

Research on the Construction of User Profiles in University Libraries and Service Optimization based on Empirical Analysis of Motivation and Behavior

¹Ziyi Yuan, ²Aoxue Zhao and ^{3,*}Xia Li,

^{1,2,3}School of Mathematics and statistics, Shandong University of Technology, Zibo, China

*corresponding author

Abstract: With the advancement of digital education in colleges and universities, libraries are transforming into "comprehensive academic services", but "unified" services are difficult to adapt to the differentiated needs of students. Existing user profiling research mostly focuses on external behavioral data and ignores intrinsic motivation and its relationship with learning outcomes, resulting in a lack of pertinence in service optimization. This study targeted college students and collected 310 valid questionnaires. Through reliability and validity analysis, it verified the motivation evaluation system including four major dimensions: academic, environment, social interaction, and atmosphere/anxiety. Then, K-means clustering (K=3) was used to divide three types of users: "full motivation-driven academic masters", "academic and social strong participants" and "low motivation wanderers". Through variance and chi-square tests, it was found that the learning efficiency and academic performance of the first two categories were significantly better than those of the third category. There was no significant correlation between the regional preferences of the three categories, but they all favored electronic/general reading rooms. Finally, the study proposes a hierarchical service optimization strategy to provide empirical support for the "user-centered" transformation of libraries.

Keywords: University Library; User Portrait; Motivation Dimension; K-Means Clustering; Service Optimization

I. INTRODUCTION

1. Research Background

The acceleration of digitalization in higher education has accelerated the transformation of libraries from traditional 'book collection and borrowing' to 'comprehensive academic services'. However, the current 'uniform' services (such as unified opening hours and differentiated resource delivery) are difficult to meet the diverse needs of students. Some students deeply use the library for learning and research, some tend to engage in socialized discussions, and others only stay for a short period to handle basic affairs. These differences arise from different motivations and behavioral habits, but they have not been fully addressed by the existing service system, leading to a mismatch of needs. Although some studies have applied user profiling technology to optimize library services, they mostly focus on external behavioral data such as borrowing frequency and spatial trajectory, ignoring the 'motivation' as the core driving factor, and have not associated with learning outcomes, making the service optimization strategy lack of specificity and effectiveness. Therefore, building user profiles from the perspective of 'motivation - behavior - effectiveness' has become a key to resolving the contradiction between supply and demand and supporting the cultivation of talents in higher education.

(二) Research objective

Taking university students as the research object, this study constructs a library usage motivation evaluation system covering four dimensions—academic, environmental, social, and atmosphere/anxiety—and tests its reliability and validity. Combining motivation data with behavioral indicators such as usage frequency and duration of stay, it uses K-means clustering to classify user types and integrates group attributes to build multi-dimensional user portraits. Through analysis of variance (ANOVA) and chi-square test, it explores differences in learning efficiency, academic performance, and regional preferences among different groups, and reveals the logic of "motivation-behavior-effectiveness". Finally, it proposes hierarchical service optimization strategies based on the needs and pain points of different groups, helping libraries achieve the "user-centered" transformation.

(三) Significance of the research

Theoretically, this study incorporates "motivation-behavior-effectiveness" into a unified analytical framework, making up for the limitations of existing research and improving the user portrait theory for university libraries. It verifies the scientificity of the four-dimensional motivation evaluation system, develops a standardized measurement tool, and optimizes the research process of "clustering + difference test", providing methodological references for similar studies. Practically, it accurately identifies three typical user groups, offering empirical basis for libraries to abandon the "one-size-fits-all" service model. The proposed personalized optimization strategies can be directly implemented to enhance the matching degree between services and needs. Additionally, it reveals the connection between motivation, behavior, and learning effectiveness, pointing out the direction for libraries to support university talent cultivation and exert their "comprehensive academic service" function.

二、 Research Design

(一) Theoretical foundations

1. User Profile Theory

The user personaError! Reference source not found. proposed by the pioneer in the field of interaction design, Alan Cooper, is a core concept that, by integrating users' attribute characteristics, behavioral data, and intrinsic psychological needs, transforms abstract user groups into virtual user prototypes with typical features. In the field of library services, the ultimate goal of introducing user personas is to achieve a shift from a 'resource-centered' service model to a 'user-centered' one. To build accurate and effective user personas, it is necessary to adhere to the basic principles of authenticity (based on real data rather than imagination), completeness

(covering multidimensional characteristics), and dynamics (updating with user changes).

An accurate user portrait relies on multidimensional data support. This includes not only relatively static user attributes such as demographic characteristics and academic background, but also dynamic behavioral data such as visit frequency, duration of stay, and preferences for resource use. It also requires in-depth exploration of psychological dimensions such as the motivations, preferences, and pain points of users entering and using the library. This study aims to build a multidimensional user portrait system by integrating students' psychological motivations for library use with their external behavioral characteristics, utilizing data-driven methods. This serves as a solid foundation for subsequent analysis of user group differences and optimization of service strategies.

II. MOTIVATION THEORY

Motivation, as an intrinsic psychological process that excites, guides, and maintains individual behavior, has a decisive impact on understanding users' library usage behavior. According to the classic Self-Determination Theory^{Error! Reference source not found.}, individual behavioral motivation can be roughly divided into intrinsic motivation and extrinsic motivation. Intrinsic motivation originates from the individual's interest, pleasure, or intrinsic satisfaction derived from the activity itself (such as pure curiosity), while extrinsic motivation is more driven by external environmental factors (such as rewards, pressure, social expectations) and the consequences of behavior.

In the specific context of university libraries, students' motivations are often complex and diverse. Based on literature review and preliminary observations, this study summarizes the core motivations driving college students to use the library into four main dimensions: academic motivation (originating from intrinsic learning and research needs, such as preparing for exams, researching literature, and completing academic tasks); environmental motivation (attracted by the physical environment of the library or driven by the need to escape from external disturbances such as dormitory environments, such as seeking a quiet and comfortable space); social motivation (based on interpersonal interaction needs, such as studying with classmates, group discussions, and meetings); and atmosphere/ anxiety motivation (influenced by the learning atmosphere of the library or driven by peer pressure, aiming to alleviate anxiety or seek a sense of psychological belonging). This multi-dimensional motivation framework aims to more comprehensively capture the driving mechanisms behind user behavior.

III. CLUSTER ANALYSIS THEORY

Cluster analysis^{Error! Reference source not found.} is one of the core methods in the fields of multivariate statistics and unsupervised machine learning. Its goal is to automatically divide similar samples into the same cluster and dissimilar samples into different clusters based on the inherent features of the sample data, without any predefined category labels, so that the samples within the cluster have high similarity and the samples between clusters have low similarity.

Among numerous clustering algorithms, the K-means clustering algorithm^[4] has gained wide application in fields such as user segmentation and market research due to its simple principle, high computational efficiency, and easy-to-understand results. The basic process of this algorithm is as follows: first, pre-set the number of clusters K; then, randomly initialize K cluster centers; next, assign each sample to the

nearest cluster center based on distance; finally, recalculate the new center of each cluster (usually the mean of all samples within the cluster), and repeat the allocation and update steps until the cluster centers no longer change significantly or reach the preset number of iterations. The ultimate goal is to minimize the sum of squared distances within clusters.

This study employs the K-means clustering algorithm to construct a portrait of university library users. The selection of the number of clusters K will be determined by combining theoretical expectations (based on predictions of user groups based on motivation types) with data-driven methods (such as the elbow method or silhouette coefficient), in order to ensure the scientificity and effectiveness of the final classification results.

(二) Research model

Based on the aforementioned theoretical foundation and focusing on the core objective of this study - constructing user profiles and verifying their differences - this research adopted a model that combines exploratory data mining with confirmatory statistical inference. Specifically, the core analytical model includes:

K-means clustering analysis: As a primary tool for exploratory analysis, this model utilizes users' multi-dimensional motivations (academic, environmental, social, atmosphere/anxiety) and key behavioral data (library visit frequency, duration of stay) to objectively and data-drivenly classify samples with similar characteristics into several user profiles.

One-way analysis of variance: As one of the confirmatory analysis tools, it is used to test whether there are statistically significant differences in the means of continuous outcome variables (such as self-reported learning efficiency, academic performance (GPA)) among the different user portraits (categorical variables) constructed.

Chi-square test: As one of the validation analysis tools, it is used to examine whether there is a statistically significant association between the distribution of the constructed different user portraits (categorical variables) and the classification result variables (such as preferred regions).

(三) Research Framework: A Three-Level Mechanism from Motivation-Driven to Behavior Differentiation

This research framework aims to reveal a complete logical chain from intrinsic motivation to external behavioral choices, ultimately forming differentiated user profiles and producing different outcomes. The framework systematically decomposes the library interaction process of college students into three core stages that are progressive, interrelated, and sequentially linked.

1. First stage: demand-driven mechanism

This stage is the starting point of the entire behavioral chain, with the core question being: 'What are the real needs of college students entering the library?' Based on motivation theory, this study believes that the driving force presents diversified characteristics, and identifies and measures four core motivations: academic motivation, environmental motivation, social motivation, and atmosphere/anxiety motivation. These motivations together constitute the initial driving force for students to enter the library.

2. Second level: behavioral choice mechanism

Different combinations of motivations, with varying degrees of intensity and focus, directly affect students' choices in external behavioral patterns. This study focuses on two key behavioral indicators, which are the external manifestations of intrinsic motivation: the frequency of visiting the library and the average duration of stay. These behavioral patterns, together with intrinsic motivation, form the basic input variables for identifying user profiles.

3.Third stage: image formation and result differentiation

Under specific motivation-behavior pattern combinations, student groups naturally differentiate into user profiles with different stable characteristics. The core of this study is to objectively identify these potential user profiles through K-means clustering analysis (RQ1). These profiles not only represent different motivation structures and behavioral habits, but also are expected to show significant differences in the final learning-related outcomes (such as perceived learning efficiency, objective academic performance) and spatial usage preferences (such as preferred areas). The variance analysis and chi-square test at this stage are used to verify the differentiation of these results (H1, H2, H3), thereby confirming the effectiveness and explanatory power of the profile construction.

(四) Research question and research hypothesis

1.Research question

Based on the aforementioned research framework, the core research question of this study is:

Based on different motivations for use (academic, environmental, social, atmosphere/anxiety) and behavioral patterns (frequency, duration), how can the user profiles of XX University Library users be categorized into typical types?

2.Research hypothesis

To test the discriminant validity of the portraits constructed through cluster analysis, that is, whether there are significant differences between different portraits on key result variables, this study proposes the following hypotheses:

There are significant differences in self-assessed learning efficiency among different user portrait groups.

H2: There are significant differences in academic performance (GPA) among different user portrait groups.

There are significant differences in the distribution of preferred areas (spatial preferences) among different user portrait groups.

三、 Research Design and Data Processing

(一) Questionnaire Design

1.Design Principles

The questionnaire design follows the principles of purposefulness, scientificity, and feasibility: guided by the research objectives, the items are designed based on motivation theory and library service scenarios; the validity of the items is tested through expert consultation and pre-research; the difficulty of answering is reduced by combining the Likert 5-point scale with objective multiple-choice questions.

2.Questionnaire Structure and Content

The questionnaire consists of three parts:

Basic information includes demographic characteristics such as gender and grade, totaling 3 questions. Core dimension scale: covering four major motivation dimensions including academic

motivation (3 questions), environmental motivation (5 questions), social motivation (5 questions), and atmosphere/anxiety motivation (5 questions), as well as behavioral indicators such as usage frequency (1 question), and duration of stay (1 question), totaling 19 questions. Results and preference indicators include the library learning efficiency (1 question), academic performance (1 question), and preferred area (1 question), totaling 3 questions.

3.Preliminary Research and Questionnaire Revision

A preliminary survey was conducted with 50 students, and two items with insufficient discriminability were removed through project analysis. Three library science experts and two statistics experts were invited to conduct a content validity evaluation, and the experts agreed that the items could comprehensively cover the research dimensions, with good content validity. The revised questionnaire structure is clear, and the item design is reasonable.

(二) Data Collection

1.Survey subjects

A stratified sampling method was employed to select college students from different grades and disciplines as the subjects of the survey. A total of 330 questionnaires were distributed, with 310 valid questionnaires collected, resulting in an effective recovery rate of 93.9%. The composition of the sample is as follows: in terms of gender, there were 152 males (49.0%) and 158 females (51.0%); in terms of grade, there were 68 first-year students (21.9%), 75 second-year students (24.2%), 82 third-year students (26.5%), and 85 fourth-year students (27.4%); in terms of discipline, there were 103 liberal arts students (33.2%), 108 science students (34.8%), and 99 engineering students (31.9%).

2.Collection method

Collect data through online methods: Distribute electronic questionnaires via the Questionnaire Star platform, ensuring anonymous responses during the survey process. Clearly inform the respondents about the purpose of the study and the use of the data, to guarantee the authenticity and confidentiality of the information.

(三) Data Preprocessing

1.Missing value handling

The multiple imputation method was used to handle missing data, and the missing situations of each item were inspected. It was found that the missing rates were all below 5%, and the imputed data met the requirements for subsequent analysis.

2.Exception value handling

Identified outliers using the Z-score method, with a criterion of $|Z| > 3$ for outlier judgment. A total of 7 outliers were found, and the method of replacing with the nearest mean was adopted to correct them, in order to avoid interference with the analysis results.

3.Data standardization

Due to the difference in scale between the motivation scale scores and the behavioral indicators, Z-score standardization is used to preprocess the data before cluster analysis. This converts all variables into standardized data with a mean of 0 and a standard deviation of 1, eliminating the influence of the unit of measurement.

(四) Reliability and Validity Analysis

1. Reliability analysis

The results of the Cronbach's alpha coefficient test for the internal consistency reliability of the scales show (Table 1): the alpha coefficients of each motivation scale range from 0.713 to 0.801, all above the acceptable standard of 0.7, with the social motivation scale having the highest alpha coefficient of 0.801, and the academic motivation scale having an alpha coefficient of 0.713. The item total statistics show that the corrected item correlations with the total score for each item are all greater than 0.49, and the alpha coefficient after deleting any item is lower than the original coefficient, indicating that the scale has good reliability and the item settings are reasonable.

Table 1: Reliability analysis results of various motivation scales

Scale Name	Item count	Cronbach's alpha coefficient	Adjusted item and total correlation range	After deleting the item, the range of alpha coefficient
Academic Motivation Scale	3	0.713	0.508-0.573	0.571-0.651
Environmental Motivation Scale	5	0.763	0.498 - 0.569	0.708-0.732
Social Motivation Scale	5	0.801	0.518-0.620	0.752-0.783
Atmosphere/Anxiety Motivation Scale	5	0.785	0.545-0.586	0.737-0.751

2. Validity analysis

The structural validity of the scale was tested using exploratory factor analysis [5]. Initially, the KMO test was conducted to assess the correlation among variables (a KMO value of ≥ 0.6 is considered acceptable, and ≥ 0.8 is considered good). The Bartlett's sphericity test was then used to determine if the variables were suitable for factor analysis (a significance level < 0.05 indicates the presence of common factors among the variables). According to the findings: the KMO values for all scales were ≥ 0.671 , and the significance levels for the Bartlett's sphericity test were all < 0.001 , meeting the conditions for factor analysis (as shown in Table 2 and Figure 1).

Table 2: Validity and applicability test of various motivation scales

Scale Name	KMO value	Bartlett's sphericity test	Significance
Academic Motivation Scale	0.671	Approximate chi-square = 175.964, degrees of freedom = 3	< 0.001
Environmental Motivation Scale	0.809	Approximate Chi-square = 331.966, degrees of freedom = 10	< 0.001
Social Motivation Scale	0.831	Approximate chi-square = 426.707,	< 0.001

Scale Name	KMO value	Bartlett's sphericity test	Significance
Atmosphere/Anxiety Motivation Scale	0.834	degrees of freedom = 10 Approximate chi-square = 372.593, degrees of freedom = 10	< 0.001

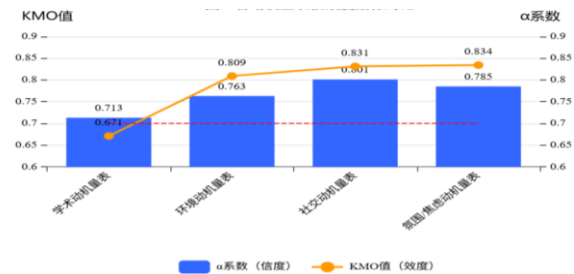


Figure 1: Comparison of reliability and validity indicators of various motivation scales

The factor extraction results show that each scale extracted one common factor, with the cumulative explained variance ranging between 51.443% and 63.579%. The analysis of common factor variance indicates that, except for one item in the social motivation scale with a commonality of 0.470, the commonality of the remaining items is all ≥ 0.494 , suggesting that the scale has good structural validity and can effectively measure the expected dimensions. Moreover, the range of common factor loadings for each scale is 0.62-0.85 (academic motivation), 0.58-0.82 (environmental motivation), 0.65-0.88 (social motivation), and 0.60-0.84 (atmosphere/ anxiety motivation), indicating that the items have a high loading on the corresponding common factors, demonstrating good convergent validity.

In summary, the KMO values, Bartlett's test of sphericity significance, communalities, and convergent validity indicators of all the motivation scales meet academic standards. This indicates that the scale can effectively measure the four dimensions of 'academic motivation', 'environmental motivation', 'social motivation', and 'atmosphere/anxiety motivation'. It provides a reliable measurement tool for the construction of user portraits based on motivation, ensuring the scientific nature of the research conclusions.

四、Construction of the User Profile of University Library Users

(一) Cluster analysis design

1. Cluster variable selection

Based on the theoretical framework and data characteristics, six clustering variables were selected: academic motivation score, environmental motivation score, social motivation score, atmosphere/anxiety motivation score, frequency of use, and duration of stay. The selected variables were tested for correlation, showing Spearman's rank correlation coefficient[6] ranging between 0.18-0.32, indicating a certain degree of independence and avoiding the problem of multicollinearity.

2. Determine the number of clusters

Adopt the elbow method and the silhouette coefficient method[7] to determine the optimal number of clusters: Plot the sum of squared within-cluster variations against the number of clusters ($K=1-5$), and a clear inflection point appears at

K=3; Calculate the silhouette coefficients for different K values, and the silhouette coefficient is the highest at K=3 (0.65), indicating that the data is suitable for being divided into 3 categories.

3. Clustering algorithm selection

The K-means clustering algorithm was selected, with an upper limit of 20 iterations and a convergence criterion of 0.001. The clustering process converged after 4 iterations, with the movement of cluster centers approaching zero, indicating that the clustering results are stable and reliable.

(二) Cluster result analysis

1. Cluster centers and group distribution

The clustering results (Table 3, Figure 2) show that there are significant differences in scores on six clustering variables among the three groups: Cluster 1 scores the highest in all motivation dimensions (4.0 points), with both usage frequency and duration at high levels; Cluster 2 scores higher in academic and social motivations (4.0 points), with environmental and atmosphere/anxiety motivations scoring moderately (3.0 points); Cluster 3 scores the lowest in all variables (2.0 points). In terms of distribution, Cluster 1 consists of 134 people (43.2%), Cluster 2 consists of 105 people (33.9%), and Cluster 3 consists of 71 people (22.9%).

Table 3: Final Cluster Centers

Cluster variable	Cluster 1	Clustering 2	Clustering 3
Academic Motivation Score	4.0	4.0	2.0
Environmental motivation score	4.0	3.0	2.0
Social Motivation Score	4.0	4.0	2.0
Atmosphere/Anxiety Motivation Score	4.0	3.0	2.0
Frequency of use	4.2	3.8	2.1
Duration of stay	4.1	3.7	2.0

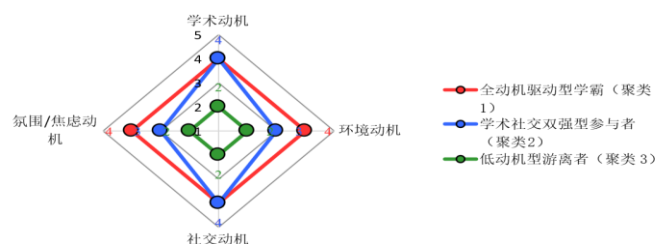


Figure 2: Radar Chart of Motivation Characteristics of Three User Groups

2. Clustering Validity Test

The variance analysis test for cluster validity shows that the between-group mean square of the 6 cluster variables is significantly greater than the within-group mean square, with all significance levels being less than 0.001. This indicates that there are extremely significant differences among the three groups on the selected variables, and the clustering results are valid.

(三) Analysis of the characteristics of three types of user portraits

1. All-motivation-driven academic overachiever (Cluster 1)

Motivation Characteristics: The four major motivations of academic, environmental, social, and atmosphere/anxiety are all at high levels, showing the 'all-dimensional strong driving' feature. This group is driven by both internal academic needs and attaches importance to the environmental conditions and social scenes of the library, while also being significantly motivated by the group's learning atmosphere.

Behavioral characteristics: High frequency of use (an average of 4.2 times per week), long duration of stay (an average of 4.1 hours per day), and a deep user of the library.

Group Characteristics: The majority of the group consists of third and fourth-year students (accounting for 68.7% of the total), with science students making up the highest percentage (41.8%), and most of them are students with moderate-to-high academic performance.

2. Academic Social Dual Strong Participants (Cluster 2)

Motivation Characteristics: Academic and social motivation scores are prominent (4.0 points), environmental and atmosphere/anxiety motivation scores are moderate (3.0 points), showing the 'core need-driven' characteristic. Academic goals and interpersonal interaction are the main reasons for using the library, with a low sensitivity to environmental conditions.

Behavioral characteristics: High frequency of use (an average of 3.8 times per week), moderate duration of stay (an average of 3.7 hours per day), and a preference for switching between the discussion area and the self-study area.

Group characteristics: The grade distribution is relatively balanced, with a slightly higher proportion of liberal arts students (38.1%), most of whom are involved in research projects or group studies.

3. Low-motivation Type Drifter (Cluster 3)

Motivation Characteristics: The scores of the four motivation dimensions are all significantly lower than the previous two categories, showing a 'weak drive' characteristic. There is a lack of clarity in academic goals, lower environmental and social needs, and minimal influence from the group atmosphere.

Behavioral characteristics: Low frequency of use (average 2.1 times per week), short duration of stay (average 2.0 hours per day), mostly brief visits.

Group characteristics: mainly composed of first and second-year students (accounting for 73.2% of the group), evenly distributed across various disciplines, and with relatively outstanding academic performance.

五、Different user group difference analysis

(一) Analysis of Learning Efficiency and Academic Performance Differences

1. Descriptive Statistics

The differences in learning efficiency and academic performance among different groups are significant (Table 4): In terms of learning efficiency, Cluster 1 (3.87 points) and Cluster 2 (3.81 points) are close in score, both significantly higher than Cluster 3 (2.21 points); in terms of academic performance, Cluster 3 has the highest score (3.15 points), with Cluster 1 (1.83 points) and Cluster 2 (1.85 points) being close to each other (Note: Academic performance is coded as 1-5 points, with lower scores indicating better performance).

Table 4: Descriptive Statistics of Learning Efficiency and Academic Performance

indicator	Cluster 1 (N=134)	Cluster 2 (N=105)	Cluster 3 (N=71)
Library study efficiency	3.87 ± 0.62	3.81 ± 0.58	2.21 ± 0.73
Academic performance	1.83 ± 0.41	1.85 ± 0.39	3.15 ± 0.52

As shown in Table 5, from the perspective of standard deviation, although the standard deviation of Cluster 3 in learning efficiency (0.893) and academic performance (0.690) is less than that of Cluster 2 in learning efficiency standard deviation (1.066), combined with the mean analysis, it can be seen that the mean learning efficiency of the low motivation

type drifters (2.21) is significantly lower than that of the previous two types, indicating that there is a situation of 'some students have low efficiency, which pulls down the overall mean'. That is, the internal differences of the group are reflected in the 'concentration of low efficiency' rather than the 'dispersion of high efficiency'. While the standard deviations of Cluster 1 (learning efficiency standard deviation 0.940, academic performance standard deviation 0.689) and Cluster 2 (learning efficiency standard deviation 1.066, academic performance standard deviation 0.744) reflect that there are certain differences in learning efficiency and academic performance within the groups of full motivation drivers and academic social dual strong types, these differences do not weaken the conclusion of 'significant differences between groups'. On the contrary, it reflects the 'reasonable fluctuations under high means' within the internal groups of the two high motivation groups.

Table 5: Descriptive Statistics of Learning Efficiency and Academic Performance among Different User Groups

		N	Average value	Standard Deviation	Standard Error	The 95% confidence interval of the mean		Minimum value	Maximum value
						Lower limit	ceiling		
How do you think your average study efficiency is in the library compared to in the dormitory or classroom?	1	134	3.87	.940	.081	3.71	4.03	1	5
	2	105	3.81	1.066	.104	3.60	4.02	1	5
	3	71	2.21	.893	.106	2.00	2.42	1	5
	Total	310	3.47	1.189	.068	3.33	3.60	1	4
What is the approximate range of your current academic performance (GPA or percentage average)?	1	134	1.83	.689	.059	1.71	1.95	1	4
	2	105	1.85	.744	.073	1.70	1.99	1	4
	3	71	3.15	.690	.082	2.99	3.32	2	4
	Total	310	2.14	.898	.051	2.04	2.24	1	4

2. Variance Homogeneity Test

The results of the Levene test for homogeneity of variance show that the Levene statistic for the library study efficiency is 0.354, with a significance level of 0.702, which is greater than 0.05; and the Levene statistic for academic performance is 0.004, with a significance level of 0.996, also greater than 0.05. Both meet the assumption of homogeneity of variance and variance analysis can be conducted.

3. Analysis of Variance Results

One-way ANOVA[8] shows (Table 5): There are extremely significant differences in the learning efficiency in the library ($F=76.630$, $P<0.001$) and academic performance ($F=96.32$, $P<0.001$) among different groups. This result verifies the research hypothesis that 'the intensity of user motivation is associated with learning efficiency and academic performance', that is, the stronger the motivation, the higher the learning efficiency and the better the academic performance. In order to further clarify the specific differences among the three groups, post-hoc multiple comparison analysis is required.

Table 6: Analysis of Variance Results

indicator	Sum of squares	degree of freedom	mean square	F value	Significance
Library study efficiency	145.574	2	72.787	76.630	<0.001
Academic performance	95.126	2	47.563	96.872	<0.001

indicator	Sum of squares	degree of freedom	mean square	F value	Significance
performance					

4. Post-hoc multiple comparisons

The results of the post-hoc comparisons using Tukey's HSD[9] method (Table 6) show that:

The learning efficiency between Cluster 1 and Cluster 2 shows no significant difference (mean difference = 0.056, standard error = 0.127, $P=0.898$), but both are significantly higher than Cluster 3 (mean difference between Cluster 1 and Cluster 3 = 1.654, $P<0.001$; mean difference between Cluster 2 and Cluster 3 = 1.598, $P<0.001$).

Academic performance: There is no significant difference between Cluster 1 and Cluster 2 (mean difference = -0.019, standard error = 0.092, $P=0.976$), but both are significantly lower than Cluster 3 (mean difference between Cluster 1 and Cluster 3 = -1.327, $P<0.001$). Considering the coding logic (lower academic performance scores indicate better performance), it can be concluded that the actual academic performance of Cluster 1 and 2 is superior to that of Cluster 3.

Table 7: Post-hoc Multiple Comparison Results (Tukey HSD)

Comparison groups	indicator	Mean difference	Standard Error	Significance
Cluster 1 vs Cluster 2	Library study efficiency	0.056	0.127	0.898

Comparison groups	indicator	Mean difference	Standard Error	Significance
Cluster 1 vs Cluster 3	Academic performance	-0.019	0.092	0.976
	Library study efficiency	1.654	0.143	<0.001
Cluster 2 vs Cluster 3	Academic performance	-1.327	0.104	<0.001
	Library study efficiency	1.598	0.150	<0.001
	Academic performance	-1.307	0.109	<0.001

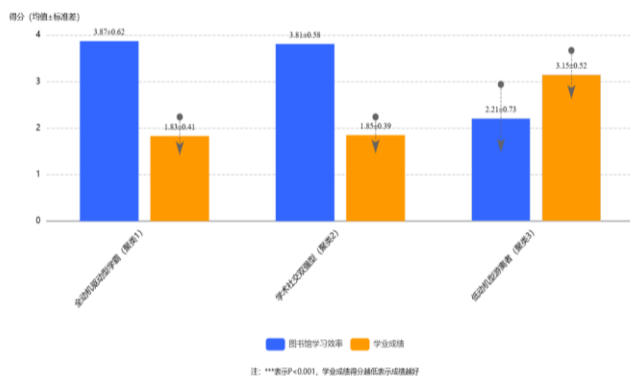


Figure 3: Comparison of Learning Efficiency and Academic Performance Among Three Types of User Groups

From the visualization results (Figure 3), it can be seen that the full-motivation-driven type (Cluster 1) and the academic-social dual-strong type (Cluster 2) both have learning efficiency scores close to 4 (high learning efficiency) and academic performance scores close to 2 (excellent academic performance). The bar charts of the two indicators for both groups are almost identical, which intuitively confirms the conclusion of 'no significant difference' in the statistical test; while the learning efficiency score of the low-motivation type (Cluster 3) is only about 2.2, and the academic performance score is about 3.15, significantly lower than the first two types, which perfectly matches the quantitative results of the post-hoc comparison.

(二) Regional preference difference analysis

1.Regional preference distribution

The regional preferences of different groups exhibit certain characteristics (Table 7): Cluster 1, Cluster 2, and Cluster 3 all have a strong preference for electronic reading rooms/general reading rooms (37.3%, 28.6%, and 28.2% respectively); the preference ratio for silent study rooms in Cluster 2 (26.7%) is higher than that in Cluster 1 (15.7%) and Cluster 3 (22.5%); the preference ratio for open spaces/café in all three groups is relatively low (all < 15%).

Judging from the visualization results (Figure 4), the fully motivated academics (cluster 1) have the highest electronic/ordinary reading room preference proportion (37.3%) among the three groups, and the height of the color block in this area in the histogram is significantly highlighted; the academic and social strong participants (cluster 2) have a silent study room preference proportion (26.7%) that is significantly higher than the other two categories, and the height difference of the color blocks is intuitively discernible; low motivation wanderers (cluster 3)'s preference for

borrowing and returning books (26.8%) is the highest among the three categories, fully echoing the table data.

Table 8: Regional Preferences Distribution of Three Groups (%)

Preferred area	Cluster 1 (N=134)	Cluster 2 (N=105)	Cluster 3 (N=71)
Silent study room (1)	15.7	26.7	22.5
Electronic/Organic Reading Room (2)	37.3	28.6	28.2
Open Space/Coffee Shop (3)	11.2	10.5	11.3
Newspaper Reading Area (4)	17.2	13.3	11.3
Borrow and return book area (5)	18.7	21.0	26.8

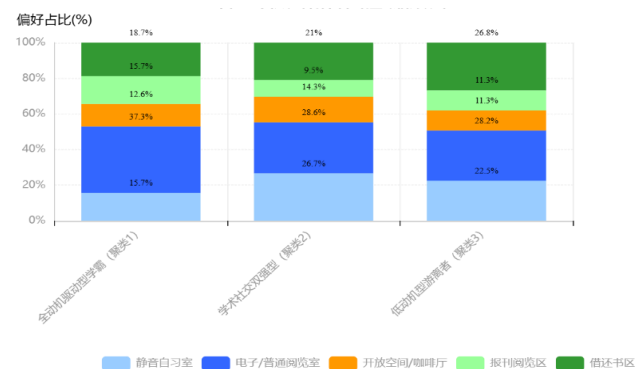


Figure 4: Analysis of Library Area Preferences for Three Types of User Groups

This regional preference difference is highly consistent with the motivational characteristics of the groups: academically driven students with full motivation, due to their clear academic goals, tend to choose electronic or regular reading rooms that are concentrated in resources and have a strong learning atmosphere; participants with strong academic and social motivations need both academic spaces and quiet study rooms with some interactivity due to their social motivation preference (convenient for quickly switching from group discussions to independent study); low motivation drifters, due to their low dependency on the library, only choose the borrowing and returning area for basic needs such as borrowing and returning books, and have a weaker preference for deep learning areas.

2.Chi-square test results

The Pearson chi-square test was used to analyze the association between clustering and regional preferences, with the results showing $\chi^2=8.114$, $df=8$, and $P=0.422>0.05$, indicating that there is no significant association between different user groups and their regional preferences. Further analysis of homogeneous subsets confirmed that the three groups did not form significant grouping characteristics in terms of regional preferences.

This result suggests that when optimizing the space of the library, it is unnecessary to allocate separate areas for specific user groups. The layout can be based on the overall positioning of 'electronic/common reading rooms as the core learning area, silent study rooms as the auxiliary interactive area, and the book borrowing and returning area as the basic functional area'. At the same time, for the low motivation groups, 'area function guidance' and 'learning scenario demonstration' can be

used to guide them to explore areas that better meet their learning needs, gradually enhancing the depth of library use.

六、Strategies for Optimizing the Services of University Libraries

Based on the portraits of three types of university library users constructed above (full motivation-driven academic masters, strong academic and social participants, and low-motivation wanderers) and their characteristic differences, combined with the analysis results of learning efficiency, academic performance, and regional preferences, this study proposed a targeted service optimization strategy from the dual dimensions of "hierarchical precision service" and "overall system optimization" to help libraries achieve "user-centered" service transformation.

(一) Optimization of services for all-mechanism-driven top students

This group has a comprehensive motivation, deep usage, strong academic needs, and demonstrates excellent learning efficiency and academic performance. The strategy should focus on enhancing the efficiency and quality of their deep learning.

1. Create a composite learning space

This group has both the need for focused study and social discussion. It is recommended to set up 'semi-open discussion areas' around silent study rooms, equipped with soundproof glass and multimedia equipment, to meet their flexible switching needs between focused study and group discussions. Optimize the hardware configuration of the electronic reading room by adding high-performance computers and professional database terminals to enhance the efficiency of resource acquisition.

2. Provide personalized academic services

Establish a 'one-to-one' subject service mechanism, regularly pushing the latest literature and academic trends according to their professional direction. Offer advanced training courses such as paper writing and data processing, inviting subject experts to give special lectures to meet their intense academic needs. Build an academic exchange platform, organize interdisciplinary seminars, and strengthen the incentive role of a group learning atmosphere.

(二) Service optimization for academic and social dual-power participants

This group combines clear academic goals with prominent social motivations, and also boasts high learning efficiency and excellent academic performance. The strategy should focus on promoting their group learning and academic interaction.

1. Optimize Social Learning Scenarios

In the core area of the library, create 'theme discussion spaces', divide the areas by academic disciplines and equip them with professional books and tools, providing a fixed location for group study. Establish an 'academic team-up' online platform to support users in posting team-up needs and study plans, enhancing the positive role of social motivation.

2. Enhanced Resource Integration Services

In pursuit of its explicit academic objectives, develop 'special topic resource packages' that integrate teaching materials, exercises, literature, and other relevant resources. Establish the position of 'subject liaison librarian' responsible for connecting with the learning needs of groups and providing customized resource search and acquisition services. Optimize the library's

navigation system, clearly marking the locations of discussion areas and resource zones, and enhancing the efficiency of space utilization.

(三) Service optimization for low-motivation wanderers

This group is mainly composed of junior students, whose motivation and frequency of use are both low, and their learning efficiency and academic performance are relatively poor. The core strategy lies in reducing the threshold for use and stimulating potential interest.

1. carry out guided services

Considering that the majority of the group are freshmen, a special activity named 'Library Newcomer Guidance' is organized to help them become familiar with the resources and spaces within the library through on-site explanations and interactive experiences. 'Resource Usage Manual' and 'Learning Scenario Guide' are also produced to lower the threshold for their use and stimulate potential demand.

2. Establish an incentive mechanism

Design a 'Library Growth Plan' to encourage increased usage through a points reward system, which can be exchanged for printing services, priority reservation for study seats, and other benefits. Invite outstanding alumni to share their library usage experience, fostering a positive learning atmosphere and gradually enhancing their motivation. For students with academic performance advantages, they can be absorbed as 'Student Service Volunteers,' deepening their understanding of the library's value through practical service.

(四) Overall service system optimization

1. Constructing a Dynamic Portrait System

Integrate borrowing data, space usage tracks, and questionnaire feedback from multiple sources to establish a dynamic user portrait database that tracks changes in user needs in real time. Develop a 'personalized recommendation' feature that automatically pushes resources and event information based on user portraits, enhancing the accuracy of service delivery.

2. Improve spatial layout and management

Based on regional preference data, reasonably adjust the area allocation and facility configuration of each functional area, increase the number of seats in the electronic reading room, and optimize the process design of the book borrowing and returning area. Adopt the 'flexible management' model, dynamically adjust the opening hours and staff allocation of each area according to different peak usage periods.

七、Conclusion and Prospects

(一) Research conclusion

This study takes the users of university libraries as the core, starting from an integrated perspective of 'motivation-behavior-effectiveness', and through empirical analysis, constructs user profiles and explores group differences, ultimately leading to the following key conclusions:

The motivation measurement system used in the library is scientific and effective. The motivation evaluation system constructed in the research, which includes four dimensions: academic, environmental, social, and atmosphere/anxiety, has been tested for reliability and validity. The Cronbach's alpha coefficients of each dimension range between 0.713 and 0.801, meeting the standard for internal consistency reliability; the KMO values are all ≥ 0.671 , and the Bartlett's sphericity test

shows $P < 0.001$, indicating good structural validity. This system can serve as a reliable tool for accurately measuring the motivation of university students to use the library and provides a standardized measurement framework for subsequent related research.

Second, the user portrait clustering results are clear and significantly different. Based on six core variables: academic motivation, environmental motivation, social motivation, atmosphere/anxiety motivation, frequency of use, and length of stay, K-means clustering algorithm was used (the silhouette coefficient was the highest when $K=3$, reaching 0.65), and three typical user groups were successfully identified: "full motivation-driven academic masters" (accounting for 43.2%), in the four major motivation dimensions and The best performance indicators are all behavioral indicators, mainly senior students; "Academic and Social Strong Participants" (accounting for 33.9%), who have outstanding academic and social motivations, high frequency of use but medium length of stay, and balanced grade distribution; "low-motivational idlers" (accounting for 22.9%), who have the lowest motivation and behavioral indicators, and are mostly junior students. The differentiation of characteristics of the three groups provides a clear basis for stratified services.

Third, the differences in learning outcomes among different groups are highly statistically significant. Through one-factor analysis of variance, it was found that there were extremely significant differences between the three groups in library learning efficiency ($F=76.630$, $P < 0.001$) and academic performance ($F=96.872$, $P < 0.001$). Post hoc multiple comparisons (Tukey HSD) further verified that there is no significant difference in library learning efficiency between "full motivation-driven academic masters" (mean learning efficiency 3.87) and "academic and social strong participants" (mean learning efficiency 3.81), but both are significantly higher than "low motivation type wanderers" (learning efficiency mean value 2.21); in terms of academic performance (the lower the score, the better the performance), the mean values of the first two groups are 1.83 and 1.85 respectively, which are significantly better than the "low motivation dissociators" (mean value 3.15), which fully confirms the internal correlation logic of "motivation-behavior-effectiveness".

Fourth, regional preferences present the characteristics of "commonality is the main factor and characteristics are supplementary". The results of the chi-square test show that there is no significant correlation between the library area preferences of different user groups ($\chi^2=8.114$, $P=0.422 > 0.05$), but there are obvious commonalities and group characteristics: the three types of groups all have a high preference for "electronic/general reading rooms" ("full motivation-driven academic masters" account for 37.3%, "academic and social strong participants" account for 37.3%) (ratio 28.6%, "low motivation type wanderers" accounted for 28.2%); at the same time, "low motivation type wanderers" had the highest preference for the "book borrowing and returning area" (26.8%), reflecting that their use of the library still remains at the basic functional level.

(二) Research innovation points

This research has achieved certain breakthroughs in the theoretical framework, research methods, and practical application, with the main innovative points as follows:

Firstly, breaking through the limitations of a single dimension, we construct an integrated analytical framework. Previous

research on university library users has mainly focused on external behavioral data (such as borrowing frequency, spatial trajectory). This study innovatively incorporates the 'motivation (internal drive) - behavior (external manifestation) - outcome (result feedback)' into a unified analytical framework. It systematically reveals the complete logical chain of users' use of the library, making up for the deficiencies of existing research that 'overemphasizes behavior and underestimates motivation' and 'overemphasizes classification and underestimates outcome'. It provides a new perspective for the theory of library user portraits.

Secondly, optimize the method of portrait construction to enhance the reliability of the results. During the portrait construction process, not only are multi-dimensional motivation data and core behavior data integrated, but also the optimal number of clusters is determined through the 'elbow method' and 'profile coefficient', ensuring the scientific nature of the clustering results. At the same time, variance analysis and chi-square test are employed to verify the discriminant validity of different portraits in terms of learning effectiveness and regional preferences, rather than merely staying at the level of 'classification description'. This significantly improves the accuracy and interpretability of the user portraits, providing a methodology paradigm for similar research.

Thirdly, focusing on practical needs, we aim to integrate theory with application. The research does not merely engage in theoretical discussions but is based on the motivation characteristics, behavioral habits, and pain points of three types of user groups. It proposes targeted hierarchical service optimization strategies, such as creating a composite learning space for 'full-motivation-driven top students' and establishing incentive mechanisms for 'low-motivation drifters'. These strategies are specific and operational, effectively bridging the gap between theoretical research and library service practice, and providing empirical support for the transformation of services centered around the user.

(三) Research limitations

Although this study has achieved certain stage achievements, it still has the following deficiencies due to the limitations of research conditions and design ideas:

The representativeness of the sample needs to be expanded. The research sample is only sourced from a single university and does not cover students from different types of institutions (such as comprehensive, science and engineering, and teacher education), as well as different levels (such as 'Double First-Class' universities, ordinary undergraduate colleges, and vocational colleges). This may limit the generalizability of the research conclusions – for example, the library usage motivation of students from science and engineering universities may differ from that of liberal arts universities, and a single sample may not fully reflect the user characteristics of different universities.

Secondly, the timeliness and dynamics of the data are insufficient. The use of cross-sectional survey design only collects user data at a specific point in time, which fails to capture the dynamic changes in user motivation and behavior over time (such as whether the motivation of freshman students transitioning to sophomores will shift from 'low motivation' to 'high motivation'), and it is also difficult to analyze the differences in user behavior at different stages such as the beginning and end of a semester. The exploration of the dynamic evolution laws of user portraits is not in-depth enough.

Finally, the factors considered are not comprehensive enough. The research and analysis framework mainly focuses on the motivations and behaviors of users themselves, without incorporating the influence of external environmental factors - such as the convenience of online learning resources (MOOCs, online databases) which may reduce students' reliance on the physical space of the library, or whether sudden public events like the pandemic will change users' habits. The absence of these factors may lead to an incomplete explanation of library usage behaviors.

(四) Prospects for the future

In response to the limitations of the research, and in conjunction with the development trends of smart libraries, future studies can be further deepened from the following directions:

Expand the scope of the sample to enhance the universality of the conclusions. Subsequent research can adopt a multi-center, stratified sampling approach, selecting universities of different types and levels for investigation. Through cross-school comparative analysis, explore the commonalities and differences in the user portraits of different university libraries, and thereby form a more universally applicable theoretical framework and service strategies. This will provide references for optimizing library services in universities of different types.

Employing a longitudinal tracking design to explore the laws of dynamic evolution. It is possible to conduct a longitudinal follow-up of the same group of students for a period of 3 to 4 years. Key nodes from freshman to senior year (such as the beginning and end of each semester) are selected for repeated research, recording the changes in user motivation, behavior, and portrait types. Analyze the impact of factors such as grade level, academic pressure, and research needs on the evolution of user portraits, and provide the library with a 'full-cycle' user service solution.

Integrate multi-source data to construct a panoramic user profile. Future research can break through the limitations of 'questionnaire data' by integrating the multi-source data of the library—such as electronic resource access records (database searches, paper downloads), online service usage data (reservation of seats, online consultation), and library entry records from the campus card system—into a panoramic user profile that integrates 'online + offline' data. This will provide a more comprehensive depiction of user needs and offer more precise data support for service optimization.

Integrating intelligent technology to promote the intelligent upgrade of services. With the application of big data and artificial intelligence in the library field, the future can develop a dynamic user portrait system based on real-time collected user data, and realize the intelligent push of resources and services through algorithms - such as pushing seminar activity information to 'academic and social dual strong participants', and pushing entry-level resource guides to 'low motivation drifters'; at the same time, combined with Internet of Things technology, optimize space management (such as real-time display of the population density in each area), and promote the continuous upgrade of library services from 'precision' to 'intelligent'.

References

- [1] Luo Shijian, Guo Huirui, Zhong Suping, et al. User Profile Generation Method for Product Design Driven by Intelligent Agents [J/OL]. Computer Integrated Manufacturing Systems, 1-20 [October 22, 2025]. <https://doi.org/10.13196/j.cims.2025.0360>.
- [2] Xu Huiwen, Zhao Qinlan, Qiu Shuang. Analysis of Multiple Influencing Factors on the Perception of Dignified Labor among Internship Nursing Students Based on Self-Determination Theory[J]. Health Vocational Education, 2025, 43(18): 81-85.
- [3] Lü Jingfang, Sun Qiumin, Dai Fen, et al. Comprehensive Evaluation of Japonica Rice Quality Based on Principal Component Analysis and Cluster Analysis[J/OL]. Zhejiang Agricultural Sciences, 1-9 [October 22, 2025]. <https://doi.org/10.16178/j.issn.0528-9017.20240835>.
- [4] Zhou Ying. User Satisfaction Analysis of Multi-source Air Conditioning Data Based on K-means Clustering [D]. Harbin University, 2025.
- [5] Umerenkova GA, Alba M, Cendales B, et al. Integrated validation of a cross-cultural emotional competence scale using Confirmatory Factor Analysis, Rasch Modeling and Exploratory Graph Analysis[J]. MethodsX, 2025, 15103627-103627.
- [6] Xia L, Ullah S, Guan L. Multivariate Extension Application for Spearman's Footrule Correlation Coefficient[J]. Mathematics, 2025, 13(9): 1527-1527.
- [7] Zhang X, Li L. Color Extraction and Artistic Matching Design Based on Silhouette Coefficient Method and Eye Tracking Technology[J]. Journal of The Institution of Engineers (India): Series C, 2025, (prepublish): 1-16.
- [8] Rahim A K, Humairi A S N S. Enhancing tribological properties of N150 mineral engine oil through MWCNT/ZnO nanoparticle additives: A comprehensive study using one-way ANOVA and experimental design[J]. Next Materials, 2025, 9(1): 1066-101066.
- [9] Chittella R, Gopalakrishnan P. A numerical simulation-based method to predict the floor-wise distribution of cooling loads in Indian residences using Tukey's honest significant difference test[J]. Advances in Building Energy Research, 2023, 17(1): 1-29.