Study on the willingness of energy conservation and emission reduction of university students in Wuhan under the background of “carbon peak” and “carbon neutral” strategy-Based on K-means clustering and principal component analysis

Yihong Xu1, Shujuan Hu1 and Rongxiao Yu1*
1college of business, Jianghan University, Wuhan, Hubei, 430056, China
*Corresponding author’s e-mail: alan.yhxu@jhun.edu.cn

Abstract. An important means to achieve the goal of "carbon peak" and "carbon neutral" is energy conservation and emission reduction. It is an indispensable part of the vision of "carbon peak" and "carbon neutral" to study the willingness of college students to save energy and reduce emissions, understand the current situation of their awareness of conservation and advocate a green low-carbon lifestyle. On the basis of literature review, 642 samples of university students in Wuhan were investigated by using probability sampling, an evaluation system of college students' willingness to save energy and reduce emission was established, SPSS was used for statistical analysis, and the reliability and validity were feasible. Based on K-means clustering analysis to analyze the cognitive characteristics of energy conservation and emission reduction of Wuhan college students, and based on the principal component analysis to comprehensively evaluate their willingness to save energy and reduce emissions, the conclusion is drawn that "it is feasible for college students to develop the willingness of energy conservation and emission reduction on the whole".

1. Introduction
In September 2020, President Xi Jinping announced at the general debate of the 75th session of the United Nations General Assembly that China would adopt strong policies and measures to enhance its national self-determination, with carbon dioxide emissions aiming to peak by 2030 and strive to achieve carbon neutrality by 2060.

As the world’s largest energy consumer and carbon dioxide emitter, China is under great pressure to cut emissions amid a worsening global climate. The achievement of the "double carbon" goal is a systematic project, in addition to the government's corresponding energy conservation and emission reduction policies, it also needs the whole society to mobilize, so that energy conservation and emission reduction will be deeply rooted in the people’s heart, and green and low carbon will become the consensus. Start with individuals, bring your own lunch boxes, handbags, refuse disposable items, make saving electricity a habit at home and at work, reduce personal energy consumption, choose bicycles or public transportation, and truly green travel, and make efforts to slow down fossil energy consumption and carbon emissions. Wuhan, as an industrial and economic hub in central China and the city with the largest number of college students, is conducive to the promotion of energy conservation and emission reduction intention. Taking college students in Wuhan as the research object, studying college students’
willingness to save energy and reduce emissions will help to guide the green low-carbon lifestyle of young college students through social practice activities, and then influence others to expand the social influence of "carbon peak" and "carbon neutrality", which will help accelerate the formation of an industrial structure, mode of production, way of life, and spatial pattern that is resource-conserving and environment-friendly, and unswervingly follow the path of high-quality development that gives priority to ecology, green and low-carbon development. It is conducive to the implementation of energy conservation and emission reduction policies and the goal of “carbon peak” and “carbon neutral”.

2. Literature review

2.1. About the meaning of "carbon peak" and "carbon neutral"
Hu Angang (2021) pointed out that "carbon peak" refers to the carbon dioxide emissions reached an all-time high, and then went through a platform period into a continuous decline process, which is the historical inflection point of carbon dioxide emissions from increase to decrease and marks the decoupling of carbon emissions from economic development, the goal of reaching the peak includes peak years and peak value.

As for the concept of "carbon neutral", there are two different views on "climate level" and "carbon emission" in academic circles. The first is the Special Report on Global Warming of 1.5°C released by IPCC (2018) based on the climate dimension, which states that net global greenhouse gas emissions can stabilize global warming within a certain range and reach a certain level between atmospheric greenhouse gas emissions and absorption; The other is based on a perspective of carbon emissions, that is to achieve carbon emission neutrality. Li Nu et al. (2021) proposed that "carbon neutrality" refers to enterprises, groups or individuals estimate the total amount of greenhouse gas emissions directly or indirectly generated within a certain period of time, through afforestation, energy conservation and emission reduction, etc. to offset their own CO₂ emissions and achieve “zero emissions”.

2.2. About the research on policies and influencing factors on energy conservation and emission reduction
"Energy conservation and emission reduction" comes from the 11th Five-Year Plan for National Economic and Social Development of the People's Republic of China, which refers to energy conservation and reduction of environmental harmful emissions. Li Kongyan (2017) believes that energy conservation and emission reduction have a broad and narrow meaning. Broadly speaking, energy conservation and emission reduction refers to saving material and energy resources and reducing waste and environmental harmful emissions. In a narrow sense, it refers to energy conservation and reduction of environmental harmful emissions.

As for the research on energy conservation and emission reduction policy, foreign scholars mainly focus on policy formulation and implementation. According to public finance theory, Decaniyo (1993) indicated that the private sector is not actually actively implementing energy conservation and emission reduction policies. At the same time, some foreign scholars explore the coping strategies of energy saving and emission reduction through the model method. Hassler et al. (2012) argue that South Korea, India, and the United States are shifting towards energy-efficient innovations in industrial technology that can help energy-intensive industries operate efficiently. Klesmann et al. (2013) analyzed the implementation and landing of policies related to risk-sensitive renewable energy technologies, and concluded that they can control the performance of financing costs and compress the cost budget under the premise of ensuring the effectiveness.

The research on energy conservation and emission reduction policy by domestic scholars mainly focuses on the effect and potential and putting forward corresponding policy suggestions. Jiang Haiyang (2017) made a detailed analysis of the energy conservation and emission reduction measures and pointed out the existing problems in the evaluation results of the paper. Jia Junjun (2018) takes energy conservation and emission reduction as the starting point and combines qualitative and quantitative research methods to evaluate the effectiveness and feasibility of existing and future energy conservation
and emission reduction policies. Cao Wei (2020) studied the regional differences and convergence of energy conservation and emission reduction efficiency in Shaanxi Province, and used models to analyze them in various cities in Shaanxi Province.

For the factors affecting energy conservation and emission reduction, the foreign scholar Marxa Mendiluee (2007) used the exponential decomposition method to analyze the trend of energy intensity change in the 15 EU countries from 1995 to 2004, and concluded that the trend experienced the process of first rising and then falling. Lynn Price (2011) evaluates the achievements of energy conservation and emission reduction during the 11th Five-Year Plan period by qualitative and quantitative methods, and believes that the improvement of energy efficiency is the direct cause of the reduction of energy consumption per unit GDP during the 11th Five-Year Plan.

Xu Shan (2016), a domestic scholar, uses the theory of applying the interpretation structure model to break down the hierarchy of the influencing factors and analyze the path of their effect on the system. Luo Nengsheng (2017) shows that environmental governance investment, proportion of three productions, technical level and other factors are conducive to the improvement of energy conservation and emission reduction efficiency, while urbanization level and energy structure inhibited the efficiency. Lu Rui (2019) found that technological progress can play a greater role in energy-intensive industries, while in low-energy industries, technological efficiency is the focus of energy efficiency improvement.

2.3. About the review of research methods
Tryon first proposed the concept of clustering in 1939, pointing out that clustering is a multi-statistical method that uses multiple indicators for mathematical classification and was used in biology and ecology research in the early stage. Guo Min (2020) believes that clustering is mainly divided into three categories, system clustering is mainly inter-sample clustering of small samples and variable clustering. Dynamic clustering is suitable for clustering among samples when the sample is large, and K-means method is commonly used. Ordered sample clustering refers to clustering of samples with arranged order, which requires that the samples with adjacent order can be clustered into a class. Zeng Chi (2020) evaluates the performance of listed companies by studying the specific defects of K-Means clustering algorithm.

Li Jingping (2015) proposed that the dimension of the original variable is too high in the actual analysis, which is not convenient for the analysis of the problem. Therefore, information on multiple original variables can be profiled by a few composite indicators and minimizes the loss of information. Xie Yuchu (2016) believes that the principal component analysis is to study how to reduce the index data. Zhang Di (2018) also proposed that the principal component analysis belongs to the multi-statistical analysis, which transforms the original related variables into unrelated variables. Although the number of variables selected is small, but are highly representative, that is, the principal component.

In summary, under the strategic background of "carbon peak" and "carbon neutral", China has formulated corresponding energy conservation and emission reduction policies. On the research of energy conservation and emission reduction policy, foreign scholars mainly focus on the formulation and implementation of policies, after the formulation of policies, they pay attention to the effect of implementation and implementation; Domestic scholars’ research is later than them, based on the study of foreign scholars, their research mainly focuses on the discussion of the effect and potential of energy conservation and emission reduction and puts forward corresponding suggestions. For the factors affecting energy conservation and emission reduction, scholars at home and abroad will use models to evaluate the impact of various factors on energy conservation and emission reduction, and analyze the path of policy implementation. However, the researches on energy conservation and emission reduction of college students, especially in Wuhan area, are rare.

3. Study samples and data sources
Taking university students in Wuhan as the research sample, a random sample was selected from 1.1256 million college students in Wuhan. A total of 667 questionnaires were sent out, 654 questionnaires of
which were actually recovered and 642 of which were valid. The effective rate of questionnaires was 98.17%.

The calculation formula of the optimal sample size \( n_0 \) is:

\[
 n_0 = \frac{u^2PQ}{d^2} \left( 1 + \frac{1}{N} \left( \frac{u^2PQ}{d^2} - 1 \right) \right)
\]

In the above formula, \( N \) is the population quantity, \( u \) is the bilateral \( \alpha \) quantile of the standard normal distribution, \( p \) is the sample scale, and \( d \) is the absolute allowable error.

According to the preliminary survey results \( p=0.4, d=0.04 \), in practice, if \( p \) is near 0.5, the sample size can be estimated according to the total variance at the peak of \( p=0.5 \). The optimal sample size can be approximated as:

\[
 n_0 = \frac{u^2p(1-p)}{d^2} = 600
\]

Assuming that the sample recovery rate is 90%, the actual sample size of the questionnaire issued is:

\[
 n = \frac{n_0}{90\%} = 667
\]

The 642 questionnaires obtained in the actual survey were stratified according to the specialty nature of the samples. The specific specialty distribution of the samples is shown in Table 1.

| Layer                | Right of layer | The sample size |
|----------------------|----------------|-----------------|
| Art                  | 0.1017         | 66              |
| Literature and History| 0.0904         | 58              |
| Institute of Technology| 0.3616       | 232             |
| Economic Management  | 0.2034         | 131             |
| Medicine             | 0.0678         | 43              |
| Political law        | 0.0452         | 29              |
| Other                | 0.1299         | 83              |
| Total                | 1              | 642             |

3.1. Differentiation test

The differentiation index of objective questions is calculated as follows:

\[
 D = P_n - P_1
\]

Among them \( D \): Discrimination index; \( P_n \): The passing rate of a problem in a high score group; \( P_1 \): The passing rate of a problem in a low-scoring group.

The differentiation index of subjective questions is calculated as follows:

\[
 D = \frac{Q_h - Q_l}{n(S_H - S_l)}
\]

Among them \( Q_h \): total score of high group candidates, \( Q_l \): total score of low group candidates, \( n \): number of high groups (or low groups), \( S_H \): highest score, \( S_l \): lowest score.

Based on SPSS, the discriminative degree of objective question is 0.34 and that of subjective question is 0.27. It is concluded that the degree of differentiation of objective questions is significantly higher than that of subjective questions.
3.2. Reliability test
In this experiment, because of the small number of questions, we need to use formula by using the coefficient method $\alpha$:

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum_{i=1}^{K} S_i^2}{S^2}\right)$$

$\alpha$: reliability coefficient; $K$: Number of questionnaire questions; $S^2$: total variance; $S_i^2$: Variance of question $i$.

Table 2. Cronbach'a coefficient of each variable.

| Variable               | Cronbach'a |
|------------------------|------------|
| Cognitive status       | 0.635      |
| Educational Impact     | 0.673      |
| Family atmosphere      | 0.746      |
| Social factors         | 0.728      |
| Economic factors       | 0.614      |

After calculation, questions in the questionnaire reached the standard, the reliability test of this questionnaire was reliable.

3.3. Validity test
This article verifies the accuracy of the questionnaire by content validity:

Table 3. Performance subscale.

| Question               | Mean    | Standard deviation | Factor load factor | Eigenvalue | Cumulative variance contribution rate (%) |
|------------------------|---------|--------------------|--------------------|------------|-------------------------------------------|
| Cognitive status       | 3.232   | 0.592              | 0.671              | 5.124      | 64.055                                    |
| Educational Impact     | 3.595   | 0.963              | 0.762              | 5.37       | 54.6                                      |
| Family atmosphere      | 3.202   | 0.588              | 0.856              | 5.267      | 72.45                                     |
| Social factors         | 3.667   | 0.553              | 0.563              | 4.54       | 64.99                                     |
| Economic factors       | 3.559   | 0.823              | 0.873              | 5.645      | 63.8                                      |

Table 4. Validity Checklist.

| KMO sampling fitness quantity | Bartlett Spherical Test |
|-------------------------------|-------------------------|
| Approximate chi-square        | Degree of freedom       | Significance        |
| 0.72                          | 3145.33                 | 100                 | 0.000        |

Because 0.7 < KMO < 0.8, it is more suitable for factor analysis, Bartlett spherical test is less than 0.05, the overall validity of this questionnaire is good.
4. Analysis of cognitive characteristics of energy conservation and emission reduction of college students based on K-means clustering analysis

In order to further understand the cognitive characteristics of different types of college students on energy conservation and emission reduction, the college students interviewed were clustered and named with the results of clustering.

4.1. Variable selection

Before clustering analysis, consider the following 6 variables for cognitive characteristics of energy conservation and emission reduction for college students (see Table 5).

| Variable               | Description                                                                 |
|------------------------|-----------------------------------------------------------------------------|
| Cognitive degree       | Your understanding of the concept of energy conservation and emission reduction |
| Interest motivation     | Do you have the habit of actively understanding energy conservation and emission reduction? |
| Watch the willingness   | If there are programs about energy conservation and emission reduction, would you like to watch them? |
| Payment will           | Are you willing to accept low-carbon products?                               |
| Family atmosphere      | Does your family have the habit of saving energy?                           |
| Inheritance will       | When you find someone in your life who destroys the environment, your attitude is |

4.2. Result analysis

Clustering analysis results were obtained by SPSS, and four types of college students were obtained. This paper named and described the four groups (see Table 6).

| Clustering Characteristics. | A  | B  | C  | D  |
|-----------------------------|----|----|----|----|
| Cognitive degree            | 1.0| 1.5| 4.3| 7.6|
| Interest motivation          | 1.0| 2.1| 6.0| 7.5|
| Watch the willingness        | 3.5| 7.0| 5.4| 8.4|
| Payment will                | 1.0| 7.1| 7.1| 7.7|
| Family atmosphere           | 1.0| 8.5| 6.0| 8.9|
| Inheritance will            | 5.5| 7.3| 7.4| 7.6|

According to the cognitive classification tables of the four groups in Table 6, the descriptive analysis is as follows:

Class A college students are not aware of carbon emission reduction, but still to a certain extent willing to watch the program and inheritance; Class B students clearly indicate that they will pass on energy conservation and emission reduction, and Class B students are more positive than Class A, but they are not strong in subjective initiative in cognition and interest of energy conservation and emission reduction. Class C students have a certain degree of awareness of energy conservation and emission reduction, and they think it is necessary for contemporary college students to continue pass on energy conservation and emission reduction related knowledge and environmental protection habits. Class D students have a high degree of awareness of energy conservation and emission reduction. In general, D students are outstanding in the knowledge and inheritance of energy conservation and emission reduction and the role models for others.
In summary, for the questionnaire cluster analysis, get the proportion of A, B, C, D four groups of college students as shown in Figure 1. Class B students accounted for the largest proportion, followed by Class C students, indicating that the current cognitive status of college students on carbon emission reduction is not optimistic, and family communication atmosphere is also low. Class A and D is relatively less, which means college students in neutral and very positive attitude are small in all respondents. That is, college students’ attitude of energy conservation and emission reduction is understanding some, but on the willing of inheriting it, the subjective initiative is low.

Figure 1. Comprehensive analysis of four types of college students.

5. Comprehensive evaluation of college students’ willingness to save energy and reduce emission

5.1. Establish an evaluation index system of college students' willingness to save energy and reduce emissions

After the relevant investigation and data analysis for college students’ willingness to save energy and reduce emissions, it was found that the correlation between certain topics and certain factors was significant, and different from the initial establishment factors, and decided to extract five core factors according to the current situation of college students in this regard, which constituted the evaluation index system of college students' willingness to save energy and reduce emissions, as shown in Table 7.

| Core factors        | meaning                                                                 |
|---------------------|-------------------------------------------------------------------------|
| Personal cognitive  | Variables that measure college students' understanding of energy conservation and emission reduction |
| Development factors | Variables that measure the impact of development on college students’ willingness to save energy and reduce emissions |
| Behavior consciousness | Variables that measure the influence of behavior consciousness on college students' willingness to save energy and reduce energy |
| Propaganda factors  | Variables that measure the influence of propaganda factors on college students’ willingness to save energy and reduce energy |
| Economic factors    | Variables that measure the influence of economic factors on college students’ willingness to save energy and reduce energy |

5.2. Comprehensive evaluation of college students’ willingness to save energy and reduce emissions based on principal component analysis

The principal component analysis was used to reduce the dimension of the questions in the questionnaire. To preliminary sorting of the original data, using SPSS statistical software and principal component analysis for evaluation of 12 class variable statistical correlation analysis, through sampling data, characteristic value of characteristic vector and the cumulative contribution rate to find five principal component factors that affects the willingness, and on this basis for quantitative evaluation of college students’ willingness to save energy and reduce emissions.
Based on 12 willingness evaluation indicators, SPSS factor analysis and principal component analysis were used to obtain Figure 2 and Table 8. Figure 2 is a gravel diagram, it can be found that basically the first five factors have covered most of the variance, so extract the first five factors as the main component, and then calculate the contribution rate of the principal component, component matrix and eigenvector matrix.

![Figure 2. gravel diagram of each factor.](image)

Table 8. Eigenvector, variances and cumulative contribution rates of each component.

| Composition | Initial eigenvector | the sum of squares of the extract load |
|-------------|---------------------|--------------------------------------|
|             | Total | Percentage of variance | Cumulative% | Total | Percentage of variance | Cumulative% |
| 1           | 2.949 | 24.578 | 24.578 | 2.949 | 24.578 | 24.578 |
| 2           | 1.728 | 14.396 | 38.974 | 1.728 | 14.396 | 38.974 |
| 3           | 1.138 | 9.480 | 48.454 | 1.138 | 9.480 | 48.454 |
| 4           | 0.997 | 8.309 | 56.763 | 0.997 | 8.309 | 56.763 |
| 5           | 0.960 | 8.001 | 64.765 | 0.960 | 8.001 | 64.765 |
| 6           | 0.914 | 7.618 | 72.383 |          |        |            |
| 7           | 0.783 | 6.526 | 78.908 |          |        |            |
| 8           | 0.701 | 5.842 | 84.751 |          |        |            |
| 9           | 0.550 | 4.584 | 89.335 |          |        |            |
| 10          | 0.474 | 3.947 | 93.282 |          |        |            |
| 11          | 0.428 | 3.566 | 96.848 |          |        |            |
| 12          | 0.378 | 3.152 | 100.000|          |        |            |

The eigenvector matrix can reflect the principal component load of the other indexes. As can be seen from Table 9, the first to the fifth principal component are personal cognition, behavioural awareness, economic factors, development factors and publicity factors.

![Table 9. Eigenvector matrix.](image)
Assume the data of the 12 variables mentioned above after standardized processing are in order, then the following principal component expression can be obtained according to Table 9: $x_1$ to $x_{12}$

$$y_1 = 0.212x_1 + 0.236x_2 - 0.063x_3 + 0.046x_4 + 0.231x_5 + 0.198x_6$$

$$y_2 = -0.273x_1 - 0.252x_2 + 0.138x_3 + 0.27x_4 - 0.108x_5 + 0.215x_6$$

$$y_3 = 0.109x_1 + 0.131x_2 + 0.103x_3 + 0.439x_4 - 0.066x_5 - 0.229x_6$$

$$y_4 = -0.096x_1 - 0.122x_2 + 0.521x_3 + 0.170x_4 + 0.025x_5 + 0.295x_6$$

$$y_5 = 0.207x_1 + 0.091x_2 + 0.663x_3 - 0.504x_4 - 0.051x_5 + 0.033x_6$$

After calculating the above five principal components based on quantified variable values, the variance contribution rate of each principal component is regarded as its weight to construct comprehensive evaluation index of college students’ willingness to save energy and reduce emission:

$$S = 0.24578y_1 + 0.14396y_2 + 0.0948y_3 + 0.08309y_4 + 0.08001y_5$$

Based on the above formula, the intensity of the collected sample data was evaluated and comprehensive evaluation score of 642 respondents were calculated, SPSS was used for statistical data, Matlab was used for plotting, and the intensity distribution diagram as shown in Figure 3 could be obtained.
As known in Figure 3, the comprehensive score of college students' willingness to save energy and reduce presents a normal distribution, with more people scoring around 4.5 points, that is, most students have average willingness, while fewer have too high or too low willingness, and those with high willingness are more than those with low willingness.

6. Conclusion
Empirical research shows that, on the whole, the score distribution of respondents' willingness to save energy and reduce emissions presents a normal distribution form. Therefore, it can be concluded that college students in Wuhan have initially formed their willingness to save energy and reduce emission. The following conclusions were drawn: (1) Respondents’ cognition of energy conservation and emission reduction is low, and their characteristics are significantly different; (2) It is generally feasible for college students to develop the willingness to save energy and reduce emissions; (3) college students can accept most effective behavior and dissemination of energy conservation and emission reduction. In addition, measures can be taken from individuals, families, schools, the government to improve college students’ the willingness to save energy and reduce emission.

References
[1] Xi Jinping. Going forward, embark on a new global journey to address climate change - Speech at the Climate Ambition Summit[N], People's Daily, 2020-12-13 (01).
[2] Hu Angang. China's Goal and Main Ways to Achieve Carbon Peak by 2030[J]. Journal of Beijing University of Technology (Social Sciences Edition), 2021,21(03):1-15.
[3] IPCC. Special report on global warming of 1.5°C[M]. Cambridge: Cambridge University Press,2018.
[4] Li Kongyan. Study on the dilemma, problems and countermeasures of energy conservation and emission reduction in Inner Mongolia Autonomous Region under the green development perspective[D]. Inner Mongolia University, 2017.
[5] Cao Yu. Study on energy conservation and emission reduction potential, regional efficiency differences and convergence in Shaanxi Province[D]. Xi'an University of Architecture and Technology, 2020.
[6] Hassler J, P. Krusell and C Olovsson. Enemy-Saving Technical Change[J]. NBER Working Paper 2012,18:20-28.
[7] Klessmann C, Rathmann M, Jager D D, et al. Policy options for reducing the costs of reaching the European renewables target[J]. Renewable Energy,2013,57(3):390-403.
[8] Jiang Haiyang. Model for the cooperation optimization of coal-fired power industry chain to promote energy conservation and emission reduction in power generation[D], Beijing North China Electric Power University, 2017.
[9] Jia Junjun. Energy conservation and carbon reduction in the residential sector: consumer behaviour, energy efficiency measures and policy tools[D]. University of Science and Technology of China, 2018.
[10] Marxa Mendiluee. Cleaner Production and eco-efficiency initiatives in Western Australia (1996-2004) [J], Journal of Cleaner Production, 2007:741-755.
[11] Lynn Price, Stephane de la Rue du Can, Jonathan Sinton. Et al. Assessment of China’s energy-saving and emission-reduction accomplishments and opportunities during the 11th Five Year Plan[J], Energy Policy, 2011, (39):2165-2178.