International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-14

https://doi.org/10.70917/ijcisim-2026-0388
Article

A Study on Analyzing Students' Psychological Characteristics and Learning Styles Using Clustering Algorithms

Tian Xu [1,*] and Shige Wang [1]

[1] Department of Education, Xi'an International University, Xi'an, Shannxi, China 710077
* Correspondence author: Wfr900114@163.com

Abstract: Learning styles are also known as learning preferences or the way learners prefer to learn. Each person has his or her own characteristics and ways of dealing with attitudes, processing and perception of problems. The study collects data by distributing questionnaires to primary and secondary school students in a city, including the collection of data related to students' psychological characteristics and students' academic performance, and preprocesses the data. Subsequently, on the basis of the traditional K-means algorithm, a data analysis method of students' psychological characteristics based on the K-means clustering algorithm improved by the genetic algorithm is proposed to explore the association between students' psychological characteristics and learning styles. The results show that the K-means clustering algorithm based on genetic algorithm significantly improves the clustering speed. According to the clustering results, the students' psychological characteristics were categorized into 4 classes, and the anxiety type had the highest percentage (47.54%), and its scores were lower. There are different learning style preferences for different students' psychological characteristics, and teachers should use learning styles as a reference to design targeted teaching activities so as to improve the teaching effect.

Keywords: genetic algorithm; K-means clustering; learning style; students' psychological characteristics

1. Introduction

Online education provides students with rich learning resources, but it has been exposed in practice with problems such as high attrition rate and poor learner learning outcomes [1]. Facing that challenge, the online education field needs to develop support services with the help of student behavioral data to help improve learner learning outcomes in the online education field [2-3]. The rapid development of online education has not only broken the time and space constraints of traditional classroom learning and attracted more educators and learners, but also left a large amount of learning behavior data [4-5]. These learning behavior data provide a strong support for realizing personalized learning services based on different psychological characteristics and learning styles [6]. The introduction of learning support services can help learners to solve the learning difficulties encountered in the learning process and provide a variety of ways and forms of services, including resources, guidance, etc., which are always learner-centered, aiming at facilitating the successful completion of learning by learners, and play an important role in improving the quality of education and reducing the attrition rate [7-9]. In addition to this, understanding learners' characteristics is the basis and key to the design and implementation of learning support services, and providing learners with targeted and precise learning support is a guarantee to achieve the desired service effect and help learners complete their studies [10-13]. Among them, psychological characteristics are a kind of personalized learning ability, which not only directly affects the target activity or task efficiency, but also directly determines whether the goal can be completed, and is the comprehensive learning quality shown by students in learning activities [14-15]. In addition, learning style is the learner's learning style and learning tendency with personality characteristics, and the learning style among different learners has obvious differences [16-17]. Fully mining education big data and analyzing it can obtain the learning behavior law of different groups of learners and propose different learning support service strategies according to the different psychological characteristics and learning styles of students, thus providing data support and decision-making references to enhance learning effects [18-22].

The field of education advocates learner-centered and precise services for learners, which puts forward requirements for the means of describing students' learning characteristics. Different learners in online learning environments have large differences in the psychological level, and exploring learners' psychological characteristics is an indispensable element in student characterization. Literature [23] combines the comparison set mining and association rule mining methods to quantitatively analyze the similarity or difference data of multi-grade learning patterns, and then deduce the learning and psychological status of students in different grades, which helps teachers to develop targeted learning programs. Literature [24] establishes a comprehensive analysis model of students' classroom behavior and psychological characteristics based on an improved feature extraction algorithm, which provides a reference for timely improvement of teaching styles by identifying and predicting students' learning status in classroom teaching.

Secondly, learning style reflects a preference of learners to accept and process information in the learning process, which is often used as an important part of the user model in online learning systems. Literature [25] proposed to introduce intelligent algorithms to construct an automated identification model of students' learning styles in order to improve the identification accuracy and create a more accurate personalized learning environment in response to the problem of low accuracy of learning style identification in the past. Literature [26] designs a student learning style inference model based on back propagation neural network, which uses observed and recorded student behavioral data to infer student learning styles and makes outstanding contributions to adaptive learning and assessment of students. Literature [27] designed an automatic and reliable learning style recognition mechanism to obtain learning style labels from student behavioral data through a multi-label fusion learning style labeling approach (LSDFA) and to enhance the effectiveness of the recognition framework through a two-layer integration model (SRGSML). Literature [28] proposes mining student behavioral data to identify students' learning styles and mapping the identified learning styles to the corresponding categories of the model using the fuzzy C-mean algorithm to improve the adaptability of the style recognition system.

In summary, by integrating data related to learners' learning and labeling learners, learner characteristics can be presented more comprehensively in an abstracted way. In this process, the clustering algorithm can unify the characterization of the required information, deeply present the connection between the elements of learner characteristics and mine the hidden learning laws, which has a stronger advantage in the design of precise online learning support service strategies.

In this paper, the primary and secondary school students in an urban area were taken as the research object, and the scales such as the Temperament Scale, Personality Scale (EPQ-RSC), Anxiety Self-assessment Scale (SAS), and Depression Self-assessment Scale (SDS) were selected to collect the students' data on the dimensions of students' psychological traits, temperament, and personality. After pre-processing the collected data, the fitness function of the genetic algorithm was used to make up for the shortcomings of K-means in finding the global optimal solution, and the learners' data were analyzed in combination with the improved K-means clustering algorithm to explore the students' psychological characteristics and learning styles. Finally, the learning types of primary and secondary school students are classified according to their psychological characteristics data and visualized and analyzed, so as to intuitively discover the characteristics of learning behaviors of learners with different learning styles, in order to achieve the purpose of promoting the development of students' metacognition, assisting teachers in supervising the learning process, and enhancing the level of scientific decision-making of educational administrators.

2. Data acquisition and pre-processing

With the continuous advancement of education informatization and electrified education, a huge amount of data emerges, and how to make good use of these data in order to better facilitate teaching and education is one of the important contents of the current educational research.

From September 2022 to December 2023, primary and secondary school students in an urban area were selected for the study. The inclusion of students required full participation in the program and complete study data and informed consent. The data required for this study involved students' psychometric data as well as data related to academic performance. Through the campus communication platform, students' psychometric data can be obtained through the designated distribution of students' psychometric questionnaires on the campus information management platform, desensitized students' basic information and academic performance data can be obtained through application. In this section, the collection of student psychometric data is described.

2.1. Selection of students' psychometric scales

2.1.1. Selection of gas quality table

Using the Chan Hoi-chang Temperament Score (TS), which consists of 60 questions, this paper tested the Temperament Score for Cholestasis (S1), Temperament Score for Polycythemia (S2), Temperament Score for Mucus (S3) and Temperament Score for Depression (S4).

2.1.2. Selection of personality scales

The Chinese version of the Eysenck Personality Questionnaire-Reduced Scale (EPQ-RSC) with 48 questions was used. The Chinese version of the Eysenck Personality Questionnaire Short Form Scale (EPQ-RSC) consists of three personality dimensions and one validity scale, which are the masking (L) scale, the psychoticism (P) dimension, the extroversion (E) dimension, and the neuroticism (N) dimension, and in this paper, we tested the Extroversion Score (E) of the Personality Scale and the Neuroticism Score (N) of the Personality Scale.

2.1.3. Anxiety self-assessment scale

The SAS, with 20 questions, is designed to quantify the extent to which anxious patients experience anxiety-related symptoms. Each question was scored from 1 to 4 based on the following responses: A: none or very little of the time, B: a small amount of the time, C: quite a bit of the time, and D: the vast majority or all of the time. Fifteen questions were worded to increase the level of anxiety, while five questions (Nos. 5, 9, 13, 17, and 19) were worded to decrease the level of anxiety. The normal score is equal to the total score multiplied by 1.25. Scores ranged from 20 to 80, with normal ranges between 20 and 44, mild to moderate anxiety levels between 45 and 59, severe anxiety levels between 60 and 74, and extreme anxiety levels between 75 and 80. The reliability and validity of the SAS have been tested in several studies, and it has been standardized in Chinese populations.

2.1.4. Self-assessment scale for depression

The SDS, consisting of 20 entries, is scored and calculated similarly to the SAS scoring method. There are 10 positively worded questions and 10 negatively worded questions. Scores range from 20 to 80. The normal range is 20 to 49, mild depression is 50 to 59, moderate depression is 60 to 69, and severe depression is 70 and above. The reliability and validity of the Depression Self-Rating Scale have been evaluated in several studies and have been standardized in the Chinese population.

The above scales are scored according to the relevant rules, and the psychological, personality and temperament scores are calculated according to the corresponding methods.

2.2. Data cleansing

Generally speaking, there are more or less problems with the collected datasets, including but not limited to missing data, wrong data format, and meaningless values. The data cleansing process is to "correct" these "dirty data". For missing data, manual data completion or partial data deletion can be performed; for redundant or duplicate data, the redundant parts are deleted directly; and data with errors in content or format also require data modification or data deletion on balance. There is also a need to standardize field names in case of conflicts when using the database for integration.

First of all, there is the issue of student information. Student information is many and mixed, not only involving course information, course nature, course examination methods, course examination results, etc., a number of problems have arisen:

(1) Mixed course names, such as within the same school, mathematical analysis courses appeared in mathematical analysis I, mathematical analysis under the mathematical analysis, mathematical analysis II and other names, screening course classes, through the comparison, manually correct and unify the course name, to facilitate the subsequent operation.

(2) Confusing course classification, by analyzing the course schedule of other classes in the college, refer to the nature of other classes for the course schedule, and modify the nature of such courses.

(3) The problem of missing grades of some students, due to the lack of usual grades and other

information for reference, there is a lack of basis for completing the missing grades, so the missing grades are directly deleted here.

2.3. Data integration and conversion

Data integration refers to the integration of data from different databases. The cleaned learning style data, students' high school performance data, students' basic information data, and students' academic performance data are integrated together. Some of the integrated data are shown in Table 1.

Table 1. Partial data after integration (Take the students' 102 parts)

Student number	Class	Learning style	Course name	Test results	Course nature	Gaokao results	Gender	Peoples
102	Class 1	2	Thinking	87	General education	605	Male	Han
102	Class 1	2	English	86	General education	605	Male	Han
102	Class 1	2	Sports	77	General education	605	Male	Han
102	Class 1	2	Management	92	General education	605	Male	Han

After cleaning the data basically does not contain logical errors already, but at this time the data also exists in inconsistent units, data type inconsistency and other issues, so you need to use the numerical means of digital conversion of these field attributes. For example, the division of classes, first according to the school for the sorting number, then grade numbering, and finally add the number to distinguish between classes, such as A school number 2, A school ninth grade number 29, A school nine class number will be 291. Course categorization and so on. The processed data is shown in Table 2.

Table 2. Conversion data(Take the student 223 data as an example)

stnum	class	style	kcmc	grade	kcxz	gk	gender	mz
223	211	2	33	87	3	603	1	1
223	211	2	34	86	3	603	1	1
223	211	2	38	77	3	603	1	1
223	211	2	44	91	3	603	1	1

In summary, a total of 438 valid questionnaires were collected for subsequent analysis.

3. Improved data processing for clustering algorithms

3.1. Fundamentals of the traditional K-means clustering algorithm

K-means algorithm to complete the data clustering based on the distance, because of easy to implement, scalable performance is widely used [29]. The idea of the algorithm is as follows: set the original dataset and arbitrarily selected k original clustering centers, in turn, all the remaining samples and the original center of the Euclidean distance; the samples will be assigned to the corresponding clustering center described in the category, the allocation is based on the principle of the minimum distance, i.e., the closer to the same class; followed by the distance of the samples of each category of the average value of the distance, the new center of this category of clustering is obtained, when the error sum function is within a reasonable range to determine the final clustering center is obtained. When the error sum of squares function is within a reasonable range, the final clustering center is determined and the clustering results are output. The disadvantage of the traditional K-means algorithm is that the size of K needs to be defined manually, which increases the computational time overhead.

Define $X = \{x_1, x_2, \cdots, x_e\}$ as a collection of clustered datasets, where $x_i = (x_{i1}, x_{i2} \cdots, x_{ig})$, $x_j = (x_{j1}, x_{j2} \cdots, x_{jg})$, at this time based on formula (1) to calculate the Euclidean distance between:

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{ig} - x_{jg})^2} \tag{1}$$

The error sum of squares function shown in equation (1) is used to constrain the clustering error:

$$H = \sum_{i=1}^{K} \sum_{j=1}^{g_i} \| x_j - e_i \|^2 \tag{2}$$

In Eq. (2), the number of clustering categories is K, and the number of samples in category i as well as the sample mean are denoted as g, e_i, respectively.

3.2. Improvement of GA-K-means algorithm flow

GA-K-means is the use of genetic algorithm to select the optimal initial seed in K-means clustering.GA-K-means algorithm is shown in Fig. 1.Firstly, the system generates an initial population, which is used to search for the globally optimal initial seed, and the genetic algorithm carries out the genetic operations such as selection, crossover, and mutation on the current population, and keeps updating the population until it satisfies the stopping condition; and then, according to the output of the initial clustering center, use K-means algorithm for clustering and output clustering results.

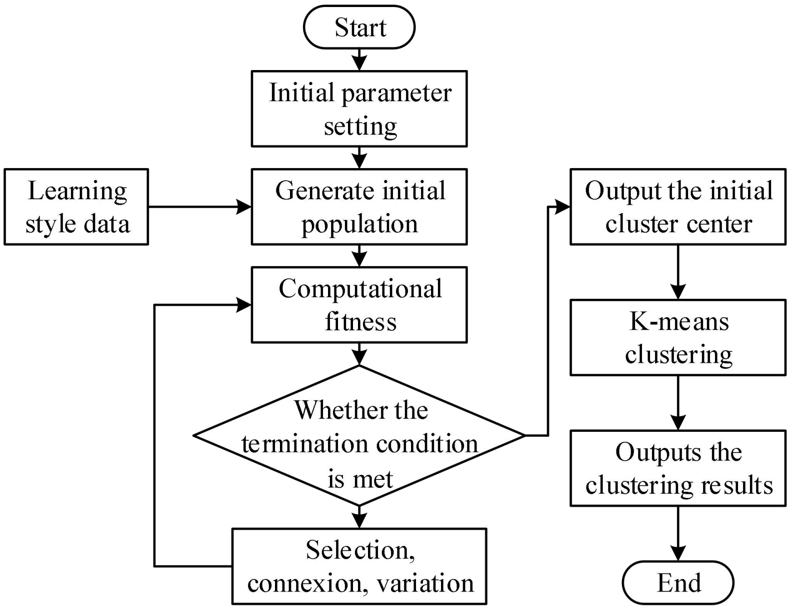


Figure 1. Ga-K-means algorithm process

3.2.1. Chromosome coding

In this paper, we adopt the real number coding method, which makes the genetic algorithm closer to the problem space, facilitates the design of specialized genetic operators, performs genetic search in a larger space, alleviates the problem of “combinatorial explosion”, and improves the computational complexity of the genetic algorithm. Each chromosome consists of K learners' psychological feature vectors, K represents the number of clustering centers, and we set the value of K to 5. Each chromosome is a 1*20 sequence of real numbers, and the structure of the chromosome is shown in Table 3, with the first 4 bits representing the psychological feature vectors of the first clustering center, and so on, and the last 4 bits representing the psychological feature vectors of the fifth clustering center.

Table 3. Chromosome coding

LS_1				**LS_2**				**LS_3**				**LS_4**				**LS_5**			
1	4	7	2	8	4	5	9	2	5	9	1	3	10	3	6	11	4	5	8

3.2.2. Initializing populations

In this paper, 100 groups of clustering centers are randomly initialized, and each group of clustering centers consists of 5 random psychological feature vectors. Genetic algorithms evolve over generations to find the best chromosome. In each generation they generate a new population from the current population by using genetic operators.

3.2.3. Adaptation function design

The fitness function is a factor that drives the genetic algorithm to converge to an optimal solution. We choose the total minimum distance function as the fitness function for chromosome strength to find the optimal initial seed for the K-means algorithm. The fitness function is defined as follows:

$$Fit = \sum_{i=1}^{k} \sum_{L_j \in G_i} | LS_j - g_i |^2 \tag{3}$$

In Eq. (3) G_i denotes the i nd cluster, L_i denotes the learner belonging to G_i, LS_i denotes the psychological feature vector of the j th learner L, and g_i denotes the cluster center of G_i. The fitness function Fit represents the sum of the distances from each learner to the respective cluster center. The smaller the fitness function Fit, the better the selection of the initial clustering center.

In this algorithm, 100 iterations are used as one of the stopping criteria. After 100 generations, the chromosome with the smallest fitness value is used as the final output. The following stopping condition is also designed: if the adaptation value of the best chromosome remains unchanged for 10 consecutive generations, the process of the genetic algorithm will stop.

3.2.4. Genetic operator selection

(1) Selection operator

Selection operation selects high quality individuals and discards low quality individuals through fitness, reflecting the biological law of "survival of the fittest". Common selection operations include: roulette selection, sorting selection, optimal individual preservation and random league selection. In this paper, roulette selection is used, and the probability P_c of a chromosome being selected is:

$$P_c = \frac{f(x_i)}{\sum f(x_i)} \tag{4}$$

In Eq. (4) $f(x_i)$ represents the fitness value of the i nd chromosome; $\sum f(x_i)$ represents the sum of fitness values of all chromosomes in the population.

(2) Crossover operator

Crossover refers to two chromosomes exchanging part of their genetic information in a certain way, thus producing two new chromosomes. Commonly used crossover methods are: single-point crossover, double-point crossover, uniform crossover and arithmetic crossover. In this paper, we use single-point crossover, in which one crossover point is randomly selected from three possible crossover points P1, P2, P3 and P4.

3.2.5. K-means clustering

(1) Use the optimal solution obtained by the genetic algorithm as the initial clustering center. (2) Calculate the distances of all data objects to these k initial clustering centers and assign the data to the class in which the closest one is located. (3) Recalculate the center of mass of each cluster that has been obtained as the new clustering center. (4) Calculate the criterion function E. If E is not satisfied, repeat steps 2 and 3 until the centers of the clusters no longer move and output the clustering results.

3.3. Analysis of clustering effects

Set the relevant parameters of the experiment: learning factors 1c and 2c are 1.5, inertia weight ω is 0.9, clustering k value is 3, population size is 100, and the number of iterations is 1,000.Comparing the data of the K-means algorithm and the GA-based K-means algorithm, each algorithm is executed for 20 times in the experiment, which yields an average running time of the K-means algorithm of 3.647527 s, while the average running time of GA-based K-means algorithm is 3.559145 s. Comparing the K-means algorithm, it can be seen that the GA-based K-means algorithm has a more significant advantage in terms of running time.

Figure 2 shows the convergence curves of the fitness values of the two algorithms, with the horizontal axis indicating the number of iterations and the vertical axis indicating the fitness value. The fitness value is an effective indicator of the quality of clustering, and the smaller the fitness value represents a higher degree of similarity of intra-class data obtained by clustering. From the figure, it can be seen that the GA-based K-means algorithm reaches stability with a smaller fitness value compared to the K-means algorithm, which indicates that the GA-based K-means algorithm has a better clustering effect.

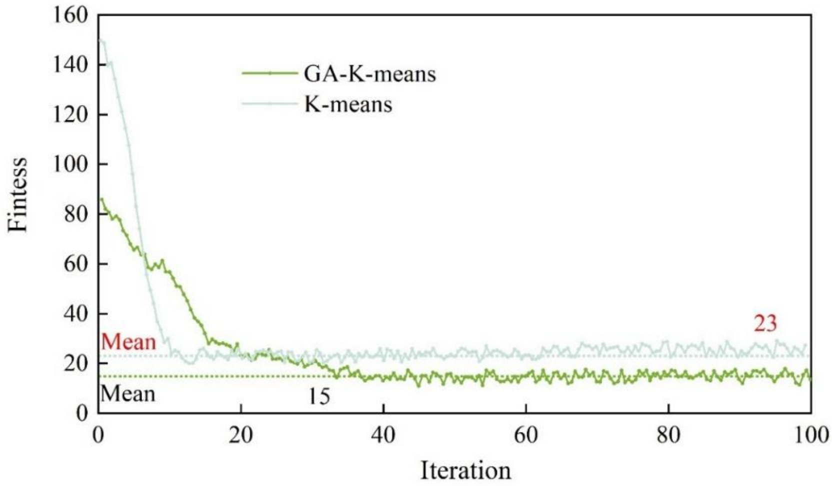


Figure 2. The two algorithms are the convergence of the logarithm of the set

4. Analysis of Students' Psychological Characteristics and Learning Styles

4.1. Analysis of psychological and personality temperament variables of respondents

4.1.1. Categorization of Respondents' Psychological and Personality Temperament Variables

The results of the factor analysis of the overall psychological and personality temperament variables of the respondents are shown in Table 4, the first four factors explained 73.74%% of the total variance, and the subsequent factors had low variance explaining power, and the roots of the characteristics were all less than 1. Therefore, the number of classifications for this group of respondents was determined to be 4.

Table 4. Analysis of the variable factor of psychological personality temperament

Constituent	**Initial eigenvalue**			**Extract the sum of squares and load**			**Rotate the squares and load**		
	Total	**Variance %**	**Cumulation %**	**Total**	**Var%**	**Cum%**	**Total**	**Var%**	**Cum%**
1	2.79	28.70	28.70	2.79	28.70	28.70	2.57	23.18	23.18
2	2.41	24.14	52.84	2.41	24.14	52.84	2.39	21.97	45.15
3	1.38	12.17	65.01	1.38	12.17	65.01	1.43	14.94	60.09
4	1.09	8.73	73.74	1.09	8.73	73.74	1.09	13.65	73.74
5	0.76	6.56	80.30						
6	0.53	5.94	86.24						
7	0.37	4.94	91.18						
8	0.34	4.67	95.85						
9	0.28	2.39	98.24						
10	0.24	1.76	100.00						

The results of the cluster analysis are shown in Table 5, where the salient features in category 1 are high extraversion and neuroticism scores, especially higher than normal neuroticism scores, and high bile, polycythemia, mucus, and depression scores, as well as high anxiety and depression scores. Category 2 was characterized by significantly higher neuroticism scores than extraversion scores, but all were within the normal range, with slightly higher scores for polycythemia and depression, and high scores for anxiety and depression; Category 3 was characterized by significantly higher extraversion scores than neuroticism scores, which were also all within the normal range, with significantly higher scores for polycythemia and mucus, and scores for anxiety and depression within the normal range; and Category 4 was characterized by significantly higher neuroticism scores than extraversion scores, and extraversion scores were lower than normal.

Table 5. Analysis of psychological characteristics variable clustering

Variable	Clustering			
	1(n=60)	2(n=131)	3(n=164)	4(n=83)
Anxiety scale score	53	58	45	55
Depression score	58	64	49	63
Personality scale extroverted score	57.45	50.21	57.75	34.48
Personality scale neurotic score	67.32	67.43	48.74	52.92
Temperament scale biliary score	17.21	4.35	5.05	0.48
Temperament scale polyblood score	18.49	7.43	10.15	5.41
Temperament scale viscosity scale	17.42	5.68	8.18	7.87
Temperament scale depression	16.81	7.96	-0.07	8.23

Cluster analysis of the final center categories of psychological personality temperament is shown in Figure 3. It can be seen from the figure that choleric scores are significantly lower than others and anxiety and depression scores are higher.

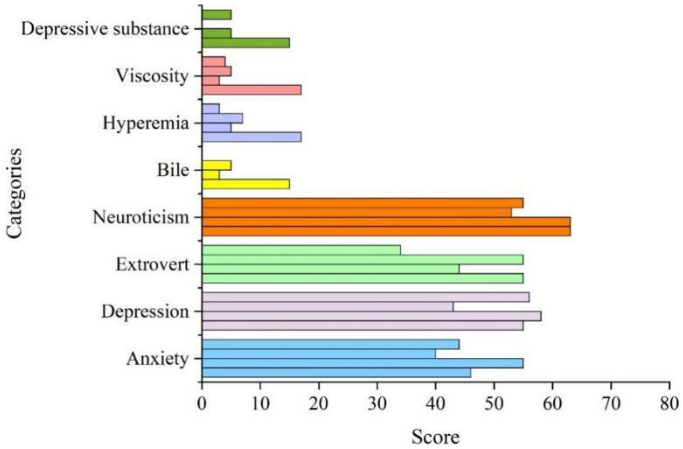
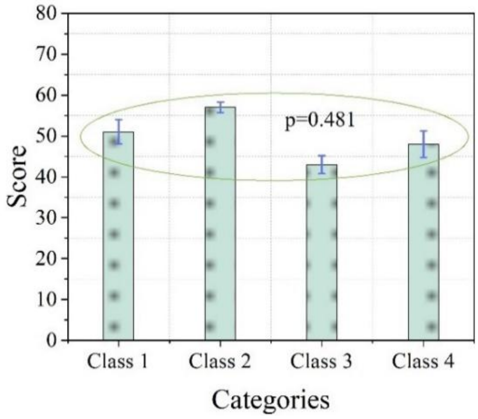


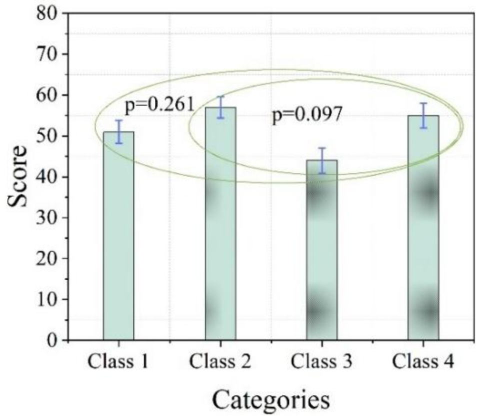
Figure 3. The cluster analysis is the final center of different psychological personality

4.1.2. Psychological and Personality Temperament Characteristics of Different Categories of Respondents

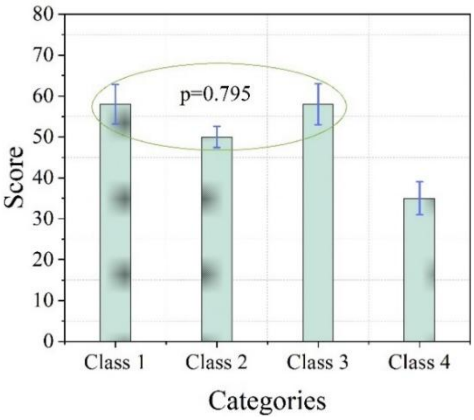
The psychological and personality temperament characteristics of different categories of respondents are shown in Figure 4, (a) to (h) for different personality temperament characteristics, respectively. The difference between two-by-two comparisons in the figure is statistically significant except for the labeling. Mild anxiety was present in both category 1 and category 4 and the difference was not statistically significant. Depression levels were not statistically significant differences between category 1 vs. category 4 and category 2 vs. category 4. The difference in extraversion scores between category 1 and category 3 was not statistically significant, and the difference in neuroticism scores between category 1 and category 2 was not statistically significant.



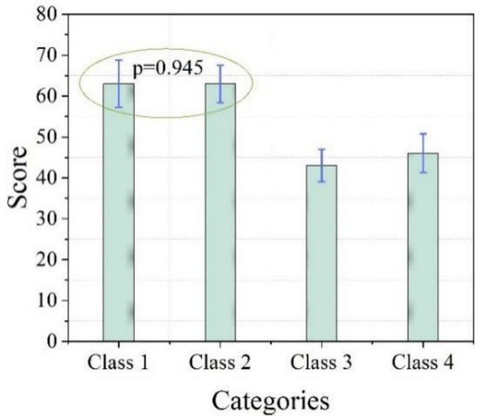
(a) Anxiety assessment scale score



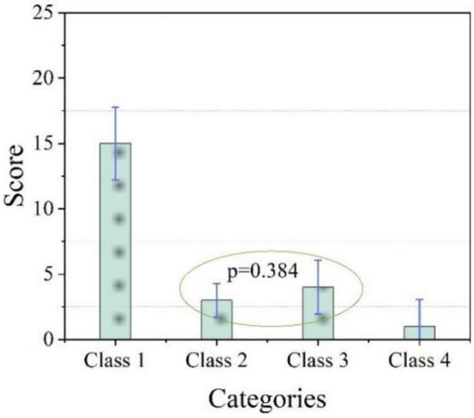
(b) Depression evaluation scale score



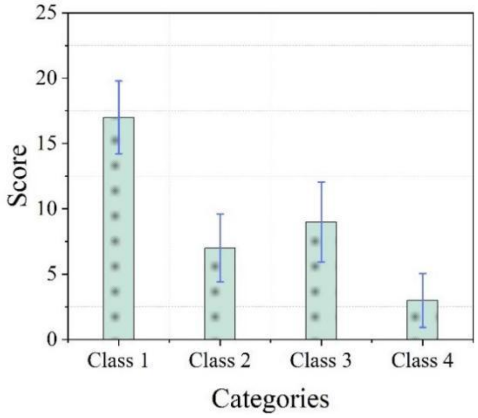
(c) Extroversion score



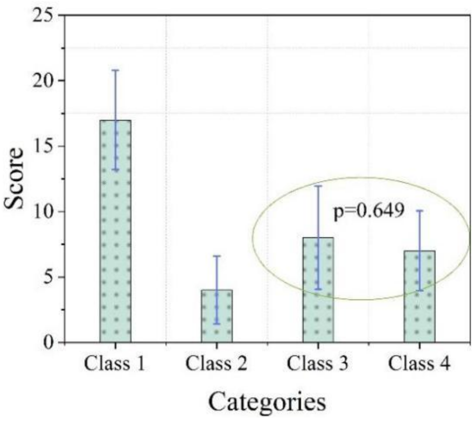
(d) Neurotic scale score



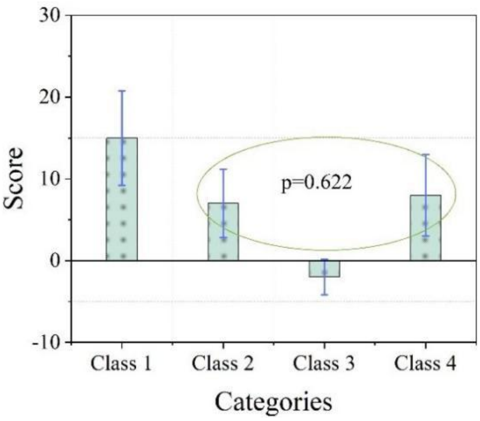
(e) Choleric score



(f) Polyblood score



(g) Viscosity score



(h) Depressive score

Figure 4. Different categories of focus, depression, personality and temperament score

4.1.3. Logistic regression analysis of different categories of respondents

The results of statistically significant ($p < 0.1$) variables of different categories analyzed by multinomial nominal logistic analysis are shown in Table 6. The reference category is category 3. It can be seen that some variables have similar effects for category 1, category 2 and category 4 relative to category 3, e.g., the effects of high stress, feeling of stomach bloating, and not having acne are positive, whereas the effects of getting good sleep, eating breakfast every day, and eating fruits regularly are negative. Some variables also have different trends in their effects for different categories, e.g., facial swelling and heartburn sensation are positive for categories 1 and 2, but negative for category 4. The opposite is true for yellowish skin, which is negative for categories 1 and 2, but positive for category 4.

Table 6. Multivariate nominal logistic analysis

Variable	Class 1			Class 2			Class 3		
	B	**Sig**	**Exp(B)**	**B**	**Sig**	**Exp(B)**	**B**	**Sig**	**Exp(B)**
High pressure	0.575	0.009	1.776	0.402	0.029	1.495	0.332	0.113	1.389
Facial swelling	0.163	0.694	1.177	0.733	0.015	2.079	-0.144	0.735	0.867
Good sleep	-0.585	0.118	0.561	-0.871	0.007	0.422	-0.112	0.715	0.894
Gastric swelling	0.032	0.906	1.033	0.589	0.005	1.795	0.351	0.136	1.422
The skin is dark	-0.594	0.155	0.556	-0.697	0.047	0.503	0.582	0.234	1.785
Have acne	0.451	0.172	1.565	0.207	0.455	1.232	0.821	0.008	2.271
Often eat fruit	-0.191	0.343	0.833	-0.175	0.294	0.844	-0.674	0.001	0.513
Heartburn	0.592	0.049	1.805	0.417	0.089	1.517	-0.192	0.542	0.832
Breakfast is eaten every day	-0.171	0.194	0.847	-0.172	0.113	0.847	-0.334	0.005	0.722

4.2. Learning style clustering results

4.2.1. Analysis of clustering results

The number of categories K of the K-means clustering algorithm needs to be set in advance, but the optimal K value needs to be determined based on the results of the contour coefficient test. Trying different values of K, the results are shown in Figure 5, the contour coefficient is the largest when the value of K is 4, so the learners' psychological characteristics can be divided into 4 categories.

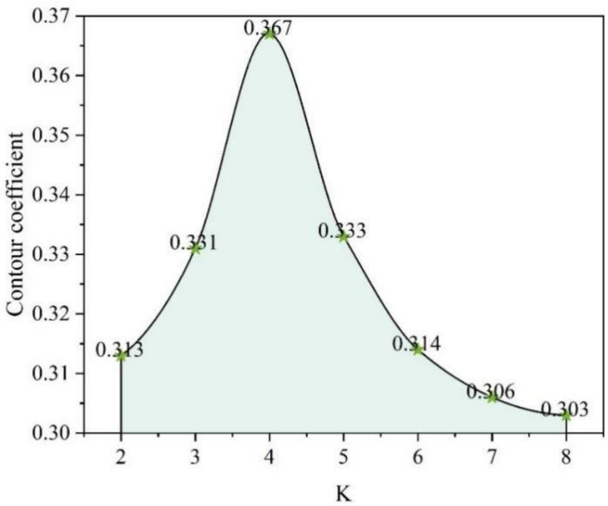


Figure 5. The contour coefficient of the different k values

The 4 types of psychological characterization indicators are shown in Figure 6, from which it can be seen that temperament type psychology has the highest score, followed by personality type psychology, and depression type psychology has the lowest score, and the ratio of the 4 types of psychology is 14.79%, 35.22%, 47.54% and 2.45%, respectively, which is unevenly distributed. ★ represents learners with less depressive mood, good learning control mood, learning behavior driven by internal motivation, so it is called positive learners. Learners represented by ◆ have low anxiety, high student learning scores, and are compliant with classroom arrangements, so they are called compliant learners. ● represents learners with higher self-personality, more extroverted learning behaviors, and students who behave positively in the classroom, so they are called extroverted learners. The learners represented by ■ have excellent learning temperament, which indicates that the students have high learning autonomy, better independent learning, and their learning behavior is driven by internal motivation, so they are called independent learners.

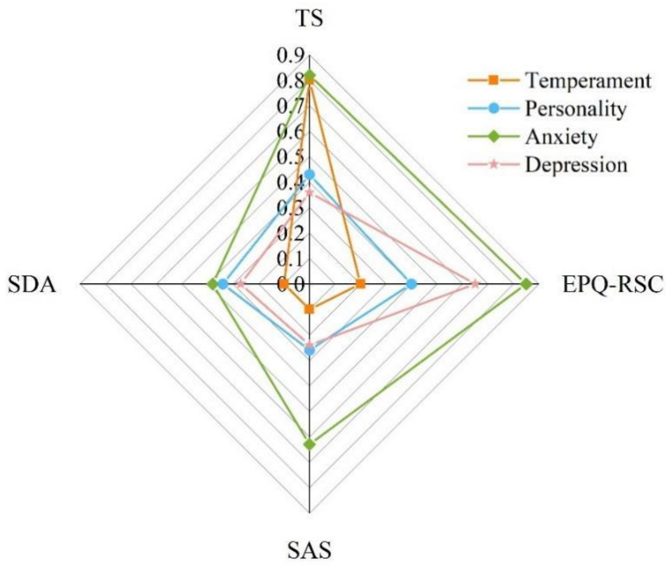


Figure 6. Psychological characteristics analysis index

4.2.2. Psychometric data for different learning styles

The results of the study of learning styles of students with different psychological characteristics are shown in Table 7. From the table, it can be seen that there is no statistical significance in the difference of learning styles of the 4 categories of learners in the anxiety category of psychological traits ($P>0.05$), and the difference of other learning styles and psychological traits are statistically significant ($P<0.05$).

Table 7. Different psychological characteristics students' learning style

Project		Learning style				Statistical value	P value
		Positive type	Adaptation type	Extroversion	Independent type		
EPQ-RSC	TS	92.84±7.45	82.23±5.11	87.68±8.58	78.49±9.24	57.317	<0.001
	Bilious temperament	12.63±2.56	14.17±4.08	11.42±3.74	9.79±4.37	167.732	<0.001
	Melancholic temperament	9(7,11)	9(8,14)	11(8,16)	54(53,61)	3.375	0.011
	Mucoid substance	22.54±2.51	22.69±2.13	22.14±2.75	20.75±2.88	65.278	0.011
	SDA	1.35	1.14	1.08	1.03	58.672	<0.001
	SAS	18.44±6.13	21.58±4.11	19.62±4.58	17.22±4.51	55.892	0.411

4.3. Visualization results of psychological characteristics of different learning styles

The psychological characteristics of students' different learning styles are shown in Figures 7 to 9, with each scatter representing the corresponding learning data of 1 learner, and different styles of learners are distinguished by different scatter shapes. In the figure, there is a positive correlation between students' personality and learning temperament, i.e., the better the students' personality, the higher the average grade, and there are positive correlations between anxiety, depression and temperament. The difference is that when anxiety breaks through 0.43, students' temperament will always be maintained at a high level, while depression has no critical point for the enhancement of temperament and appears to be smoother.

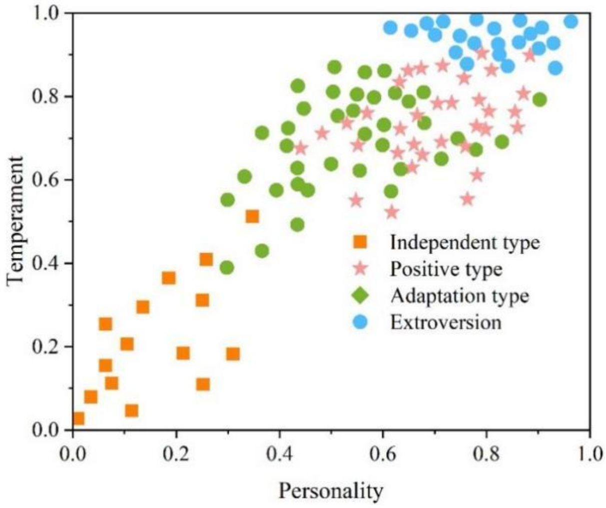


Figure 7. The relationship between personality and temperament

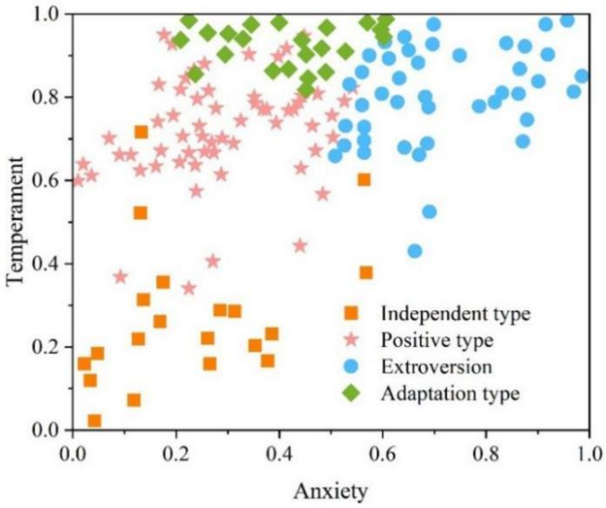


Figure 8. The relationship between anxiety and temperament

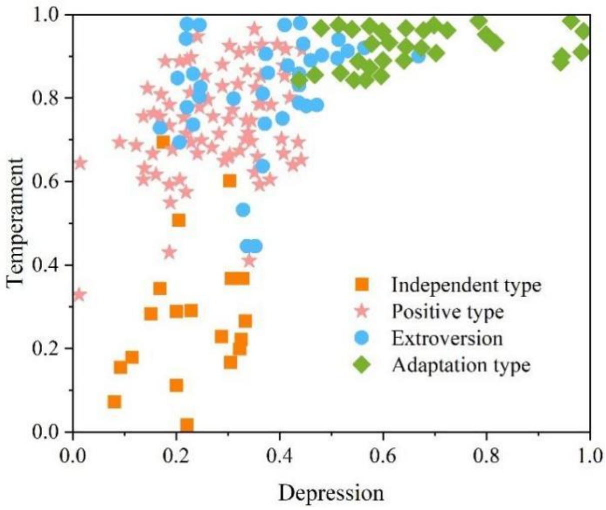


Figure 9. The relationship between depression and temperament

5. Conclusion

The study used the Temperament Scale (TS), Personality Scale (EPQ-RSC), Self-Assessment Scale for Anxiety (SAS), Self-Assessment Scale for Depression (SDS) to collect data on the psychological characteristics of primary and secondary school students in a certain city during September 2022-December 2023 and analyze the students based on the genetic algorithm of K-means relationship between psychological characteristics and learning styles. The results of the study show that GA-K-means algorithm has excellent clustering effect compared with the traditional K-means algorithm, while the psychological personality temperament characteristics of the respondents can be divided into 4 categories, and the ratios of the 4 categories of psychological share are 14.79%, 35.22%, 47.54% and 2.45%, respectively. Students of different categories differed in life habits, common symptoms and so on. Finally, students were categorized into 4 learning styles: independent, compliant, positive and extroverted. Students' psychological characteristics affect learning styles and there is a positive correlation between students' personality and learning temperament.

References

1. Falode, O. C., Chukwuemeka, E. J., Bello, A., & Baderinwa, T. (2020). Relationship between flexibility of learning, support services and students' attitude towards distance learning programme in Nigeria. European Journal of Interactive Multimedia and Education, 1(1), e02003.
2. Kintu, M. J., Zhu, C., & Kagambe, E. (2017). Blended learning effectiveness: the relationship between student characteristics, design features and outcomes. International Journal of Educational Technology in Higher Education, 14, 1-20.
3. Kapoor, G., Vostanis, A., Mejía-Buenaño, S., & Langdon, P. E. (2023). Using precision teaching to improve typically developing student's mathematical skills via teleconferencing. Journal of Behavioral Education, 1-30.
4. Shoaib, M., Sayed, N., Singh, J., Shafi, J., Khan, S., & Ali, F. (2024). AI student success predictor: Enhancing personalized learning in campus management systems. Computers in Human Behavior, 158, 108301.
5. Shemshack, A., Kinshuk, & Spector, J. M. (2021). A comprehensive analysis of personalized learning components. Journal of Computers in Education, 8(4), 485-503.
6. Ramey, D., Lydon, S., Healy, O., McCoy, A., Holloway, J., & Mulhern, T. (2016). A systematic review of the effectiveness of precision teaching for individuals with developmental disabilities. Review Journal of Autism and Developmental Disorders, 3, 179-195.
7. Regan, P. M., & Jesse, J. (2019). Ethical challenges of edtech, big data and personalized learning: Twenty-first century student sorting and tracking. Ethics and Information Technology, 21, 167-179.
8. Zeide, E. (2017). The structural consequences of big data-driven education. Big Data, 5(2), 164-172.
9. Liu, D. Y. T., Bartimote-Aufflick, K., Pardo, A., & Bridgeman, A. J. (2017). Data-driven personalization of student learning support in higher education. Learning analytics: Fundaments, applications, and trends: A view of the current state of the art to enhance e-learning, 143-169.
10. Ketui, N., Wisomka, W., & Homjun, K. (2019). Association rule mining with permutation for estimating students performance and its smart education system. Journal of Computers, 30(2), 93-102.
11. Lavrijsen, J., & Verschueren, K. (2020). Student characteristics affecting the recognition of high cognitive ability by teachers and peers. Learning and Individual Differences, 78, 101820.
12. Akareem, H. S., & Hossain, S. S. (2016). Determinants of education quality: what makes students' perception different?. Open review of educational research, 3(1), 52-67.
13. Larmuseau, C., Desmet, P., & Depaepe, F. (2019). Perceptions of instructional quality: Impact on acceptance and use of an online learning environment. Interactive Learning Environments, 27(7), 953-964.
14. Shishova, E. O. (2019). E-Learning Based on Student's Psychological Characteristics. European Proceedings of Social and Behavioural Sciences.
15. Li, Y. (2022). Analysis of psychological disorders and adaptive influence of blended learning of college students. Computational Intelligence and Neuroscience, 2022(1), 5418738.

16. Bajaj, R., & Sharma, V. (2018). Smart Education with artificial intelligence based determination of learning styles. Procedia computer science, 132, 834-842.
17. Peng, P., & Fu, W. (2022). A pattern recognition method of personalized adaptive learning in online education. Mobile Networks and Applications, 27(3), 1186-1198.
18. Xu, T., Deng, W., Zhang, S., Wei, Y., & Liu, Q. (2023). Research on recognition and analysis of teacher–student behavior based on a blended synchronous classroom. Applied Sciences, 13(6), 3432.
19. Xinhan, N. (2021). Intelligent analysis of classroom student state based on neural network algorithm and emotional feature recognition. Journal of Intelligent & Fuzzy Systems, 40(4), 7171-7182.
20. Jisi, A., & Yin, S. (2021). A new feature fusion network for student behavior recognition in education. Journal of Applied Science and Engineering, 24(2), 133-140.
21. Jingchao, H., & Zhang, H. (2021). Recognition of classroom student state features based on deep learning algorithms and machine learning. Journal of Intelligent & Fuzzy Systems, 40(2), 2361-2372.
22. Wang, Z., Li, L., Zeng, C., & Yao, J. (2023). Student learning behavior recognition incorporating data augmentation with learning feature representation in smart classrooms. Sensors, 23(19), 8190.
23. Kong, J., Han, J., Ding, J., Xia, H., & Han, X. (2020). Analysis of students' learning and psychological features by contrast frequent patterns mining on academic performance. Neural Computing and Applications, 32(1), 205-211.
24. Li, M. (2021). Recognition of Psychological Characteristics of Students' Behavior Based on Improved Machine Learning. Journal of Sensors, 2021(1), 8135942.
25. Bernard, J., Chang, T. W., Popescu, E., & Graf, S. (2017). Learning style Identifier: Improving the precision of learning style identification through computational intelligence algorithms. Expert Systems with Applications, 75, 94-108.
26. Rasheed, F., & Wahid, A. (2019). Learning style recognition: A neural network approach. In First International Conference on Artificial Intelligence and Cognitive Computing: AICC 2018 (pp. 301-312). Springer Singapore.
27. Ni, Q., Mi, Y., Wu, Y., He, L., Xu, Y., & Zhang, B. (2023). Design and implementation of the reliable learning style recognition mechanism based on fusion labels and ensemble classification. IEEE Transactions on Learning Technologies, 17, 241-257.
28. El Fazazi, H., Samadi, A., Qbadou, M., Mansouri, K., & Elgarej, M. (2018, October). A learning style identification approach in adaptive e-learning system. In International Conference Europe Middle East & North Africa Information Systems and Technologies to Support Learning (pp. 82-89). Cham: Springer International Publishing.
29. Jinyong Yang. (2024). Study of an Adaptive Financial Recommendation Algorithm Using Big Data Analysis and User Interest Pattern with Fuzzy K-Means Algorithm. International Journal of Computational Intelligence Systems,17(1),310-310.