is a positive correlation between learning duration, course completion rate and score distribution, which is an important indicator for measuring the effect of online learning; behavior analysis technology has developed into a deep learning model, and the prediction accuracy rate reaches more than 95%; learners show different behavior patterns such as high-input, low-input, jumping, strategic, etc.; in Online teaching design has an important impact on learning behavior. Based on the conclusion, optimization suggestions are put forward from the three levels of platform, teaching design and teaching management: the platform should strengthen the ability of data collection and analysis, and establish a learning early warning and personalized recommendation mechanism; teaching design should create situations, enrich resources, strengthen interaction, and improve feedback; teaching management should strengthen behavior supervision and establish early warning machines. System, carry out teacher training, and establish a learning incentive mechanism. This research provides a scientific basis for personalized recommendation and curriculum design of online education platforms, which is of great value for improving the quality of online education.

Key words: Online learning behavior, learning duration, course completion rate, score distribution, learning analysis, educational data mining

Table

1. Introduction 1

2. Literature review 2

2.1 The theoretical basis and data source of online learning behavior analysis 2
2.2 Characteristics and analysis methods of learning behavior 2
2.3 Relevant research on learning effect prediction 3

3. Theoretical framework 4

3.1 The concept of theory 4
3.1.1 The behavior of online learning 4
3.1.2 Development of learning analysis and educational data 4
3.2 The basis of relevant theories 4
3.2.1 Learn the theory of analytical technology 4
3.2.2 Educational Data Mining Theory 5
4.1 Data collection and processing 6
4.2 Characteristics of behavior 6
4.3 Predictive model construction 6
4.3.1 Classification algorithm 6
4.3.2 Integrated learning algorithm 7
4.3.3 Deep learning algorithm 7
4.4 Cluster analysis method 7

5. Case analysis 8

5.1 Descriptive statistical analysis of learning behavior 8
5.1.1 The relationship between study time and grades 8
5.1.2 Analysis of the characteristics of major online learning platforms 10
5.2 Analysis of learning time and learning frequency 12
5.2.1 Distribution of learning time 12
5.2.2 Distribution of learning frequency 13
5.3 Course completion rate and score distribution analysis 15
5.4 Analysis of the market size of online education 16
5.5 Evidence of the learning effect based on the prediction model 17
5.6 Cluster analysis of learner group characteristics 18
5.7 Online teaching design and behavioral intervention revelation 18

6. Conclusion 20

Declarations 22

Special Reminder 22
Competing interests 22

Ethics approval and consent to participate 22
Availability of data and materials 22
Author contributions 23
Funding 23

7. References 24

1. Introduction

With the rapid development of computers, the Internet and digital media, online learning has become an emerging way of learning, which is profoundly changing the pattern of traditional education[1]. The construction of educational informatization has accelerated, and various teaching methods combined with informatization have emerged one after another, and educational big data has been formed[7]. Online learning provides learners with a free and convenient way to learn, but it also faces the problems of low pass rate and high dropout rate[8]. Therefore, by analyzing a large amount of learning behavior data generated by learners in the process of online learning and exploring learning laws, it has important theoretical and practical value for improving the quality of online education and optimizing teaching design.

This study focuses on the analysis of online learning behavior, focusing on the three core dimensions of learning time, course completion rate and score distribution, and exploring the effectiveness of online learning platforms and its optimization direction. Through in-depth exploration of learners' behavioral characteristics, it aims to provide a scientific basis for personalized learning recommendations and curriculum design of online education platforms and provide guidance for teaching practice.

2. Literature review

2.1 The theoretical basis and data source of online learning behavior analysis

Online learning behavior analysis relies on educational big data learning analysis technology. The educational informatization process produces a large amount of data. The role of data is gradually valued, and the data format is also more standardized[7]. Network data information builds a complex network. Students' learning data reflects learning behavior, and the learning results are affected by it[1] .

The data source covers various online teaching platforms, such as the Superstar Learning MOOC platform, etc. These platforms record the completion of learning log videos and watch data tasks. Gao Tianding Linhua[6] uses Python crawler technology to obtain students' online learning data and provide a data basis for behavior analysis. Wang Jiaoyu and others[4] based on the Superstar Learning Platform, in-depth analysis of students' behavior in the course teaching process.

2.2 Characteristics and analysis methods of learning behavior

The mining method of learning behavioral characteristics has undergone development and evolution, from statistical analysis to machine learning to deep learning. Early research mostly adopts descriptive statistical methods. Lu Liyuan and others use content analysis to analyze the online learning logs of 193 learners, and use ROST CM 6.0 for word frequency analysis, semantics Network analysis[5].

In the process of technological development, machine learning methods were gradually introduced. Dai Hong and others[9] used integrated learning methods to predict academic performance. Wu Wenkai[8] improved the Stacking algorithm to build a more effective classification model.

In recent years, the deep learning method has begun to be applied. An Menglei and others have adopted the CNN-LSTM prediction model based on the attention mechanism[3]. The student learning behavior data is normalized and processed into the input model after correlation

analysis. The attention mechanism distinguishes the importance of information, and the accuracy and reliability of the model are improved.

2.3 Relevant research on learning effect prediction

Online learning behavior analysis is an important application direction, and learning effect prediction is one of them. In the early research, the multivariate regression algorithm was widely adopted. Gao Tian and Ding Linhua[6] calculated and analyzed students' learning behavior data from multiple dimensions, and the prediction of students' learning effect was based on this.

At present, integrated learning deep learning has become the mainstream method. Wu Wenkai[8] adopts the improved Stacking algorithm, and the classification and prediction of the public data set of the ed X network learning platform is completed. The accuracy rate F1 value is improved by 2.46% 2.11% 1.97%. The experimental results show this point. , An Menglei and others[3] proposed a performance prediction model, and the accuracy rate of students' performance prediction reached 95.69%. Meng Jiana and others[1] used classifiers to predict whether students can be judged through learning, and the learning effect estimation provides students with.

3. Theoretical framework

3.1 The concept of theory

3.1.1 The behavior of online learning

Online learning behavior refers to the learners' recordable behavioral activities in the online learning environment to complete the learning goals, including course access, video viewing, online test homework submission, discussion and interaction, etc. These behavioral data truly reflect the learner's learning process, and the learning investment degree learning strategy is also reflected in this[8].

3.1.2 Development of learning analysis and educational data

Learning analysis refers to measuring, collecting, analyzing and reporting learner data to understand and optimize the learning environment. Educational data mining extracts information knowledge from educational data. The two are closely related and have different focuses. Learning analysis focuses on real-time feedback intervention, while educational data mining focuses on model discovery and prediction[10].

3.2 The basis of relevant theories

3.2.1 Learn the theory of analytical technology

The theory of learning analysis technology emphasizes technical means to standardize the information teaching process, collect students' real and detailed learning data, analyze data in real time, and use targeted benign intervention to improve students' learning effectiveness[7]. The theory believes that sleeping data comes to life, big data technology is introduced in the education field, and teachers become mentors through reverse classrooms to help students become better.

3.2.2 Educational Data Mining Theory

The theory of educational data mining is based on the machine learning statistical data mining method. The hidden learning mode law is found through in-depth mining of educational data. This method emphasizes the construction and verification of data preprocessing characteristic engineering models[3],[8], and the optimization of personalized learning and teaching is therefore supported by data.

4. Research method

4.1 Data collection and processing

Data collection mainly relies on the background log database of the online learning platform. The collection content includes learners' basic information course access records, video viewing time, number of task point completion, online test score discussion participation, etc.

Data preprocessing covers data cleaning, missing value processing, abnormal value detection, data normalization, An Menglei, etc. conducts students' learning behavior data through correlation analysis[3], normalization processing, and data quality is ensured in the input model.

4.2 Characteristics of behavior

Behavioral feature engineering is the process of transforming raw data into meaningful features. The main characteristics include:

(1) Characteristics of learning time: total learning time, single learning time, learning frequency, etc.;

(2) Characteristics of course completion rate: video viewing completion rate, homework completion rate, task completion rate, etc.;

(3) Characteristics of the distribution of grades: test scores, homework scores, final grades, etc.;

(4) Characteristics of interaction: the number of discussions, the number of questions, the number of answers, etc.

4.3 Predictive model construction

4.3.1 Classification algorithm

Classification algorithms predict whether learners can pass the course, logical regression decision tree support vector machines, etc. are commonly used. Meng Jiana and others[1] use

classifiers to predict students' learning results. Estimated learning results provide reference for students, and at the same time provide a scientific basis for personalized learning recommended course design on online education platforms.

4.3.2 Integrated learning algorithm

The integrated learning algorithm combines multiple base learners, and the prediction performance can be improved. Wu Wenkai[8] carefully studied the stacking algorithm theory. Based on the performance advantages, he used it as a learner category prediction algorithm to improve the hierarchy, improve the representation of data feature attributes, improve the combination strategy, and improve the classification algorithm. The experimental results show that the improved Stacking algorithm has performed well in terms of evaluation indicators such as accuracy and accuracy and accuracy F1 value.

4.3.3 Deep learning algorithm

Deep learning algorithms can automatically learn data characteristics. An Menglei and others[3] adopt the attention mechanism CNN-LSTM prediction model. The attention mechanism distinguishes the importance of information, and the accuracy and reliability of the model are improved. The experimental results show that the accuracy of student performance prediction reaches 95.69%.

4.4 Cluster analysis method

Cluster analysis is used to discover the characteristics of learner groups. An Menglei and others[3] adopt the improved K-means clustering algorithm based on Canopy. The prediction value of student learning behavior characteristics is pre-culustered by the Canopy algorithm, and the pre-culeration results are clustered in combination with K-means to achieve effective classification of learner groups.

5. Case analysis

5.1 Descriptive statistical analysis of learning behavior

5.1.1 The relationship between study time and grades

Through descriptive statistical analysis of the online learning platform data, the basic behavior patterns of learners can be revealed. Wang Jiaoyu and others[4] analyze the teaching behavior of students' courses, study the performance effect of students' learning times and task point completion, and summarize the learning characteristics and hobbies of higher vocational students.

Zhang Ying and others[2] showed that the viewing rate of students from video one to video three decreased, and the rate of students who watched back also decreased. Learners' attention to the course content decreased with the progress of learning. Optimizing the teaching design to improve the continuous participation of learners needs to be considered.

The analysis results of Lu Liyuan et al.[5] pointed out that online learners pay attention to the common emotional understanding of learning experience, and the interactive design of the learning process is conducive to the occurrence of group belonging. More learners are willing to take the initiative to use resources for independent learning. Learners pay attention to teachers' feedback and evaluation. Moderate motivation is conducive to the continuous positive development of learning behavior.

According to the collected online learning platform data, the statistical results of learners' behavior are shown in Table 5-1-1 [1]:

Table 5-1-1 [1] Statistical table of learners' behavior [13]

Learning behavior	Mean value	Standard deviation	Min-value	Max-value	Range
Learning progress (%)	93.20	5.05	78.00	95.00	17.00
Duration of video learning (Min)	1346.50	8.28	1055.50	1725.80	670.30
Online homework (points)	20.58	4.25	15.88	23.70	7.82
Discuss the score (score)	3.50	3.17	2.20	4.50	2.30
Offline test results (points)	32.50	8.35	29.00	33.50	4.50

It can be seen from the above table that the average learning progress of learners has reached 93.2%, and the overall completion rate is relatively high. The average video learning time is 1346.5 minutes, which shows that the time investment is large. The online homework discussion score is relatively low, and there is still room for improvement in interactive participation. As shown in Figure 5-1-1 [1].

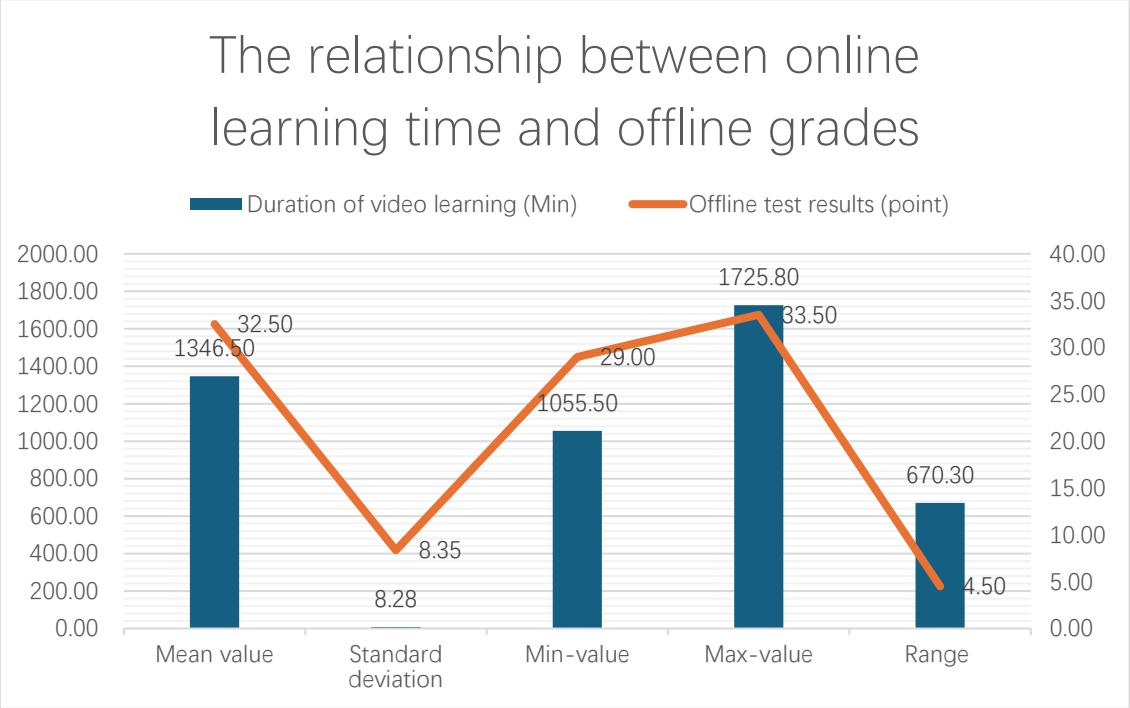


Figure 5-1-1 [1] Column diagram of the relationship between online learning time and offline performance [13]

It can be seen from Figure 5-1-1[1] that learners' investment in online learning is sufficient. However, the offline scores are low and the difference in scores between different people is relatively large, which shows that although learners "value" online learning, the knowledge absorption of learning is insufficient. This also shows that although people participate in online courses, there are different people who listen carefully. It also shows that people don't attach much importance to it. This also shows that learners still have a lot of room for improvement in knowledge absorption or attention to online courses.

5.1.2 Analysis of the characteristics of major online learning platforms

There are differences in course design and learning experience between different online learning platforms. The following are the average estimated duration statistics of major platforms: as shown in Table 5-1-2 [1].

Table 5-1-2 [1] Different online learning platforms in curriculum design and learning [15]

Platform	Times (H)	Duration range (H)
Coursera	20	25
edX	18	25
Udemy	5	3
LinkedIn Learning	4	/

According to the data, the duration of each course on academic platforms such as Coursera and edX is relatively long, with an average of 18H-20H, which is suitable for in-depth learning and for those learners who have enough time to learn. But at the same time, it also reflects that if learners do not have a large amount of study Time, so learning on these platforms is likely to sacrifice a lot of time in daily life because it takes a lot of time, which reduces the absorption of knowledge and makes the efficiency of learning less; while practical platform courses such as Udemy and LinkedIn Learning are shorter and more suitable. Fragmented learning is aimed at those learners who usually need to work or do not have long spare time. They are very suitable for these platforms to make learning efficient. As shown in Figure 5-1-2 [1].

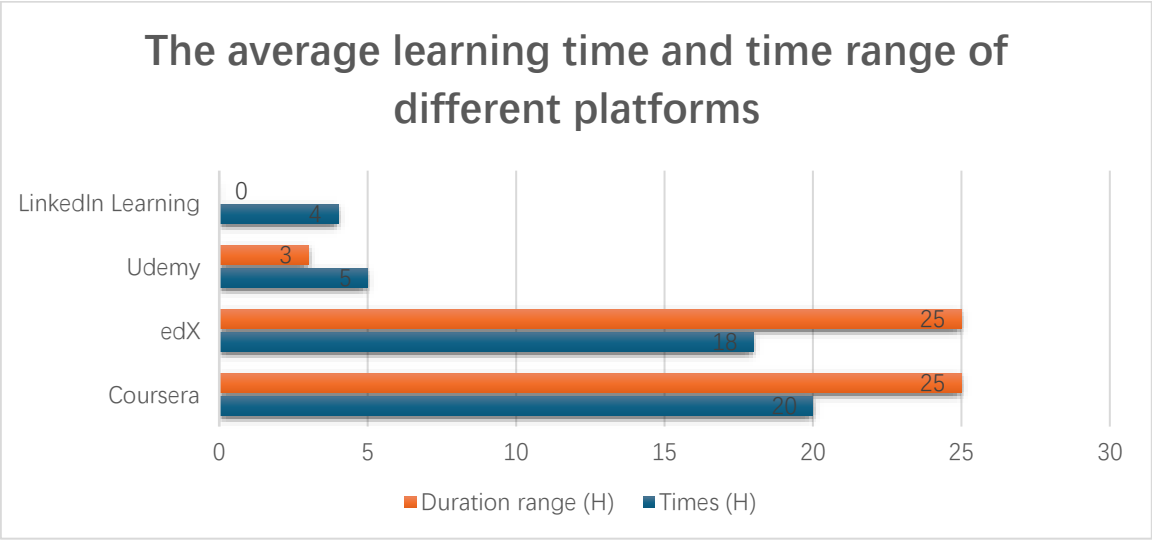


Figure 5-1-2 [1] Average learning time and duration range bar chart of different platforms [15]

Wang Jiaoyu and others deeply analyzed the behavior of students in the course teaching process, studied the number of students' learning times, the completion of task points, the performance effect and other contents[4], and summarized the learning characteristics and hobbies of senior vocational students.

Zhang Ying et al.[2] found that from video to video three, the student's viewing rate and reviewing rate decreased in turn, and the attention to the course content decreased with the progress of learning. The optimization of teaching design and whether to design a timing system is needed. The solution of these problems is very necessary to improve the attractiveness of online learning, which can be mentioned. High learners' continuous participation and improve the degree of absorption of knowledge.

The analysis results of Lu Liyuan et al.[5] point out that online learners pay attention to the learning experience, the sympathy between learners and teachers, and the perception of knowledge. The learning process of interactive design is conducive to the group's sense of belonging, and more learners are willing to take the initiative to use resources for independent learning. In addition, learners will also pay more attention to teacher feedback and teacher evaluation. Moderate motivation is conducive to learning behavior, and sustainable development is positive.

5.2 Analysis of learning time and learning frequency

5.2.1 Distribution of learning time

The distribution of learning time of online learning users in China in 2020 is shown in Table 5-2-1 [1]:

Table 5-2-1 [1] Distribution Table of Learning Time of Line Learning Users [12]

Times(Min)	Ratio(%)
30-60	60
60-90	28

Times(Min)	Ratio(%)
90-120	6.4
<30	4.8
>=120	0.8

Data shows that the vast majority of learners (60%) spend 30-60 minutes, 28% of learners spend 60-90 minutes, and only a few learners spend more than 90 minutes, which shows that learners have limited time to concentrate. There are still many learners who have very little fragmented time. Even if they have some fragmented learning time, they will not be fully used for learning, and may be attracted by other external factors. The real implementation in the time of learning may be stretched to the limit. As shown in Figure 5-2-1 [1].

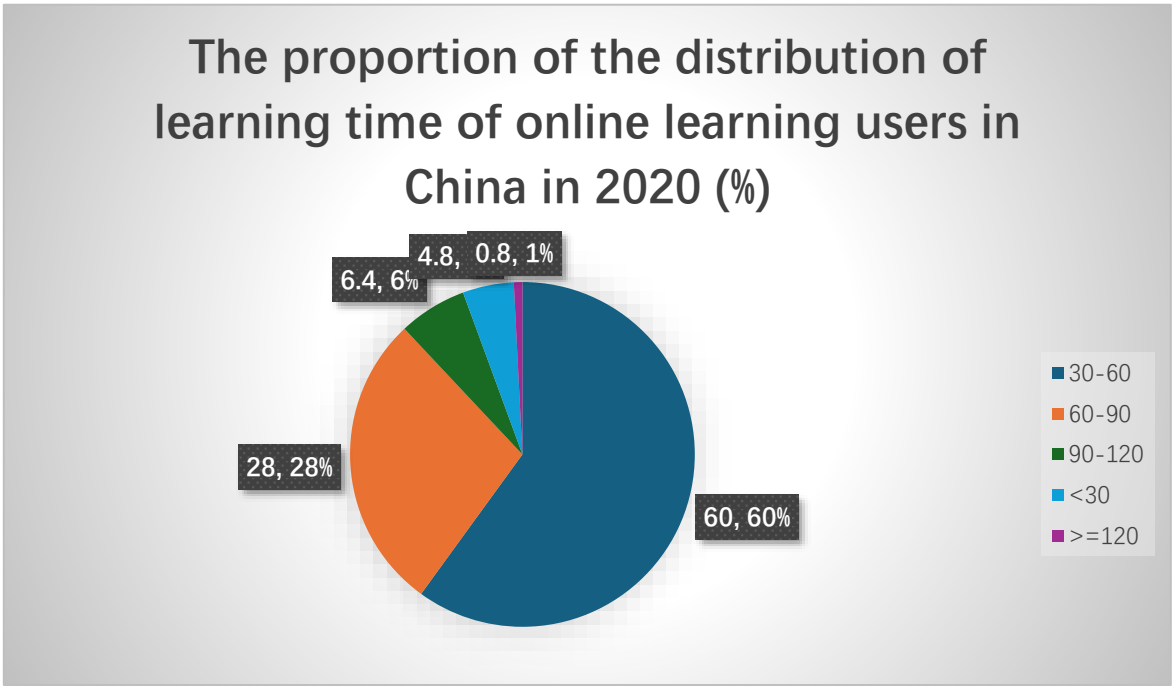


Figure 5-2-1 [1] Pie chart of the percentage of each score [12]

5.2.2 Distribution of learning frequency

The distribution of learning frequency of online learning users in China in 2020 is shown in the following table: as shown in Table 5-2-2 [1].

Table 5-2-2 [1] Distribution Table of Learning Frequency of Online Learning Users in China in 2020 [12]

Times	Ratio(%)
Several times a week	55.2
Once a week	23.2
Almost every day	11.2
Many times in a month	8.8
Many times a year or longer	1.6

More than half (55.2%) of learners study online several times a week, indicating that most learners still have good learning habits. They will seize their spare time to learn and improve themselves. However, there are also nearly 10% of learners who learn less frequently. They not only need to strengthen learning management, but also need to strengthen the fun and interactive experience of online courses, reduce unnecessary timing systems, and not treat learners as "slaves" to the course. This will not only make learners psychologically resistant, but also dispel the new learners' interest in learning. As shown in Figure 5-2-2 [1].

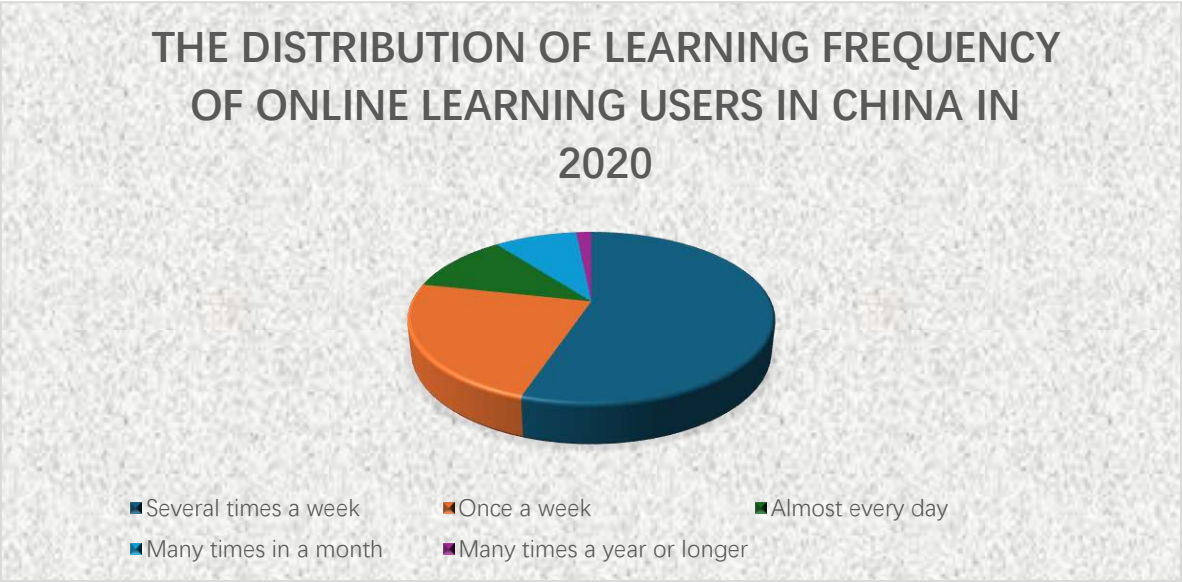


Figure 5-2-2 [1] Pie chart of learning frequency distribution of online learning users in China in 2020 [12]

5.3 Course completion rate and score distribution analysis

The distribution of learners' grades is shown in Table 5-3 [1]:

Table 5-3 [1] 2020 China Online Learning Learner Performance Distribution Table [14]

Score range	Ratio(%)	Distinguishing feature
0-20	40	Insufficient investment in learning
60-80	30	Intermediate level
80-100	10	Good level
>=100(Outstanding)	20	Excellent performance

Data shows that the score distribution shows a polarization trend: low-segments (0-20) account for 40%, medium-high-segments (60-80) account for 35%, and excellent segments (>=100) account for 15.63%, indicating that there are significant differences in the group of learners. Some learners can learn efficiently and absorb efficiently. The knowledge in the course is to improve one's knowledge at ordinary times, and to improve one's knowledge diligently, so the results reflected on paper are relatively high. However, some learners may not be focused enough on the course at ordinary times, or turn aside after listening to the online course. Without review, the knowledge has not yet "stabed firm". New knowledge pours into their minds like a flood, which makes the old knowledge not mastered and applied flexibly, and new knowledge comes. This is a low score on paper. As shown in Figure 5-3 [1].

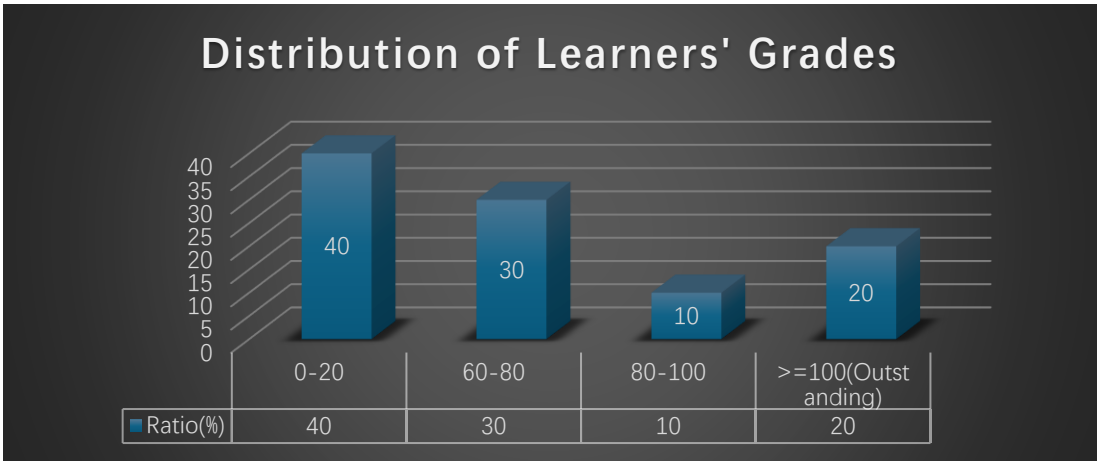


Figure 5-3 [1] Distribution of Learners' Grades [14]

5.4 Analysis of the market size of online education

The size and growth of China's online education market in 2016-2025 are shown in Table 5-4 [1]:

Table 5-4 [1] Table of China's Online Education Market Size and Growth in 2016-2025 [11]

Year	Scale (ten thousand people)	Growth rate (%)
2016	744	-
2017	981	31.9
2018	1388	41.5
2019	1892	36.3
2020	2606	37.7
2021	3557	36.5
2022	4692	31.9
2023	5980	27.5
2024	7433	24.3
2025	9066	22.0

Data shows that China's online education market will grow from 7.44 million in 2016 to 90.66 million in 2025, with an annual compound growth rate of more than 25% and strong development momentum. This not only shows that the platform in the market needs online learning, and there are many people who want to improve their own needs in fragmented time. At the same time, it is equally important to maintain learners' interest in learning and retain new customers on the platform and maintain old customers. Although the growth rate has slowed

down in recent years, the market size continues to expand, and the market outlook is still smooth. As shown in Figure 5-4 [1].

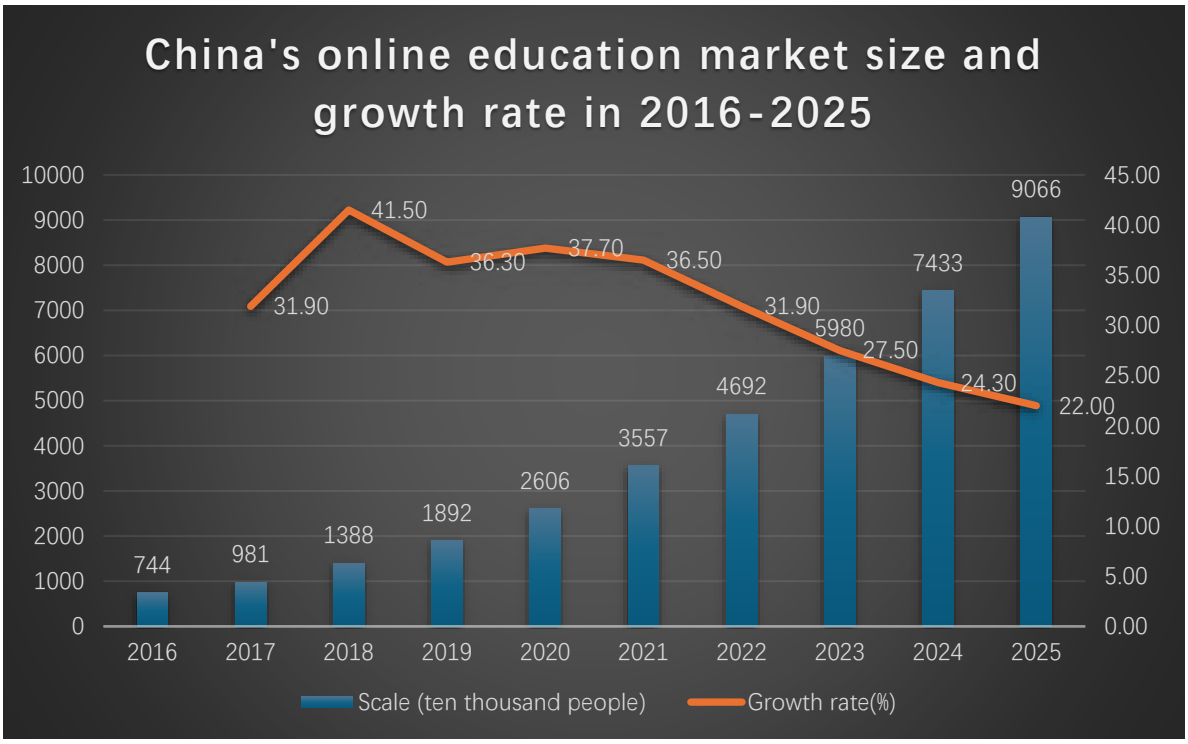


Figure 5-4 [1] Column chart of China's online education market size and growth in 2016-2025 [11]

Wang Bo[7] pointed out that the technology standardizes the information-based teaching process, collects students' real and detailed learning data, analyzes data for real-time feedback, adopts targeted and benign intervention to improve students' learning effectiveness, and teachers reverse the classroom to become mentors to help students become better talents.

5.5 Evidence of the learning effect based on the prediction model

Empirical research verifies the validity of the learning effect prediction model. Gao Tianding Linhua[6] adopts the multivariate regression algorithm to calculate and analyze students' learning behavior data from multiple dimensions, predict the learning effect, and give timely warning of students' online learning behavior.

As a result of the experiment of An Menglei and others, the accuracy of the performance prediction model reached 95.69%, and the learning effect prediction tool was provided with high accuracy[3].

Wu Wenkai's[8] experiment shows that the improved Stacking algorithm has improved the accuracy rate, accuracy rate and F1 value on the ed X data set by 2.46%, 2.11% and 1.97% respectively, verifying the effectiveness of integrated learning.

5.6 Cluster analysis of learner group characteristics

Through analysis, learners can be divided into different groups to implement teaching strategies for different groups of people. Wu Wenkai implements different teaching programs according to different types of learners, thus providing a scientific basis for the efficient teaching of online learning platforms[8].

The results of cluster analysis can usually find the following types of learners:

(1) High-quality people: long learning time, high completion rate, good grades;

(2) Low-quality people: short learning time, low completion rate, poor grades;

(3) Fluctuating people: medium learning time, unstable completion rate, and large fluctuations in grades;

(4) Highly efficient people: moderate learning time, high completion rate and excellent grades.

5.7 Online teaching design and behavioral intervention revelation

Based on the results of learning behavior analysis, it can provide important help for online teaching and behavioral intervention.

Lu Liyuan suggested that the design of online teaching should start from creating online learning situations, enriching online learning resources, creating good classroom interaction, and moderate text motivation, so as to improve the quality of online education [5].

The research of Zhang Ying and others shows that students' pre-class video learning behavior is an important factor affecting their adaptability to the flipped classroom. In teaching practice, the management and supervision of students' pre-class video learning behavior should be strengthened[2].

Wang Bo pointed out that using the technology to standardize the information teaching process, collect students' real and detailed learning data, provide real-time feedback through data analysis, and use targeted benign intervention to improve students' learning effectiveness. Teachers can really become mentors by reversing the classroom to help students become better[7].

6. Conclusion

Through the systematic analysis of online learning behavior, this study draws the following main conclusions:

An important indicator of online learning effect is the distribution of course completion rate of learning time. The three show a positive correlation. The longer the learning time, the higher the course completion rate, and the higher the score distribution, the higher the score segment. Online learning behavior analysis technology has developed from simple statistical analysis to complex deep learning models. The prediction accuracy has been continuously improved, and the attention mechanism has been integrated with learning and other technical applications, so that the accuracy of learning effect prediction has reached more than 95%. In the process of online learning, learners show different behavior patterns and group characteristics. Cluster analysis can effectively identify different types of learners and provide a basis for personalized teaching. The optimization of online teaching design has an important impact on learning behavior. The creation of good learning situations, enriching learning resources, strengthening interactive design, moderate incentives and other measures can effectively improve the participation and learning effect of learners.

Based on the above conclusions, the optimization suggestions are put forward as follows. In terms of online learning platform, the data collection and analysis ability needs to be improved, the learning analysis system needs to be established, and the behavior of learners can be monitored in real time. The prediction model is used for learning early warning, and people with learning difficulties can be found in time and provided with targeted help. Cluster analysis results guide personalized recommendations, different types of learners, suitable for resource path push. The platform function is optimized, the user experience is improved, and the technical threshold is lowered. In terms of teaching design, a good online learning situation is created, and learners' sense of belonging and participation is enhanced. Online learning resources are rich, and diversified and multi-level materials are provided. Classroom interaction design is strengthened, and learners' communication and collaboration are promoted. Reasonable design of learning tasks, video length control, segmented learning support provided.

Improve the feedback mechanism and establish a timely feedback and motivation for learners' behavior. In terms of teaching management, the supervision of learners' pre-class behavior management should be strengthened, and the completion of video learning tasks should be paid attention to. An early warning mechanism for learning is established, and those who have a short learning time and a low completion rate should intervene in time. Teacher training is carried out, and data analysis ability and online teaching ability are improved. The establishment of a learning incentive mechanism, certificate points and other ways to motivate learners to continue learning.

Declarations

Special Reminder

The literature cited in this report comes from CNKI, and the reference format may be slightly different. If you still need to rectify, please email me. Email: 2941308283@qq.com. Thank you very much.

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki. The need for ethics approval was waived by the Ethics Committee of the Ethics Committee of Fuzhou Melbourne Polytechnic. All methods were performed in accordance with the relevant guidelines and regulations of the National Health Commission of the People's Republic of China. Informed consent was obtained from all participants and/or their legal guardians.

Availability of data and materials

The datasets analysed during the current study are not publicly available due to ethical restrictions but are available from the corresponding author on reasonable request. Part of the data were derived from the following sources:

1. The market size data of China's online education sector were obtained from 199IT (Chinese Internet Data and Information Center), available at: https://www.199it.com/archives/1264397.html.

2. The user behaviour data of online learning were sourced from iiMedia Research (data.iimedia.cn), available at: https://www.iimedia.cn/c1061/69228.html.

The authors confirm that they have checked the terms of use of these source websites to ensure that the use and analysis of these data for this study are permitted.

Author contributions

Y.W. (Yunyi Wang) contributed to the conceptualization, data curation, formal analysis, investigation, methodology, project administration, validation, visualization, and writing of the original draft, as well as the review and editing of the manuscript. The author confirms sole responsibility for the following: study conception and design, data collection and analysis, interpretation of results, and manuscript preparation.

Funding

The authors declare that no funding was received for this research. No funds, grants or other support was received.

7. References

[1] Meng Jiana, Wang Zheng, Song Haiyu, etc. Analysis of students' learning behavior for online teaching platforms [J]. Higher Education Journal, 2025,11(27):91-94.

[2] Zhang Ying, Kong Jinge, Wang Yun. Analysis of students' pre-class video learning behavior and its impact on the flipped classroom learning experience [J]. China Higher Medical Education, 2025,(6):4-5+20.

[3] An Menglei, Han Meng, Ren Rong, etc. Online learning behavior analysis and performance prediction methods [J]. Information Technology and Informatization, 2024, (9):17-21.

[4] Wang Jiaoyu, Gao Zhiliang, Wang Xuan. Online learning behavior analysis research based on superstar learning platform [J]. Scientific Consulting, 2024, (14):108-111.

[5] Lu Liyuan, Liu Ganhong, Wang Runhua. Online teaching design optimization research based on online learning behavior analysis [J]. Modern Information Technology, 2024,8(6):173-177.

[6] Gao Tian, Ding Linhua. Research on learning effect prediction methods for online learning behavior analysis [J]. Computer Programming Skills and Maintenance, 2022, (7):36-38.

[7] Wang Bo. Research on online education learning behavior analysis technology based on big data [J]. Electronic World, 2020, (24):100-101+104.

[8] Wu Wenkai. Research on Network Learning Behavior Analysis and Prediction [D]. Anshan: Liaoning University of Science and Technology, 2020.

[9] Dai Hong, Wu Wenkai, Ren Ling, etc. Research on Behavior Analysis and Prediction of Online Learning [J]. Communication World, 2019,26(10): 28-29.

[10] Sun Yanlong. Online Learning Behavior Analysis Research Review [J]. Shandong Industrial Technology, 2017, (24):221-222.

[11] Internet Data Information Network-199IT 2016-2025 China's online education market size and growth rate (with original data table) [EB/OL]. https://www.199it.com/archives/1264397.html (accessed on 2026-04-01).

[12] Online learning industry data analysis of Aimedia.com: 60% of Chinese online learning users learn for 30-60 minutes [EB/OL]. https://www.iimedia.cn/c1061/69228.html (2026-04- 01 Visit).

[13] Educational scientific literature. Full document of Issue 4 of 2025 2025:93-94. htt ps://www.oacj.net/ upfile/files/2025%E4%B8%AD%E6%96%87/Z04/%E3%80%8A%E6% 9 5%99%E8%82%B2%E7%A7%91%E5%AD%A6%E6%96%87%E7%8C%AE%E3%80%8B2025%E5%B9%B4%E7%AC%AC4%E6%9C%9F%E5% 85 %A8%E6%9C%AC%E6%96%87%E4%BB%B6.pdf (2026-04-01 access).

[14] International Review of Research in Open and Distributed Learning Using SurvivalAnalysis to Identify Populations of Learners at Risk of Withdrawal: Conceptualization and Impact of Demographics Juan Antonio Martínez-Carrascal, Martin Hlosta and Teresa Sancho-Vinuesa,2023. https://doi.org/10.19173/irrodl.v24i1.5801 (2026-04-01 Visit).

[15] Исмагилова, А. Р., Исмагилова, В. С., & Морозова, Р. Ф. Роль и перспективы онлайн-обучения в современном образовании[C]// Педагогика и психология. 20 19. (Original work in Russian) https:// pdfs.semanticscholar.org/c2d1/15624b5ea6d05 e2c6d84305cc6d133374efa.p df?skipShowableCheck=true (2026-04-1 Visit).