Statistical Analysis on the Preferences of MOOC Learners

Qisang Lin, Xiangqi Zheng, and Kunping Zhu(✉)

Department of Mathematics, East China University of Science and Technology, Shanghai 200234, China
kpzhu@ecust.edu.cn

Abstract. The paper aims to explore learner preferences for different online videos of open platforms. The research collects and analyzes qualitative and quantitative data by extracting audio and video features and questionnaire survey. Correlation analysis, cluster analysis and ANOVA are used as tools to analyze the collected data on variables influencing learner preferences. The results indicate that *learners prefer videos with a lecturer in camera* and videos with appropriate *loudness*, while the *speaking speed* and *the duration of single episode* may also be the factors influencing learner preferences. Based on the conclusion, some suggestions are put forward for online teaching.

Keywords: MOOC videos · Learning preferences · Statistical analysis · Online learning

1 Introduction

Since the outbreak of the global epidemic, online learning gradually became one of the most important teaching modes. Pelletier and Brown et.al [1] pointed out that curriculum design should emphasize the learning experience of students. Many colleges and universities started online classes or flipped classes to help learners achieve higher-level learning goals [2, 3]. Moreover, educators realize the necessity of development and utilization of technology to provide students with more effective courses [4]. At the same time, Pelletier and McCormack et.al [5] indicated that teachers should have the ability to effectively carry out mixed teaching and receive training on digital tools and teaching methods.

There are researches focusing on the effectiveness of online classroom teaching. Amin and Sundari [6] used the questionnaire to evaluate the effectiveness of the class. Hunyan et.al [7] used the systematic evaluation to compare online and offline classes. Wang [8] took MLR as a method to find out the relationship between SRL, online behavior and scores. Rasheed et.al [9] concluded that many learners preferred to use external commercial website resources for online materials and learning. Based on methods and conclusions of former researches, we try to explore the relationship between variables of videos and learner preferences.

Y. Yau and F. Hage Chehade (Eds.): DESD 2022, ASSEHR 691, pp. 302–311, 2023.
https://doi.org/10.2991/978-2-494069-37-4_38

Because the content and teaching mode differ among different disciplines, the influencing factors are also different. So we only take *Probability and Statistics* as an example, using the data of 16 MOOC videos on Bilibili *(Bilibili, similar to YouTube, is the most popular video-sharing website among young people in China. Videos of 16 courses can be found in this website of Bilibili.*https://search.bilibili.com/all?keyword=%E6%A6%82%E7%8E%87%E8%AE%BA%E4%B8%8E%E6%95%B0%E7%90%86%E7%BB%9F%E8%AE%A1&from_source=webtop_search&spm_id_from=333.851*)* courses as the benchmark to analyze learner preferences. However, since this study only involved a special group of students who had taken courses in *Probability and Statistics* in the mathematics discipline, the results of the study must be applied to other groups and disciplines with discretion.

We collect the data by extracting the audio of the course, using Adobe Audition to extract and calculate the *average loudness* of the audio, using the loudness detector program to obtain the *loudness range.* Moreover, the lecturer's teaching content is converted to text by the YUE recording *(A software used to convert speech into text.)*, *speaking speed* is calculated. We also collect the *like, view, coin (Like, coin and other dimensions are the evaluation of videos by users of Bilibili, the more like they get, the more popular they are.)* and other index, and conduct a questionnaire on students who participated in courses to investigate their satisfaction. After the data is collected, we use SPSS for correlation and cluster analysis to target the indicator that best reflects learner preferences corresponding to learner preferences [10]. In addition, some videos with high views but not included in the *high-quality MOOC courses* are compared with the accreditation standards.

The paper is organized as follows: Sect. 2 is devoted to the data collection. Section 3 focuses on the analysis of data, including the cluster analysis, correlation analysis and ANOVA. Section 4 is a summary of findings.

2 Data Collection

2.1 Course Selection

Before sampling, we select qualified courses and put them into the sampling box according to several criteria:

1) Among different versions of the same course video uploaded by different influencers, we choose the version with the highest views;
2) Usually a course video is divided into different episodes, we only choose the ones with total duration of all episodes together should be no less than 1 h;
3) The online course should cover main topics of *Probability and Statistics.*

Therefore, 157 courses are firstly selected. Then, 16 course videos are selected for analysis through simple random sampling.

2.2 Data Acquisition

The whole data of 16 courses is presented above in Table 1. Because it involves the privacy of influencers, we use letters instead of names to collect.

Table 1. Basic data, video and audio feature of *Probability and Statistics*

Name	Duration (episode)	View	Bullet subtitle	Like	Coin	Collection	Comment	Camera	Method	Speaking speed	Average loudness	Loudness range
A	30	20911000	640000	410000	397000	549000	13000	1	1	224	-23	5.5
B	30	2155000	3995	46000	35000	60000	2335	0	0	245	-22.9	5.4
C	108	1675000	23000	65000	53000	80000	4026	0	0	315	-30.3	4.8
D	10	1235000	22000	25000	24000	48000	1604	0	0	271	-18.2	5.4
E	10	1866000	26000	21000	8087	71000	1150	1	0	172	-12.4	7.5
F	13	2620000	11000	39000	33000	73000	6476	0	0	270	-24.2	8.3
G	39	630000	10000	13000	12000	26000	539	1	1	254	-22.2	9.4
H	19	262000	768	4952	3267	16000	139	1	1	250	-18.4	11
I	50	147000	3590	1863	691	8209	201	1	2	236	-26.3	14
J	11	55000	257	1051	525	2395	99	1	2	314	-28.9	6.4
K	60	81000	1129	1077	870	5175	188	1	1	167	-26.9	12.8
L	10	23000	126	304	100	1264	42	0	0	258	-18.6	8.5
M	11	38000	91	1106	1124	3540	87	0	2	210	-18.3	8.6
N	120	27000	37	507	231	1691	114	1	1	180	-27.6	15.5
O	30	28000	169	321	104	947	29	1	2	249	-12.5	3
P	40	45000	298	407	306	1735	124	1	1	124	-8.7	6.3

The duration of single episode, view, bullet subtitle, like, coin and *collection* can be directly collected from Bilibili. From Table 1, it is clear that there are four levels of *view*:

1) One course with more than 20 million views (A),
2) Five courses with views between 1 million and 3 million (B, C, D, E, F),
3) Three courses (G,H,I) with 100,000 to 1 million views,
4) Seven courses with 10,000 to 100,000 views (J, K, L, M, N, O, P).

As for video features, *teaching method* and *whether the lecturer is in the camera* are collected by classification.

For audio features, speed refers to the speaking speed of lecturers in online courses [11]. We select 100 sections of the lecturer's voice, calculate the speaking speed of each section, and then arrange the speaking speed in descending order. Then, we take 60% of the middle speaking speed segments to calculate their mean value to get the lecturer's speaking speed. Because the relationship between volume and audio waveform is complex, listeners have different sensitivity to different frequencies of audio subjectively [12]. Loudness refers to the energy of the audio produced by the device, indicating the amplitude of the audio signal *(According to the value of the audio standard of TV programs in China, the average Loudness target value for the audience to feel comfortable should be between -26 LUFS and -22 LUFS.)*. The *loudness range* represents the contrast of the loudness of the audio. The loudness range is obtained by subtracting the average of the maximum loudness from the average of the minimum loudness. We use the loudness analysis of Adobe Audition to get the *average loudness* of each course lecturer. *Loudness range* refers to the contrast of audio loudness, the dispersion of audio short-time loudness level. We use loudness detector to calculate the *loudness range* of each course audio.

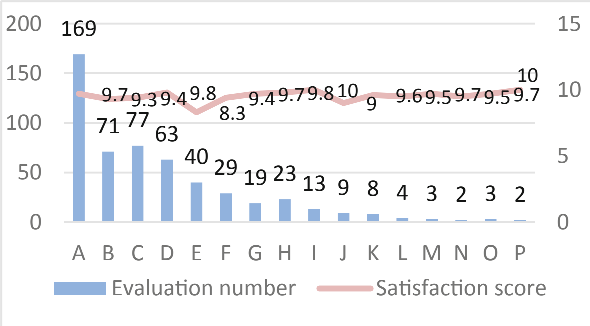

Fig. 1. The questionnaire result

2.3 Questionnaire Collection

The questionnaire consists of two questions:

1) Which lecturer have you watched Probability and Statistics on Bilibili (multiple choices);
2) Please grade this course on a scale of 1 to 10 (Fig. 1).

In order to test the stability and reliability of the data, the reliability test is carried out on the data by SPSS for reliability analysis [13]. The Cronbach's alpha is 0.719, indicating that the stability of the questionnaire is good.

3 Data Analysis

3.1 Cluster Analysis

This paper adopts Q-type clustering, readers may refer to [10] for more details.

Firstly, the basic data (*view, coin, like, collection, bullet subtitle, comment* and satisfaction) is analyzed by clustering with SPSS. Since the action of *coin, like, comment, collection* and *bullet subtitle* cannot be taken without viewing the video, dividing the data by *view* is a better indicator of learner preferences.

Secondly, from Table 1, it is clear that the order of magnitude of *view* is too large compared with other data. No matter what cluster method and interval is defined, the course always depends on *view* when it is divided into two categories. Therefore, the method is improved to maximize all basic data and dimensionless variables, so that the index and order of magnitude of each data would no longer have an impact on cluster analysis. The maximization method used is to divide the value of the variable by the full distance of the value of the variable, so that the value of the variable is calculated between 0 and 1. The calculation formula is as follows:

$$\hat{X} = \frac{X_i - min}{max - min} = \frac{X_i - min}{R}$$

Table 2. Cluster center

Categories	Satisfaction	View	Bullet subtitle	Like	Coin	Collection	Comment
1	0.8	0.233378016	0.50327056	0.5589347	0.716965939	0.346882778	0.260932535
2	0.684491979	0.030709237	0.21885701	0.1969592	0.173134802	0.297942815	0.349535707

Cluster analysis is conducted on the above data, and the following results are obtained in Table 2.

Category 1 has 5 courses (A,C,D,G,M), Category 2 has 11 courses (B,E,F,H,I,J,K,L,N,O,P).

ANOVA is conducted for each variable, and the following results are obtained:

For variable satisfaction*, coin* and *collection,* the significance P-value is less than 0.05, indicating that these variables have significant differences among the categories divided by cluster analysis;

For other variables, the significance P-value is more than 0.05, indicating that there are no significant differences on variables among the categories divided by cluster analysis.

To sum up, among the 16 course videos, there are significant differences in the variables of course satisfaction, *coin* and *collection* among the 16 course videos. Most of courses in the first category have high satisfaction scores above 0.82. And the *view, coin, bullet subtitle, like* and *collection* are higher than courses in the second category of courses. It can be inferred that the first type of courses are more favored by learners.

3.2 Correlation Analysis

According to the above cluster analysis, variables with significant differences can be obtained, and the difference in the center value of *coin* between the first and second categories is the largest. Therefore, *coin* is used as a dependent variable, and the *average loudness*, *loudness range, duration of single episode* and *speaking speed* are respectively used as independent variables for regression analysis. To prove the validity of dividing two categories by *coin*, Kendall coordination coefficient is used to analyze the consistency between variables.

We do the Kendall coefficient consistency test by two steps:

1) First test the significance level of variables, and judge whether they are significant by P-value ($P < 0.05$ or 0.01). If they are significant, it indicates that there is consistency between data;
2) Analyze the positive and negative direction of Kendall coefficient and its correlation degree;
3) Summarize the above analysis results.

As a result, in four tests, the Kendall coefficient consistency test results all show that the significance P-value of the overall data is 0.000, presenting significance on the level and rejecting the null hypothesis, so the data presents consistency. Meanwhile, the Kendall coordination coefficient W value of the model is 1.0, so the degree of correlation is almost complete consistency. Hence, it is reasonable to divide the two categories by *coin/view*.

The correlation between basic data and *average loudness, duration of single episode, loudness range* and *speaking speed* is analyzed by Pearson's coefficient to test the impact of these variables on learner preferences.

We do the Pearson's coefficient test by two steps:

1) First test whether there is a significant relationship between X and Y in the statistical category, and judge whether there is a significant relationship by P-value (* $P < 0.05$, ** $P < 0.01$);
2) If the P-value shows the significance of the results, there is correlation between X and Y; if not, there is no correlation between the two variables;
3) Analyze the positive and negative direction of the correlation coefficient and the degree of correlation.

Cluster analysis is conducted on the above data, and the following results are obtained in Table 3. Table 3 is about correlation numbers. The first line in the Table is F-value and P-value is in parentheses. If P-value is less than 0.05, it indicates that there is correlation between the two variables. According to Table 3, *speaking speed* and *like/view*, *loudness range* and *collection/view*, t*he duration of single episode* and *comments/view* have the correlation.

From Table 3 we can get that among the above basic variables, the amount of *collection/view* is significantly different between the two categories, so the available *loudness range* is the influencing factor of learner course preference, while *speaking speed* and *duration of single episode* may be the influencing factor. This just confirms the realistic view that the *average loudness* and *speaking speed* can be adjusted on the video platform, so it has little influence on the video preferences of learners. As different *loudness range* bring listeners noisy or comfortable feelings, it is reasonable that *loudness range* can be one of the influencing variable.

Loudness range between 5 to 9 can have a better effect on learners. Because such a moderate range of loudness will bring comfortable listening feelings to learners, we guess that these vivid and cadenced tones will motivate students more.

3.3 Analysis of Variance

For the video features of each course as shown in Table 1, since they are classified by two-classification or three-classification method and variables with significant differences can be obtained from the above cluster analysis, the amount of *coin/view* is used as quantitative data because the amount of *coin/view* from the first and second categories has the largest difference in the centre value. One-way ANOVA is used to test *whether the lecturer is in the camera* and *teaching method* have a significant impact on *coin*. The steps of one-way ANOVA are as follows:

1) Classify the quantitative variable (Y) through the fixed class variable (X) and conduct the normality test respectively to observe whether the overall distribution of the data

Table 3. Correlation analysis result

Variable	Satisfaction	View	Bullet subtitle/view	Like/view	Coin/view	Collection/View	Comment/view	Duration (episode)	Average loudness	Loudness range	Speaking speed
Satisfaction	1.000 (0.000***)	0.030 (0.913)	0.162 (0.549)	0.002 (0.995)	0.143(0.599)	0.211(0.433)	0.124(0.646)	0.148 (0.584)	0.037 (0.890)	0.171 (0.527)	-0.073 (0.787)
View	0.030 (0.913)	1.000 (0.000***)	0.619 (0.011*)	0.086 (0.750)	0.220(0.413)	-0.411(0.114)	-0.321(0.226)	-0.072 (0.791)	-0.092 (0.735)	-0.267 (0.317)	-0.001 (0.996)
Bullet subtitle/view	0.162 (0.549)	0.619 (0.011**)	1.000 (0.000***)	-0.015 (0.956)	0.108(0.691)	-0.272(0.309)	-0.416(0.109)	0.040 (0.883)	-0.143 (0.597)	-0.028 (0.918)	0.006 (0.983)
Like/view	0.002 (0.995)	0.086 (0.750)	-0.015 (0.956)	1.000 (0.000***)	0.914(0.000***)	0.273(0.306)	0.119(0.661)	0.355 (0.177)	-0.495 (0.051*)	-0.192 (0.476)	0.515 (0.041**)
Coin/view	0.143 (0.599)	0.220 (0.413)	0.108 (0.691)	0.914 (0.000…)	1.000(0.000***)	0.228(0 395)	0.036(0.894)	0.173 (0.522)	0.331 (0.211)	0.253 (0.345)	0.352 (0.182)
Collection/view	0.211 (0.433)	-0.411 (0.114)	-0.272 (0.309)	0.273 (0.306)	0.228(0.395)	1.000(0.000***)	0.367(0.162)	0.172 (0.524)	-0.131 (0.628)	0.563 (0.023*)	-0.195 (0.469)
Comment/view	0.124 (0.646)	-0.321 (0.226)	-0.416 (0.109)	0.119 (0 661)	0.036(0.894)	0.367(0.162)	1.000(0.000***)	0.610 (0.012*)	-0.299 (0.260)	0.395 (0.130)	-0.258 (0.334)
Duration (episode)	0.148 (0.584)	-0.072 (0.791)	0.040 (0.883)	0.355 (0.177)	0.173(0.522)	0.172(0.524)	0.610(0.012*)	1.000 (0.000***)	-0.504 (0.047**)	0.384 (0.143)	-0.121 (0.655)
Average loudness	0.037 (0.890)	-0.092 (0.735)	-0.143 (0.597)	-0.495 (0.051*)	-0.331(0.211)	-0.131(0.628)	-0.299(0.260)	-0.504 (0.047*)	1.000 (0.000***)	-0.369 (0.159)	-0.484 (0.057,)
Loudness range	0.171 (0.527)	-0.267 (0.317)	-0.028 (0.918)	-0.192 (0.476)	-0.253(0.345)	0.563(0.023**)	0.395(0.130)	0.384 (0.143)	-0.369 (0.159)	1.000 (0.000***)	-0.351 (0.183)
Speaking speed	-0.073 (0.787)	-0.001 (0.996)	0.006 (0.983)	0.515 (0.041**)	0.352(0.182)	-0.195(0.469)	-0.258(0.334)	-0.121 (0.655)	-0.484 (0.057*)	-0.351 (0.183)	1.000 (0.000***)

Table 4. Results of ANOVA test for camera

Variable	Camera	Sample Size	Average	Standard Variation	F	P (two-tailed)
Coin/view	1.0	10	0.010	0.006	5.305	0.037
	0.0	6	0.019	0.010		
	Sum	16	0.013	0.009		

Table 5. Results of ANOVA test for method

Variable	Method	Sample Size	Average	Standard Variation	F	P (two-tailed)
Coin/view	1.0	6	0.013	0.005	0.133	0.876
	0.0	6	0.015	0.010		
	2.0	4	0.012	0.012		
	Sum	16	0.013	0.009		

is normal distribution. If the test fails, the normality test can be used for further analysis;

2) Classify the quantitative variable (Y) according to the fixed variable (X) and conduct homogeneity test of variance. By observing whether the P-value is less than 0.05 or 0.01, if the P-value is greater than 0.05 (0.01), use ANOVA test to check whether the P-value is significant (less than 0.05 or 0.01); Theoretically, the data must pass normality test and homogeneity test of variance before one-way ANOVA can be conducted. Otherwise, non-parametric test is used.;
3) If P is less than 0.05, the data results are significant. There is a correlation between two variables. Otherwise, there is no correlation between two variables. The differences between the data can be analyzed by means of mean $\pm$ standard deviation; if they are not significant, there is no difference;
4) If a significant result is obtained from one-way ANOVA, the difference can be quantitatively analyzed through effect quantitative analysis.

Through the above steps, results can be seen in Table 4 and Table 5 that there are significant differences in the quantitative analysis of effects and significant differences in the analysis of variance on *whether lecturers are in the camera*, but there is little difference in *teaching method*. Therefore, it can be concluded that whether lecturers are in the camera has a significant impact on the learning preferences of learners. Learners prefer lecturers in the camera. This may because that the presence of teachers makes students get immersed in classroom atmosphere better.

4 Conclusion

The paper discusses leaner preferences on video and audio features. The results indicate that learners prefer videos with *a lecturer in camera* and videos with *loudness range* between 5 to 9, while the *speaking speed* and the *duration of single episode* may also be the factors influencing learner preferences. Through the analysis of characteristics, we have that differences between *high-quality MOOC courses* and preferred *non-high-quality MOOC courses* lies in indicators such as the standardization of *duration.* At the same time, most *high-quality MOOC courses* uses PPT to improve the classroom efficiency, while some preferred courses are presented in the form of blackboard writing with lecturers appearing in the camera. It is hoped that preferences proposed in this paper will be applied, and more complete requirements for MOOC can be made by educators in the future.

Acknowledgments. The paper is supported by the Online Course Construction Project 2022 of East China University of Science and Technology.

References

1. Pelletier K, Brown M, Brooks D C, et al. EDUCAUSE horizon report teaching and learning edition. EDUCAUSE [J]. 2021.
2. Akgün M, Atıcı B. The effect of flipped classroom on learners' academic achievements and views [J]. Kastamonu Education Journal, 2017, 25(1): 329-344.
3. Awidi I T, Paynter M. The impact of a flipped classroom approach on student learning experience [J]. Computers & Education, 2019, 128: 269-283.
4. Luo J, Wang T. Analyzing Students' Behavior in Blended Learning Environment for Programming Education[C]//Proceedings of the 2020 The 2nd World Symposium on Software Engineering. 2020: 179–185.
5. Pelletier K, McCormack M, Reeves J, et al. 2022 EDUCAUSE Horizon Report Teaching and Learning Edition[R]. EDUC22, 2022.
6. Amin F M, Sundari H. EFL students' preferences on digital platforms during emergency remote teaching: Video Conference, LMS, or Messenger Application[J]. Studies in English Language and Education, 2020, 7(2): 362-378.
7. hunyan Liu, Dan Li, Baoren Zhang, Teaching Effectiveness of SPOC Flipped Classroom in College: A Systematic Review and Meta-Analysis. Open Education Research, vol.25(1), pp. 82-91, 2019.
8. Wang F H. On prediction of online behaviors and achievement using self-regulated learning awareness in flipped classrooms[J]. International Journal of Information and Education Technology, 2019, 9(12): 874-879.
9. Rasheed R A, Kamsin A, Abdullah N A, et al. Self-regulated learning in flipped classrooms: A systematic literature review[J]. International Journal of Information and Education Technology, 2020, 10(11): 848-853.
10. Thinsungnoena T, Kaoungkub N, Durongdumronchaib P, et al. The clustering validity with silhouette and sum of squared errors[J]. learning, 2015, 3(7).
11. Lowit A, Marchetti A, Corson S, et al. Rhythmic performance in hypokinetic dysarthria: Relationship between reading, spontaneous speech and diadochokinetic tasks[J]. Journal of Communication Disorders, 2018, 72: 26.

12. Umapathy K, Krishnan S, Rao R K. Audio signal feature extraction and classification using local discriminant bases[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(4): 1236-1246.
13. Bolarinwa O A. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches[J]. Nigerian Postgraduate Medical Journal, 2015, 22(4): 195.