

Research on Influencing Factors of Online Education under Epidemic Situation Based on Regression Analysis

Liyang Qu*

School of Data Science and Technology, Heilongjiang University, Heilongjiang, China

* Corresponding Author Email: yaoyu@cau.edu.cn

Abstract. Since the COVID-19 pandemic, large-scale online teaching has officially kicked off around the world. Carrying out online teaching is not only an emergency measure in the context of epidemic prevention and control but also an important opportunity to transform teaching thinking and promote teaching innovation. The study looked at students' responses to online classes in the Indian capital, Delhi. This study used regression analysis and some data visualization methods to conduct statistical analysis of the data. The results showed that students were basically satisfied with the effect of online learning during the pandemic, and the study found that "Time spent on Online class" and "Smartphone" are the main factors to evaluate students' learning effects. According to the influencing factors obtained from the results, the "Government - School - Student" model is proposed to promote the learning efficiency and quality of students, hoping to become an effective strategy to improve students' home learning.

Keywords: Covid-19; Online class; Regression; Visualization; Influencing factors.

1. Introduction

1.1 Background

Due to the outbreak of COVID-19, offline education cannot be carried out. Various educational institutions have launched online education, which has played a great role in promoting the development of online education. There is online education feedback in the market. This makes it difficult for online education institutions to systematically and regularly analyze the collected data.

1.2 Method and Target

Hundreds of millions of children and adolescents have been affected by school closure policies in countries affected by the pandemic. In March 2020, the Organization for Economic Cooperation and Development conducted a global survey on how education systems are responding to the pandemic and the challenges facing mass online education [1]. The post-epidemic age has presented both fresh opportunities and several obstacles for teaching and learning. This time, the need and urgency of educational change are emphasized. In the future, school education should change from teaching students to reading "net", and from imparting hard knowledge to constructing soft knowledge together. Educational evaluation should change from learning input evaluation to learning output evaluation [2].

In order to prevent and control the COVID-19 epidemic, online teaching using online platforms has become the focus of the whole society. Epidemic prevention and control during the "closed not suspended" online teaching way analysis model with emphasis on "Internet + education" thought, reflect the epidemic prevention and control during the education demand and respecting student learning rule that occupies the home, on the basis of this model is formed during the epidemic prevention and control of "closed not suspended" online analysis dimension and its connotation, the way of teaching It has the characteristics of systematization, standardization, and operability, and can be used to guide the planning and practice of online teaching [3].

The integration of online and offline education, the distinction between self-learning and self-management, the synergy between home and school education, the fairness of infrastructure, and information literacy are the primary issues that the education reform is now confronting. Therefore,

the new countermeasures to education reform include using big data to enhance education governance, enhancing teachers' capacities to integrate online and offline learning, fostering the growth of personalized learning through IT, fostering the improvement of education through accurate evaluation of the educational process, and fostering effective communication between families and schools through IT.

The development of the Internet provides technical support for the transformation of teaching models. In particular, the arrival of COVID-19 has accelerated the public's recognition and acceptance of online learning. The study takes students of Delhi-NCR online learning platform as the research object and explores the relationship between various influencing factors based on regression analysis [4]. In order to improve the quality of online teaching and the quality of students' home learning better, suggestions and guidance strategies can be put forward to promote the quality of students' learning based on the correlation between various factors obtained from the regression analysis, hoping to provide a feasible way to improve the quality of students' home learning [5].

2. Method

2.1 Data from Kaggle “Covid-19” Survey Student Responses

2.1.1 Source

The two sets of data that would be used to explore are both from a website called Kaggle.

One called “COVID-19 Survey Student Responses” was found on “<https://www.kaggle.com/datasets/kunal28chaturvedi/covid19-and-its-impact-on-students>”. In this study, a cross-sectional survey of 1182 students from various age groups and educational institutions in Delhi National Capital Region (NCR) in 2020 was done. The dependent variable is the rating of online class experience and there are 18 independent variables.

The other one is called “Responses”, which was found on “<https://www.kaggle.com/datasets/bsoyka3/educational-stress-due-to-the-coronavirus-pandemic>”. This data describes the differences in kinds of stress among 40 students from several countries before and after covid-19 pandemic in 2020. The dependent variable is stress and there are 5 independent variables.

2.1.2 Columns

In “COVID-19 Survey Student Responses”, the size of data is 1182 * 19 columns. It includes 19 variables. The following Fig.1 is the data shown in Excel.

Variable	Category
ID	Numeric
Change in your weight	Categorical
Region of residence	Categorical
Stress busters	Categorical
Age of Subject	Numeric
Time spent on Online Class	Numeric
Rating of Online Class experience	Numeric
Do you find yourself more connected with your family,close friends,relatives	Categorical
Medium for online class	Categorical
Time spent on self study	Numeric
Number of meals per day	Numeric
Time spent on fitness	Numeric
Time spent on sleep	Numeric
Time spent on social media	Numeric
What you miss the most	Categorical
Prefered social media platform	Categorical
Time spent on TV	Numeric
Health issue during lockdown	Categorical
Time utilized	Categorical

Fig. 1 Description Of Variables In “COVID-19 Survey Student Responses”
(Photo credit: Original)

In “Responses”, the size of data is 41 * 15 columns. It includes 15 variables. The following Fig.2 s the data shown in Excel.

Variable	Category
Category	Categorical
Country	Categorical
State	Categorical
Age	Numeric
Gender	Categorical
Before-Environment	Categorical
Before-ClassworkStress	Numeric
Before-HomeworkStress	Numeric
Before-HomeworkHours	Numeric
Now-Environment	Categorical
Now-ClassworkStress	Numeric
Now-HomeworkStress	Numeric
Now-HomeworkHours	Numeric
FamilyRelationships	Numeric
FriendRelationships	Numeric

Fig. 2 Description Of Variables In “Responses”
 (Photo credit: Original)

2.2 Method of finding interfering factors that influence the responses from student

2.2.1 Source

Data preprocessing is the basis of data analysis. Data analysis is to mine the internal connections hidden in a large amount of disorganized data, to help find reasons, to aid decision-making, and so on. On the whole, it can be divided into four parts: data collection, data preprocessing, feature engineering, and model building.

The purpose of data preprocessing is to aim at the problems such as missing values and outliers that may exist in native data, data preprocessing is used to improve data quality and reliability of data analysis. The main steps of data preprocessing are as follows. The first step is data cleaning: Addressing data discrepancies, including outlier removal, noisy data smoothing, and missing value processing. Integration of the data is the next phase. the collection and storing of several data sources in a single location, such as a data warehouse. The third step is data transformation: standardizing, discretizing, and sparsing data based on analysis requirements. And the last step is data reduction: on the basis of maintaining data integrity, the data size is effectively reduced to obtain the reduced representation of data sets

2.2.2 Columns

After data preprocessing, the data is clean and regular. It is feasible to do some data statistics to show the general situation of data. This can help to learn the whole data and make sure which factors can be explored meaningfully. Python and Excel are good tools to do data statistics.

2.2.3 Regression analysis

In statistics, regression analysis refers to a statistical analysis method used to determine the quantitative relationships between two or more variables that are dependent on each other [6]. Depending on how many variables are being considered, regression analysis may be separated into single regression analysis and multiple regression analysis. Simple and multiple regression analyses may be classified based on the number of dependent variables, while linear and nonlinear regression analyses can be classified based on the nature of the connection between the independent and

dependent variables. Linear regression analysis is suitable for this exploration. Through calculating, the formula (1) can be obtained as follows.

$$Y = c + k1 * x1 + k2 * x2 + \dots + kn * xn + \varepsilon \tag{1}$$

In the formula, Y means The dependent variable and xn means The independent variables. Kn means the coefficient of xn, c means constant and ε means error.

3. Results

3.1 Visualization

3.1.1 Stress before and after covid-19 pandemic

Using data “Responses” to draw three-line charts and these three charts show the classwork stress, homework stress, and homework hours respectively from students in the US and some other countries as shown in the Fig.3-5.

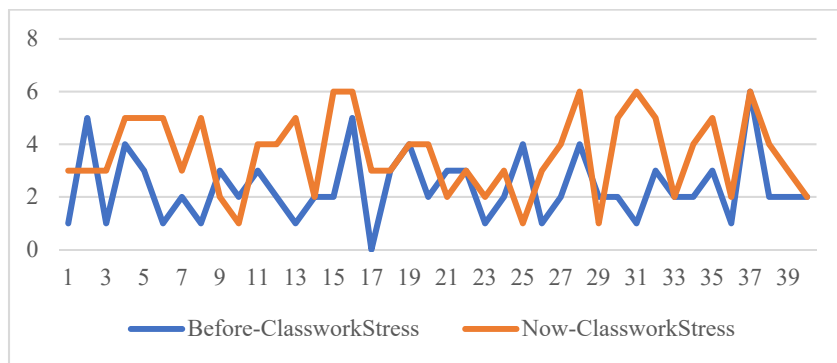


Fig. 3 Classwork Stress
 (Photo credit: Original)

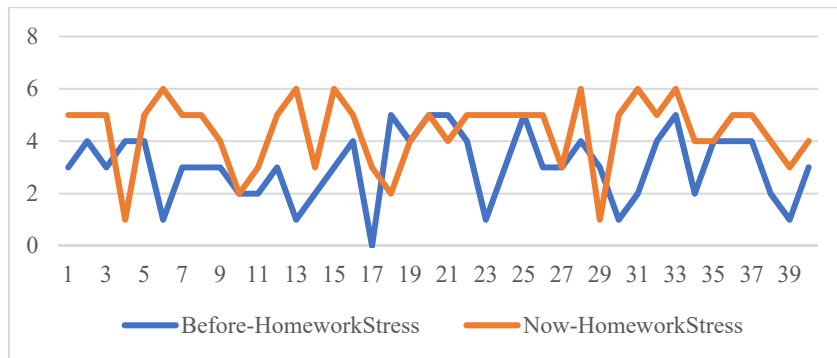


Fig. 4 Homework Stress
 (Photo credit: Original)

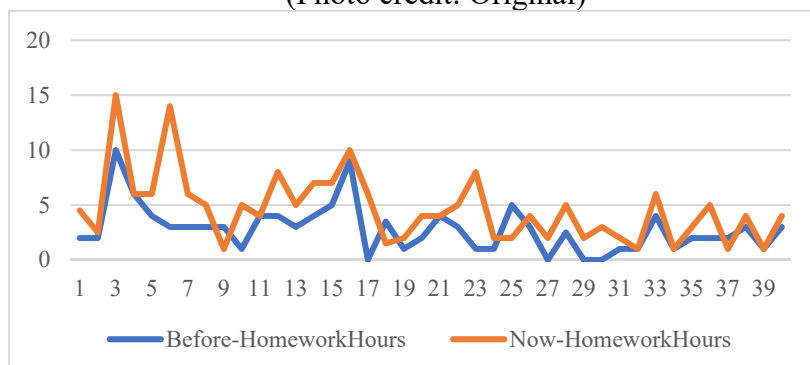


Fig. 5 Homework Hours
 (Photo credit: Original)

The orange line represents after covid-19 and the blue line is before. According to these three figures, it is obvious that the orange lines are almost above the blue lines and this explains that the stress of classwork and homework and homework hours after covid-19 pandemic are larger than before. It is not a good trend.

3.1.2 Different regions and responses

In the data “COVID-19 Survey Student Responses”, there is a column called Region of residence. And it can be divided into two parts, Delhi-NCR and Outside Delhi-NCR. According to the two categories, use Excel to calculate the average rating of online class experience and draw the chart as shown in Fig.6.

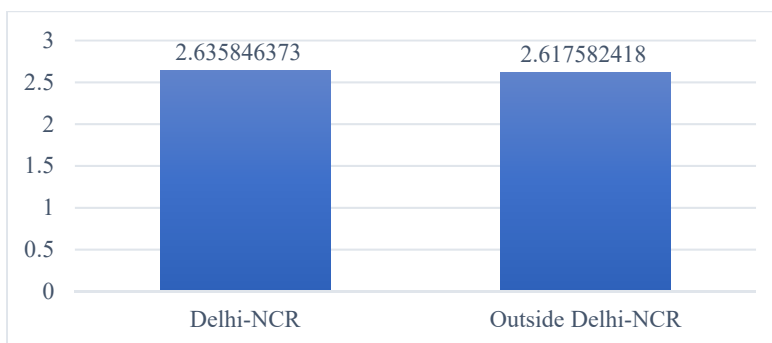


Fig. 6 Average rating
(Photo credit: Original)

From the figure, the average rating of online class experience in Delhi-NCR is 2.6358 and Outside Delhi-NCR is 2.6176. There is not much difference between the two so it is feasible to consider that the region of residence has no relation to the rating of online class experience. It is meaningless to explore further.

3.1.3 Different grades and responses

In the data “COVID-19 Survey Student Responses”, according to the age, a column named “Grade” need to be set. Those younger than 12 years old are “Primary”. Those younger than 12 and less than 15 are “Junior”. those younger than 15 and less than 18 are “Senior”. those younger than 18 years old are “University”.

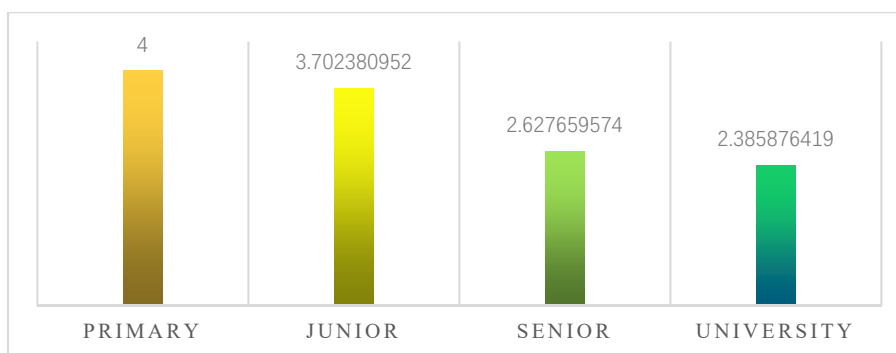


Fig. 7 Average rating of online class experience
(Photo credit: Original)

Primary school students have the best Online Class quality and Rating of Online Class and university students have the worst Rating. As the grades do up, the rating of online class experience gets lower and lower, as shown in Fig.7.

3.1.4 Different hours spent in studying and responses

The relationship between grades and time spent on online class is unexpected. University, which is supposed to be the higher level of school, but it has lowest time spent on online class, even lower

than primary school level. Junior and Senior have the most average time on online class, as shown in Fig.8.

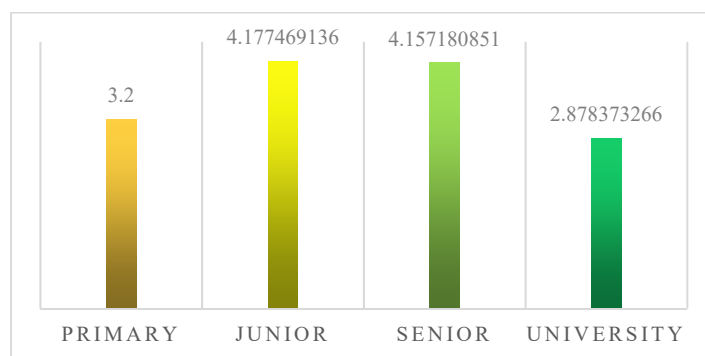


Fig. 8 Average time of different grades
(Photo credit: Original)

3.1.5 Different mediums used in studying and responses

In data “COVID-19 Survey Student Responses”, it is easy to know that Laptop/Desktop is used most and Tablet is the least. Fig.9 shows the rating of online class of each medium.

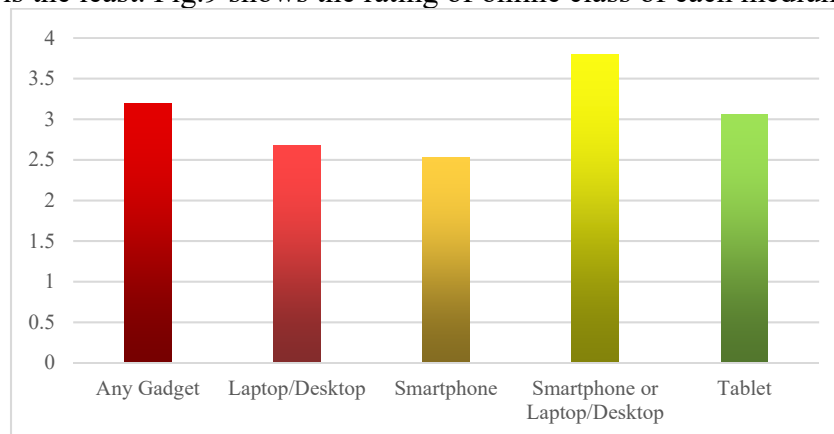


Fig. 9 Average rating of different mediums
(Photo credit: Original)

According to Fig.9, it is obvious that using a smartphone or Laptop/Desktop is more popular among these students because the average rating of online class is the highest. On the contrary, using smartphones separately has the worst effect.

3.2 Regression

3.2.1 Self-study and online study

It is meaningful to see if online class and self-study have influence on the responses about rating of online class from students and which one is larger. Use Regression in Excel to calculate these values. Table 1 is the results.

Table 1. Regression Result of Self And Online Study

	Coefficients	Standard error	t Stat	P-value
Intercept	2.284840126	0.087376023	26.14950916	9.0695E-119
Time spent on Online Class	0.101672554	0.019240893	5.284190935	1.5084E-07
Time spent on self-study	0.003842269	0.018808196	0.204286967	0.838165245

In the results, P-value of “Time spent on self-study” is 0.8381 which is larger than 0.05 so it has no influence on the responses. P-value of “Time spent on Online Class” is 1.51E-07 which is less

than 0.05 so it has strong influence on the responses. And through the results, obtain the equation (2) below.

$$\begin{aligned} \text{Rating of Online Class experience (Y)} \\ = 2.2848 + 0.1017 * \text{Time spent on Online Class (X)} + \varepsilon \end{aligned} \quad (2)$$

The coefficient of “Time spent on Online Class” is 0.1017, larger than 0, which is a positive correlation. The more time spent on Online Class is more, the quality of responses in higher. It can be inferred that students consider that, within a certain time frame, the quality of education is better with longer time spent on online class.

3.2.2 Different ways of spending time and rating of online class experience

Except for time spent online, there are plenty of parts of time spent on different events. And it is feasible to explore if the proportion of time spent on other events has an influence on the responses. Use Regression in Excel to calculate these values. Table 2 is the results.

Table 2. Regression Result of Time Spent On Different events

	Coefficients	Standard error	t Stat	P-value
Intercept	2.59769222	0.230802541	11.25504168	5.92072E-28
Time spent on self-study	0.012895518	0.019181289	0.672296728	0.501529626
Time spent on fitness	0.149201239	0.054277628	2.748853343	0.006073872
Time spent on sleep	-0.001151756	0.024955859	-0.046151715	0.963197326
Time spent on social media	-0.115270683	0.02286234	-5.041946	5.34843E-07
Time spent on TV	0.15733632	0.031307405	5.025530565	5.81609E-07

Through observing the P-value in results, the P-value of “Time spent on self-study” and “Time spent on sleep” are larger than 0.05 so there is no relation between two variables and responses. However, P-value of “Time spent on fitness”, “Time spent on social media” and “Time spent on TV” are 0.0061, 5.3484E-07, 5.8161E-07 respectively. This means they have an influence on responses. And through the results, obtain equation (3).

$$\begin{aligned} \text{Rating of Online Class experience (Y)} \\ = 2.59 + 0.1492 * \text{Time spent on Fitness (X1)} - 0.1153 \\ * \text{Time spent on social media(X2)} + 0.1573 \\ * \text{Time spent on TV(X3)} + \varepsilon \end{aligned} \quad (3)$$

The coefficients of “Time spent on fitness” and “Time spent on TV” are 0.1492 and 0.1573 respectively, larger than 0, which means it is a positive correlation. On the contrary, it is amazing that the coefficient of “Time spent on social medium” is -0.1153 which is a negative correlation. It can be inferred that students consider that, they prefer the online class which can give them more time spent on fitness and TV.

3.2.3 Different mediums students prefer to use and rating of online class experience

Except exploring time spent on different events, it is feasible to know if there is a relation between the kinds of mediums and responses from students. Use Regression in Excel to calculate the values of each medium. Table 3 is the results.

Table 3. Regression Result of Different Mediums

	Coefficients	Coefficients	t Stat	P-value
Intercept	3.8	0.607230846	6.25791661	5.53499E-10
Laptop/Desktop	-1.128440367	0.61000995	-1.849872068	0.064593861
Smartphone	-1.269387755	0.610040815	-2.080824307	0.037675593
Tablet	-0.745945946	0.646960265	-1.153001175	0.249154604
Any Gadget	-0.6	0.858754099	-0.698686622	0.484892174

Through observing the P-value in results, only the P-value of Smartphone is less than 0.05 and its coefficient is -1.12, less than 0, which means negative correlation. So by observing other variables' coefficients it is reasonable to consider that students do not like using Smartphone to study online class.

3.2.4 Grade, change in weight and rating of online class experience

It is unknown if a change in weight can influence different grades to give the rating of online class experience. Use Regression in Excel to calculate values of Senior and University with change in weight (Senior = 0, University = 1; Increased = 0, Remain Constant = 1, Decreased = 2). Table 4 is the results.

Table 4. Regression Result of Different Mediums

	Coefficients	Standard error	t Stat	P-value
Intercept	2.668674699	0.150010843	17.78987874	1.51404E-61
Grade	-0.319528806	0.164903769	-1.937668305	0.052951597
Change in your weight	-0.048192771	0.136217891	-0.353791787	0.723571261
Grade*Change in your weight	0.094722016	0.150443847	0.629617081	0.529092349

Through the results, P-value of the three are all larger than 0.05 so there is no relation between Grade and Change in your weight to influence the responses.

4. Discussion

In the first regression “Self-study and online study” it is known that “Time spent on Online Class” has strong influence on the students’ responses, however, “Time spent on self-study” has no relation with responses. It means that students in Delhi-NCR pay more attention to online class instead of self-study. But self-study is also an effective way of acquiring knowledge. In the process of online teaching, it is impossible to achieve the same student management and supervision as offline classroom, which tests students’ autonomous management ability and puts forward higher requirements for students’ autonomous learning ability. Students should plan reasonably, try their best to avoid the interference of irrelevant factors to complete the learning task and take the initiative to find solutions to solve problems, and gradually develop the ability of independent learning. However, for the phenomenon that students prefer online class, teachers’ teaching should be improved more necessary [7]. In the process of online teaching, many teachers lack sufficient online teaching experience, and teachers and students are not in the physical space, cannot really grasp the specific situation of students, online learning time is still not long, the guidance of students is not fully played.

In the third regression and the fifth visualization, it is obvious that the students’ responses to using smartphone or Laptop/Desktop are the highest, then is Any Gadget, and only Smartphone medium has a relation with students’ responses. These can be called learning conditions. “Learning condition” is the main factor that affects the effect of students' online learning. The learning environment of online teaching is mainly composed of a learning platform network and learning equipment. Due to the sudden outbreak of the epidemic, many students do not have complete learning equipment, such as some students do not have computers at home, or the network is not smooth, which will greatly affect students' online learning experience and directly affect their learning results. For this phenomenon, government and schools can take some measures as they can provide some cheap products for an online class and these products can be used in a circle. After the graduates are used, they can be taken back and given to the freshmen to use [8, 9].

It is feasible that accurate online teaching and an effective home learning model could be built, which help students achieve academic development and personal growth by providing high-quality online instruction and personalized tutoring. Teachers should make the teaching design as easy to understand, precise and accurate as possible, and the teaching task should be scientific and reasonable.

According to the teaching goal selection, production, integration of easy-to-use and applicable teaching resources, as far as possible to choose the existing high-quality resources, have specific needs to make their own resources, the production of resources to be small and fine. According to the teaching objectives and tasks, the existing resources can be appropriately introduced or transformed to create an extended resource base for students to facilitate their extended learning. In autonomous learning, students are the main body of learning [10]. They create personal learning goals in accordance with their own circumstances and requirements, guided by the overarching teaching objectives, and they accomplish the goals independently through autonomous analysis, inquiry, practice, questioning, invention, and other techniques.

5. Conclusion

This study mainly focuses on the responses of students in Delhi about online education classes and analyzes the influencing factors of online education. Both the time spent on online study and other events and the mediums used for online study has a strong relationship with the responses of students. And the homework stress and classwork stress ascend obviously. This is not a good trend which needs to be improved. What's more, as the grades do up, the rating of online class experience gets lower and lower. Education in university has the worst responses. One of the reasons for this phenomenon may be that university students have less supervision than other grade students and the self-awareness of learning is lower than that of lower grade students.

This study profits schools and teachers to make more careful teaching plans and government can provide more and better material support so that it can improve the quality of education. During the epidemic period, online teaching is an effective measure to ensure that students have access to education. In order to promote the real occurrence of learning, teaching design should consider the relationship and interaction in the learning process, pay attention to the construction of the learning environment through the guidance of teachers and enhance the interaction between teachers and students. As "lonely learners" who study at home during the epidemic, students need teachers' guidance and parents' support in both academic growth and psychological support. After a period of online teaching and home learning, personalized reports should be generated for students, pointing out the problems, and providing targeted individual counseling for students accordingly. The problems in the online teaching mode should be fed back to the educational administrative department and teaching guidance personnel in time to further improve the teaching mode.

As long as people have enough mental preparation and ability reserve, stick to the concept of student development as the center, and return to teaching effectiveness, online teaching will be able to achieve results. At the same time, people can make use of the home environment to promote students' labor education and family education and strengthen life education and health education.

References

- [1] Xu Jinjie the response of the global education system and the challenge of online education under the COVID-19 epidemic -- findings and Reflections Based on the OECD global survey results. *Comparative education research*, 2020,42 (06): 3-10.
- [2] Wang Zhuli. How should education be transformed in the post epidemic era? *Research on audio visual education*, 2020,41 (04): 13-20.
- [3] Xie youru, Qiu Yi, Huang Yuling, Wang qinlei. Characteristics, problems and innovation of the online teaching method of "no class, no school" during the epidemic prevention and control. *Research on audiovisual education*, 2020,41 (03): 20-28.
- [4] Dai Hong, Yang Xuefei. Analysis of influencing factors of large-scale online learning behavior intention during the epidemic period. *Journal of higher education*, 2022,8 (21): 65-69.
- [5] Song Lingqing, Xu Lin, Li Yayu. Precise online teaching + Home Learning Mode: a way to improve students' learning quality during the epidemic period. *China audio visual education*, 2020 (03): 114-122

- [6] Chen Xiang, Cheng Dehua, Chen Zufen. Epidemic prevention and control during the student online learning effect and its influencing factors to explore the. *Journal of Chinese modern education equipment*, 2020 (11): 14-17 + 21.
- [7] Marčec Robert, Majta Matea, Likić Robert. Satisfaction and attitudes of Zagreb medical students towards online education during the COVID-19 pandemic. *Liječnički vjesnik*, 2020, 142(S5).
- [8] Luo Changyuan, Si Chunxiao Will online education widen the gap between students with different family conditions --Quasi natural experiment based on COVID-19. *Financial research*, 2020,46 (11): 4-18.
- [9] Xiong Yong, Ling Que, Li Xiaoli. Ubiquitous e-Teaching and e-Learning: China's Massive Adoption of Online Education and Launching MOOCs Internationally during the COVID-19 Outbreak[J]. Hindawi Limited, 2021.
- [10] Wang Yunwu, Wang Yuru, Li Yanxin, Li Dan. Countermeasures and suggestions for improving the quality of online education during epidemic prevention and control. *China medical education technology*, 2020,34 (02): 119-124 + 128.