



Short Video Service Evaluation System Design from the Viewpoint of QoE

Wan Chen, Lei Chen^(✉), Tian Zhou, Jiayi Zhu, and Ping Cui

Jiangsu Province Key Laboratory of Intelligent Industry Control Technology, Xuzhou University
of Technology, Xuzhou 221018, China
chenlei@xzit.edu.cn

Abstract. The main purpose of this study is to understand the short video industry, guide the short video provider to improve the service quality of short video. In this study, literature method, questionnaire survey method, interview method, statistical analysis method and analytic hierarchy process were used, and SPSSAU data analysis was employed. A total of 106 questionnaires were collected in this survey, and the main respondents were young people. Survey, according to the results of the weight for the primary index reflects the users more attention to the safety and reliability, we conclude that primary index can help improve the value-added services, because in some service products research, found a lot of products are set up members value-added services, users can buy them to enhance personal experience.

Keywords: Short video · Questionnaire · SPSSAU · Evaluation index · Quality of experience

1 Introduction

In recent years, with the rapid development of network technology, the short video industry in China has been developing for 10 years. So far, there are numerous short video platforms in the market, and the domestic short video market is becoming increasingly fierce [1]. Our research is mainly to analyze several popular short video platforms in the market, find out the advantages and disadvantages of different short videos, and put forward suggestions for improvement. User experience [2] is based on the users in the use of products in the process of a purely subjective feeling, the user experience is the main part of the user, the object is the product, find products through the user's user experience, the advantages and disadvantages of improve the quality of the product development and promote the development of products for the better, so the quality of the user experience is very important.

General evaluation application will release questionnaire to study the user's experience, our study also used the questionnaire survey method, after the collected data using SPSSAU data analysis it is concluded that the weight of indicators and secondary indicators, comprehensive weight, index of RI values, consistency check, etc., an analysis of the different short video platform user experience satisfaction; In the selection of

indicators, we use the SERVPERF (Quality of Service) model as the basis, and use the questionnaire survey combined with the previous scholars' research to get the indicators.

2 Investigation Methods

In this study, literature method, questionnaire survey method, interview method, statistical analysis method and analytic hierarchy process [3] were adopted, and SPSSAU data was used for analysis. In the selection of indexes, the principles of usability [4], objectivity and practicability are followed to ensure the authenticity and effectiveness of indexes. For the selection of first-level indicators, we refer to the SERVPERF model [5]. The short video service based on user experience provides users with online services through the Internet, mobile phones, tablet computers and other terminal devices. In addition, short video service has certain particularity compared with other services. In the process of researching user's short video business in the mobile Internet market, it is found that short video has established value-added business, and users can obtain personalized services different from ordinary users by purchasing value-added business. According to the SERVPERF model, the quality of service [6] is analyzed, and the following indicators are finally selected as the main indicators, including reliability, assurance, security, use experience and functionality. If the order n of the judgment matrix is greater than 2, it will bring some difficulties to obtain the consistency result of the constructed judgment matrix. Therefore, conformance testing is necessary. First, the formula is:

$$\lambda_{\max} = \frac{1}{n} \sum_{i=1}^n \frac{(A * w)}{w_i} \quad (1)$$

The maximum eigenroot of the judgment matrix A of order n is calculated, and the relative weight vector of the judgment matrix A of order n is given in the formula. Then calculate the maximum characteristic root λ_{\max} of each judgment matrix, and then calculate the magnitude of the consistency index $CI = \frac{(\lambda_{\max} - n)}{n - 1}$ by the following formula:

$$CI = \frac{(\lambda_{\max} - n)}{n - 1} \quad (2)$$

By introducing the $RI = CI/CR$ value of the index, the consistency of the judgment matrix is verified by checking the RI value of the corresponding order. When the order of the judgment matrix is 1 or 2, the judgment matrix has strong consistency. When the order of the judgment matrix is greater than 2, the consistency index $CI = \frac{(\lambda_{\max} - n)}{n - 1}$ of the judgment matrix is equal to the average value of the random consistency index RI of the same order.

In this preliminary survey, the coefficient of variation and the arithmetic mean of each indicator will be calculated based on the results of user evaluation to determine the extent of user adjustment and the consistency of user opinions. The smaller the arithmetic mean, the lower the perception of the metric in the eyes of some users, so it

can be deleted. The degree of coordination is inversely proportional to the coefficient of variation, the greater the coefficient, the greater the divergence of users, and vice versa.

The degree of users' opinions $M_y = \frac{1}{m_y} \sum_{x=1}^{m_x} C_{xy}$ can be calculated by the following formula:

$$M_y = \frac{1}{m_y} \sum_{x=1}^{m_x} C_{xy} \quad y = 1, 2, 3, \dots, n \tag{3}$$

$M_y = \frac{1}{m_y} \sum_{x=1}^{m_x} C_{xy}$ is the arithmetic average value of the score of Y index. M and C_{XY} respectively represent the number of users and the score of the index. The larger the value of M_Y is, the more important the index is in the evaluation of short video service by users.

For the coordination degree of users' opinions, $V_y = \frac{\beta_y}{M_y}$ can be calculated by the following method. First, the standard deviation of the evaluation result of Y index can be calculated by the following formula:

$$\beta_y = \sqrt{\frac{\sum_{x=1}^{m_x} (C_{xy} - M_y)^2}{m_y - 1}} \quad y = 1, 2, 3, 4, \dots, n \tag{4}$$

Then the coefficient of variation of Y index score is calculated as follows:

$$V_y = \frac{\beta_y}{M_y} \quad y = 1, 2, 3, 4, \dots, n \tag{5}$$

Analytic Hierarchy Process (AHP) is a kind of decision evaluation method which can determine the weight of evaluation index. It can analyze the evaluation problem of multi-level and multi-dimensional index efficiently. The construction of judgment matrix is to determine the importance degree of each index, according to the judgment matrix can calculate and obtain the weight of evaluation index, and the comprehensive weight of the index can be calculated to order the importance degree of indicators. The key point of the selection of the two indicators is obviously different from the selection of the first level indicators. The two level indicators are integrated again on the basis of the first level indicators. The selection of secondary indicators is mainly based on the literature method and interview method. The information in the literature is combined with relevant theories, and then the indicators are modified and improved through the interview results of users. After selecting the indicators, the indicators are explained to the users so that they can understand the indicators and answer their doubts. Finally, the interview results are analyzed in order to obtain more effective information. According to the understanding of literature and relevant theories, combined with the interview content, a preliminary evaluation index of short video service based on user experience is constructed. Finally, five first-level indicators are selected (Table 1), among which each first-level indicator contains three second-level indicators (Table 2).

3 Result Analysis

3.1 Indicator Analysis

Through consistency test, the index is analyzed, and the AHP results of the index are obtained: the feature vector of the security of the first level index is 2.174, and the weight value is 43.476%; The eigenvector of first-level index reliability is 1.307, and the weight value is 26.138%. The feature vector of the first-level indicator user experience is 0.299, and the weight value is 5.973%. The feature vector of first-level index functionality was 0.451, and the weight value was 9.022%. The characteristic vector of the guarantee of the first-level index is 0.770, and the weight value is 15.390%. The maximum eigenvalue was 5.426 and CI was 0.106. Through the first-level index weight consistency test, $\lambda = 5.426$ and $CR = 0.095$ were obtained, which passed the consistency test.

Table 1. First level indicators.

The overall goal	Level indicators
Research on Evaluation Index System of Short Video Service Based on User Experience	Security
	Reliability
	Usage experience
	Functional
	Guarantee

For the weight results of secondary indicators, the weight value of privacy protection of secondary indicators of primary index security is 0.405, the weight value of identity information and authorization security is 0.479, and the weight value of market regulation is 0.115. The weight of system stability, disaster recovery strategy and sustainability is 0.600, 0.200 and 0.200 respectively. The weight of mobile demand, humanized experience and recommendation preference is 0.633, 0.260 and 0.106, respectively. The weight of backup synchronization is 0.600, the weight of feature function is 0.200, and the weight of resource search and sharing is 0.200. The weight of compensatory index is 0.655, the weight of version update is 0.186, and the weight of product brand and strength is 0.157. According to the results of consistency test, the random consistency ratio CR was all less than 0.10, so the indicators passed the consistency test.

For the comprehensive weight of indicators at all levels, in the study, the comprehensive weight is the weight factor of each second-grade rating indicator in the overall goal. Among them, the product of the weight distribution of the second-level index in the first-level evaluation index, the weight distribution of the first-level evaluation index in the overall objective evaluation and the weight distribution of the price index is the combined weight of the second-level evaluation index. And the calculation tells us.

Table 2. Secondary indicators.

Level indicators	The secondary indicators
Security	Privacy protection
	Identity information and authorization security
	Market regulation
Reliability	System stability
	Disaster strategy
	Persistent
Usage experience	Mobile demand
	Humanized experience
	Be fond of the recommended
Functional	Backup synchronization
	Features
	Resource search and sharing
Guarantee	Compensatory
	Version update
	Product brand and strength

The weight of criterion layer factor security is 0.434, the weight of dimension layer factor privacy protection is 0.405, and the comprehensive weight is 0.175. The weight of criterion layer factor security is 0.434, the weight of dimension layer factor identity information and authorization security is 0.479, and the comprehensive weight is 0.208. The weight of safety factor in criterion layer is 0.434, the weight of market regulation factor in dimension layer is 0.115, and the comprehensive weight is 0.050. The weight of reliability of criterion layer factor is 0.261, the weight of stability of dimension layer factor is 0.600, and the comprehensive weight is 0.157. The weight of reliability of criterion layer factor is 0.261, the weight of disaster recovery strategy of dimension layer factor is 0.200, and the comprehensive weight is 0.052. The weight of criterion layer factor reliability is 0.261, the weight of dimension layer factor persistence is 0.200, and the comprehensive weight is 0.052. The weight of the criterion layer factor's use experience is 0.059, the weight of the dimension layer factor's mobile demand is 0.633, and the comprehensive weight is 0.037. The weight of criterion layer factor's use experience is 0.059, the weight of dimension layer factor's humanization experience is 0.260, and the comprehensive weight is 0.015. The weight of use experience of criterion layer factor is 0.059, the weight of preference recommendation of dimension layer factor is 0.160, and the comprehensive weight is 0.009. The weight of criterion layer factor functionality is 0.091, the weight of dimension layer factor backup synchronization is 0.600, and the comprehensive weight is 0.054. The functional weight of criterion layer factor is 0.091, the characteristic synchronization weight of dimension layer factor is 0.200, and the comprehensive weight is 0.018. The weight of criterion layer factor functionality is

0.091, the weight of dimension layer factor resource search and sharing is 0.200, and the comprehensive weight is 0.018. The guarantee weight of criterion layer factor is 0.153, the compensatory weight of dimension layer factor is 0.655, and the comprehensive weight is 0.100. The guarantee weight of criterion layer factor is 0.153, the version update weight of dimension layer factor is 0.186, and the comprehensive weight is 0.028. The weight of assurance of criterion layer factor is 0.153, the weight of product brand and strength of dimension layer factor is 0.157, and the comprehensive weight is 0.024.

As for the research on the degree of unity of user opinions and the degree of coordination of user opinions, the preliminary survey results show that the degree of unity of user opinions of all indicators is greater than 4.0. The above data indicates that the evaluation index system of short video service based on users is highly approved. Preliminary research in the first round of the users of coordination degree, all user coordinate degree not greater than 0.2, the mean evaluation index differences in team for the first round of preliminary research is relatively small, so all data, according to the established evaluation index system of short video services based on user is of high rationality, the index can be used in later research.

Reliability analysis is used to study quantitative data. First, analyze the X coefficient. If this value is higher than 0.8, it indicates high reliability. If the value is between 0.7 and 0.8, the reliability is good. If the value is between 0.6 and 0.7, then the reliability is acceptable. If this value is less than 0.6, it indicates poor reliability. If CITC value is lower than 0.3, this item can be deleted; If the value of “the deleted alpha coefficient of the item” is significantly higher than the alpha coefficient, then the item can be deleted and re-analyzed.

Table 3. Questionnaire reliability test α value.

The dimension	Cronbach alpha coefficient	N of Items
Security	0.949	3
Reliability	0.959	3
Usage experience	0.932	3
Functional	0.910	3
Guarantee	0.935	3
The overall questionnaire	0.980	15

It can be seen from Table 3 that Cronbach α value of each indicator is greater than 0.8, and the overall questionnaire is greater than 0.9, so it can be verified that the reliability of the indicator system is high.

Table 4. Cronbach validity analysis.

Name	Correction Total correlation (CITC)	Item deleted α factor	Cronbach α coefficient
Privacy Protection	0.865	0.978	0.980
Identity Information and Authorization Security	0.818	0.978	
Market supervision	0.880	0.978	
System stability	0.894	0.978	
Disaster tolerance strategies	0.877	0.978	
Continability	0.903	0.978	
Mobile requirements	0.854	0.978	
Humanized experience	0.911	0.978	
Favorable Experience	0.886	0.978	
Backup synchronization	0.798	0.979	
Features Features	0.864	0.978	
Resource Search and Sharing	0.886	0.978	
Compability	0.847	0.979	
Version Update	0.827	0.979	
Product brand and strength	0.869	0.978	

Standardization Cronbach α coefficient: 0.980

As can be seen from Table 4, the reliability coefficient value is 0.980, greater than 0.9, indicating that the reliability quality of the research data is very high. For the “coefficient with deleted items”, the reliability coefficient will not increase significantly after any item is deleted, so it indicates that the item should not be deleted.

For the “CITC value”, the CITC value of the analysis items is all greater than 0.4, indicating that there is a good correlation between the analysis items and a good reliability level. In conclusion, the reliability coefficient value of the research data is higher than 0.9, which comprehensively indicates that the reliability quality of the data is high and can be used for further analysis.

Table 5. Validity analysis is the result.

Name	Factor load factor	Common degree (common factor variance)
	Factor 1	
Privacy Protection	0.884	0.782
Identity Information and Authorization Security	0.844	0.712
Market supervision	0.898	0.806
System stability	0.911	0.830
Disaster tolerance strategies	0.896	0.802
Continability	0.918	0.843
Mobile requirements	0.875	0.766
Humanized experience	0.925	0.856
Favorable Experience	0.902	0.813
Backup synchronization	0.822	0.675
Features Features	0.883	0.779
Resource Search and Sharing	0.901	0.811
Compability	0.864	0.747
Version Update	0.847	0.717
Product brand and strength	0.886	0.784
Eigenvalue (before rotation)	11.725	–
Variance interpretation% (before rotation)	78.168%	–
Accumulated variance interpretation% (after rotation)	78.168%	
KMO value	0.934	–
Bart spherical value	2182.027	–
df	105	–
p value	0.000	–

Table 6. KMO and Bartlett.

KMO and Barlett values		
KMO value	0.934	
Barlett sphericity test	Apcard square	2182.027
	df	105
	p value	0.000

As can be seen from Table 5 and Table 6, the corresponding common degree values of all the research items are all higher than 0.4, indicating that the research item information can be effectively extracted. In addition, the KMO value is 0.934, greater than 0.6, indicating that the data has validity. The variance interpretation rate values of the other 1 factor were 78.168%, and the cumulative variance interpretation rate after rotation was 78.168% > 50%. It means that the information of research items can be extracted effectively.

3.2 QoE Analysis

This time, three representative short videos, Douyin, Kuaishou and People's Short Video, were selected as the research objects. After the sample is determined, each second-level indicator is judged and scored. The evaluators of this questionnaire are mainly college students. 106 questionnaires will be released to conduct the survey and ensure its validity. In the selection of secondary indicators, the literature information and the information improved by the interviewees were integrated. Table 7 shows the information of the interviewees.

In qualitative index of the evaluation, the score of each secondary index is 5 points, score from 1 to 5 is expressed as "very dissatisfied, not satisfied, in general, satisfied and very satisfied", participants according to their own understanding, each index of the degree of satisfaction and the sample rate, and the reference description, it is concluded that each index of the score, then multiplied by the weight of the index finally concluded that the overall score, The score of each index is calculated by weighting the score results of 106 surveyors, and then multiplied by the weight of the corresponding index to get the score of each index. Finally, the comprehensive score is summarized (Table 8).

Table 7. Interviewee information.

Number	Age	Sex	Education	Way	Frequency	Usage time
1	21	Woman	Undergraduate course	Face to face	Occasionally	Half a year
2	20	Woman	Undergraduate course	Face to face	Every day	Two years
3	21	Woman	Undergraduate course	Face to face	Every day	A year and a half
4	20	Man	Undergraduate course	Face to face	Often	A year
5	21	Man	Undergraduate course	Face to face	Occasionally	Half a year
6	21	Man	Undergraduate course	Face to face	Every day	Two years

Table 8. Statistical table of study sample scores.

Level indicators	The secondary indicators	The total weight	Douyin	Kuaishou	Quanmin
Security A 0.434	Privacy protection A1	0.175	4.5	4.3	4.2
	Identity information and authorization security A2	0.208	4.2	3.8	4.1
	Market regulation A3	0.050	4.2	4.1	4.0
Reliability B 0.261	System stability B1	0.157	4.3	4.2	4.2
	Disaster strategy B2	0.052	4.1	3.9	3.9
	Persistent B3	0.052	3.6	3.8	3.9
Usage experience C 0.059	Mobile demand C1	0.037	4.3	4.0	3.9
	Humanized experience	0.015	3.8	4.1	4.0
	Be fond of the recommended C3	0.009	4.0	4.1	3.9
Functional D 0.091	Backup synchronization D1	0.054	4.1	4.2	3.9
	Features D2	0.018	3.8	4.0	4.0
	Resource search and sharing	0.018	4.1	4.0	3.8
Guarantee E 0.153	Compensatory E1	0.100	4.0	4.1	3.9
	Version update E2	0.028	4.2	4.2	4.0
	Product brand and strength E3	0.024	3.9	4.2	4.1

According to the index level design of the user-based index system of short video service, and combined with the weight calculation results of A1–E3, the first-level index is calculated and shown in Table 9:

Table 9. Statistical table of A–E results for primary indicators.

	Douyin	Kuaishou	Quanmin
Security A	1.870	1.747	1.787
Reliability B	1.075	1.058	1.064
Usage experience C	0.253	0.246	0.239
Functional D	0.363	0.321	0.351
Guarantee E	0.612	0.629	0.600
Composite scores	4.173	4.001	4.041

Based on the collated data set, the score of each secondary index of the sample product is firstly calculated, and then the weighted score is calculated for the total score. Other sample products are calculated according to this method, and Table 10 is obtained:

Table 10. Statistical table of A1–E3 score of secondary indicators.

The secondary indicators	The total weight	Douyin	Kuaishou	Quanmin
A1	0.175	0.787	0.752	0.735
A2	0.208	0.873	0.790	0.852
A3	0.050	0.210	0.205	0.200
B1	0.157	0.675	0.659	0.659
B2	0.052	0.213	0.202	0.202
B3	0.052	0.187	0.197	0.203
C1	0.037	0.160	0.148	0.144
C2	0.015	0.057	0.061	0.060
C3	0.009	0.036	0.037	0.035
D1	0.054	0.221	0.227	0.211
D2	0.018	0.068	0.072	0.072
D3	0.018	0.074	0.072	0.068
E1	0.100	0.400	0.410	0.390
E2	0.028	0.118	0.118	0.112
E3	0.024	0.094	0.101	0.098

4 Conclusion

So far, there have been many short video platforms in the market, so our research on short video service platform is mainly to understand the development status of short video and users' satisfaction with some service indicators of short video platform, further improve product quality, and maximize the advantages of short video platform. Comprehensively, we have carried out a careful study on the safety, reliability, experience, functionality and assurance of five level 1 indicators of short video. In the process of building and applying the user-based evaluation index system for short video service, it is found that there are still many deficiencies in short video and great room for improvement. The evaluation results of samples will be analyzed in the following sections. Through the analysis of sample data results can be concluded that compared to the market the same kind of short video services, douyin occupy a certain advantage in the market, get appreciated and welcomed by the masses [7], this is because the douyin convenient features, attracts more users to use, its search share function is very powerful, and to provide users open live entertainment and commercial activities, convenient and quick. Therefore, users have high requirements for the safety and reliability of short video platform, so the later maintenance in this aspect is very worthy of in-depth research. Secondly, in terms of user experience and functionality, all short video platforms are virtually the same, so there is not much difference in user experience in this aspect. Therefore, in view of our experimental results, we also put forward some suggestions, hoping to provide effective information for the study of short videos.

Acknowledgements. This work was supported in part by Xu Zhou Science and Technology Plan Project (Grant No. KC21309).

References

1. Shao, T., Wang, R., Hao, J.: Visual destination images in user-generated short videos: an exploratory study on douyin. In: 2019 16th International Conference on Service Systems and Service Management (ICSSSM), pp. 1–5 (2019). <https://doi.org/10.1109/ICSSSM.2019.8887688>
2. Wehner, N., Wassermann, S., Casas, P., Seufert, M., Wamser, F.: Beauty is in the eye of the smartphone holder a data driven analysis of youtube mobile QoE. In: 2018 14th International Conference on Network and Service Management (CNSM), pp. 343–347 (2018)
3. Zeng, L., Liu, D.: A study on the model of furniture aesthetic value based on fuzzy AHP comprehensive evaluation. In: 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, pp. 1173–1175 (2010). <https://doi.org/10.1109/FSKD.2010.5569152>
4. Zeng, Y., Lou, Z.: The new PCA for dynamic and non-Gaussian processes. In: 2020 Chinese Automation Congress (CAC), pp. 935–938 (2020). <https://doi.org/10.1109/CAC51589.2020.9327354>
5. Pinasthika, S.J., Bukhori, S., Prasetyo, B.: Hybrid lean SERVPERF-WebQual-IPA for measuring IT service quality. In: 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), pp. 13–18 (2019). <https://doi.org/10.1109/ICOMITEE.2019.8921252>

6. Gong, Q.: Notice of retraction: quality, satisfaction and loyalty: a comparison of SERVQUAL and SERVPERF for educational service in China. In: 2010 International Conference on Management and Service Science, pp. 1–7 (2010). <https://doi.org/10.1109/ICMSS.2010.5576967>
7. Wei, B., Chenxi, L.: Study on the win-win strategy of douyin and its users. In: 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE), pp. 183–186 (2020). <https://doi.org/10.1109/ICISCAE51034.2020.9236835>