Evaluating the Health of Higher Education: A Hierarchical Clustering and Multi-Model Approach

Jiawei Kong[a*], Jingwen Du[b], Chunhe Wu[c]

Dalian University of Foreign Languages, Dalian, China

[a*]1554531248@qq.com, [b]13664293091@163.com
[c]2247648413@qq.com

Abstract. To foster societal advancement through higher education improvements, we developed a model to evaluate the health of national higher education systems. Our approach began with collecting data on 13 indicators from the US, using Hierarchical Clustering to categorize these into five factors: Gender Ratio, Cost, Research & Development Funding, Academic Degrees, and Access. We applied the Entropy Weight Method to determine indicator weights and conducted a Fuzzy Comprehensive Evaluation to assess the health levels of higher education across selected nations. Further, we expanded our analysis to 13 additional countries with varying economic statuses, applying the same clustering method to assess their education health, categorized into five echelons. Specifically, we established a Multiple Linear Regression Model to identify key factors influencing its educational health. This multifaceted approach not only resolved specific research problems but also provided a robust framework for assessing and improving national higher education systems.

Keywords: Education system; Fuzzy Comprehensive Evaluation; Multiple Linear Regression Model.

1 Introduction

The higher education system (H-edu-system) in contemporary society serves as a pivotal conduit for the dissemination of knowledge and skills, acting as the backbone of professional training and intellectual development. As the primary provider of advanced education, the H-edu-system plays a crucial role not only in cultivating a competent workforce but also in enhancing a nation's competitiveness on the global stage. From the perspective of social development, the health and robustness of a country's higher education system indirectly determine its status in the international community, influencing its economic, scientific, and cultural exchanges [1].

Despite their critical importance, the higher education systems across the globe exhibit significant variability. These disparities are not merely structural but also reflect in the quality of education and the capacity to adapt to the rapidly changing demands of the global economy [2]. Each system has its unique set of strengths and weaknesses [3], which contribute to the overall health and effectiveness of national educational

Y. Feng et al. (eds.), *Proceedings of the 4th International Conference on Internet, Education and Information Technology (IEIT 2024)*, Atlantis Highlights in Social Sciences, Education and Humanities 26,
https://doi.org/10.2991/978-94-6463-574-4_35

outcomes [4]. For instance, while some countries excel in technological advancements and research output, others prioritize making higher education more accessible to their populations. This diversity necessitates that nations with less developed higher education systems confront the challenge of reforming their educational frameworks to bridge gaps in quality and performance. As globalization intensifies, the need for competitive and adaptive educational structures becomes crucial, with some systems striving to enhance their global standing through reforms and innovations [5, 6]. Additionally, the shift towards a knowledge-based economy demands that higher education institutions not only transfer knowledge but also actively participate in creating it [7], thereby influencing economic development and societal well-being [8].

Our research aims to undertake a comprehensive evaluation of national higher education systems. We focus on assessing the 'health' of these systems through a multidimensional analysis that considers various performance indicators. This evaluation is not only geared towards identifying the current deficiencies and strengths but also aims to outline strategic directions for sustainable and effective development.

2 Evaluation System

2.1 Index Clustering Analysis

The healthy development level of higher education is a relatively vague concept. Given this, we have selected an appropriate evaluation model [9] to evaluate each country's healthy development level of higher education. Before choosing this evaluation model, we first need to determine the set of factors. The cluster analysis model can divide the sample into multiple classes composed of similar objects. We use SPSS to carry out hierarchical clustering analysis to determine the set of factors for the next step in the fuzzy comprehensive evaluation of national higher education's healthy development level [10]. Due to the different dimensions of the data, we also need to standardize the data. According to the elbow rule, we can estimate the optimal number of clusters for this model.

Each class degree of distortion sum: each class degree of distortion sum is equal to the sum of the square of the distance from the class centroid to its internal members. Assuming a total of the n divided samples to Kth class ($K \leq n-1$, there are at least two elements in one class) and use C_k to represent the kth class $(k = 1, 2, \dots, K)$, and the position of the center of gravity of this class is recorded as u_k, then the distortion degree of the Kth class is $\sum_{i \in Ck} |x_i - u_k|^2$.

Define the total degree of distortion (aggregation coefficient) for all classes:

$$J = \sum_{k=1}^{K} \sum_{i \in Ck} |x_i - u_k|^2 \quad (1)$$

when the K value is from 1 to 5, the degree of distortion changes the most. After more than 5, the degree of distortion changes significantly reduce. Therefore, the elbow is $K = 5$, so the number of categories can be set to 5. Apparently, based on the cluster analysis data in the figure, we can make specific dis-tinctions between the indicators that measure the health of the national H-edu-system. The selected 13th indicators accordingly are divided into five factors, as the following Table 1.

Table 1. The division of indicators

Factor	Index	Factor	Index
Gender ratio	GP	Academic degree	PC
	UF		PG
	P		S
Cost	CE		D
	G		M
R&D funding	RD	Access	SE

2.2 Fuzzy Comprehensive Evaluation

To access the health condition of the H-edu-system in the US over the last decade. First, determine the membership function. We use the existing objective scale. Since the maximum value of the SE and S variables is objectively specified as 100%, the utility function is used as the membership function. According to expert opinion, when the GP value is close to 1, the effect is better. Too large or small is not ideal. Based on this, we get the membership function of this variable as:

$$\mu = \begin{cases} 0, & x < 0 \\ x, & 0 < x < 1 \\ 2 - x, & 1 < x < 2 \end{cases} \tag{2}$$

It can be seen from the structure of the membership function that the membership functions of all factors have undergone positive processing; that is, the greater the degree of membership, the more advantageous it is. Use the entropy method to get the weight of each factor. Definition of information entropy:

$$I(x) = \ln(p(x)), 0 \leq p(x) \leq 1, I(x) \geq 0 \tag{3}$$

Entropy defines event X is $H(x) = \sum_{i=1}^{10}[p(x_{10})I(x_{10})]$, when $p(x_i) = p(x_2) = \cdots = p(x_{10}) = \frac{1}{10}$, the $H(x)$ maximum value, at this time $H(x)$=ln 10. We have 10 targets to be evaluated, 13 evaluation indexes (already forward of) forward of the matrix consisting of the following:

$$X = \begin{bmatrix} x_{1\times1} & \cdots & x_{1\times13} \\ \vdots & \ddots & \vdots \\ x_{10\times1} & \cdots & x_{10\times13} \end{bmatrix} \tag{4}$$

Then, the standardized matrix is denoted as Z, and each element in Z is $z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x^2{}_{ij}}}$, judge whether there is a negative number in the Z matrix. If there is, further processing is needed to obtain the final non-negative matrix. It is easy to verify: $\sum_{i=1}^{10} p_{ij} = 1$, which ensures that the sum of the probabilities corresponding to each indicator is 1.

2.3 National Clustering

So far, we have used the data of the past ten years in the United States to obtain essential factors and their weights for the American H-edu-system's health. We then use this method to randomly collect data related to 14 countries [11], including developed countries, developing countries, and less developed countries. Bringing the data into the evaluation model to evaluate. The H-edu-system's health, as shown in the Fig.1.

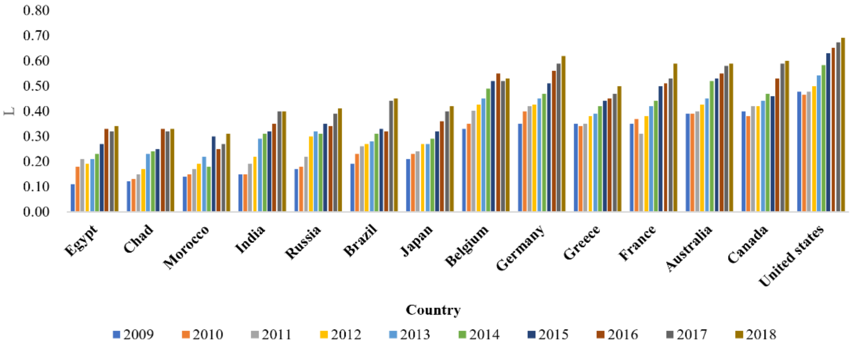

Fig. 1. Health assessment values of higher education in the last ten years

We have developed a persuasive model based on U.S. data [12]. However, we do not have specific data to develop a measure to demarcate the country's H-edu-system's health accurately. Therefore, based on the above model, we introduce the clustering method again, dividing the country type according to the evaluation value L. Through careful analysis of data we have obtained five final cluster centers, which divide the national H-edu-system's health into five parts. We can make specific distinctions about the health of the national H-edu-system which divided the selected 14 countries into five echelons based. And we also search for information to describe the five echelons' national strength, as shown in the Fig.2.

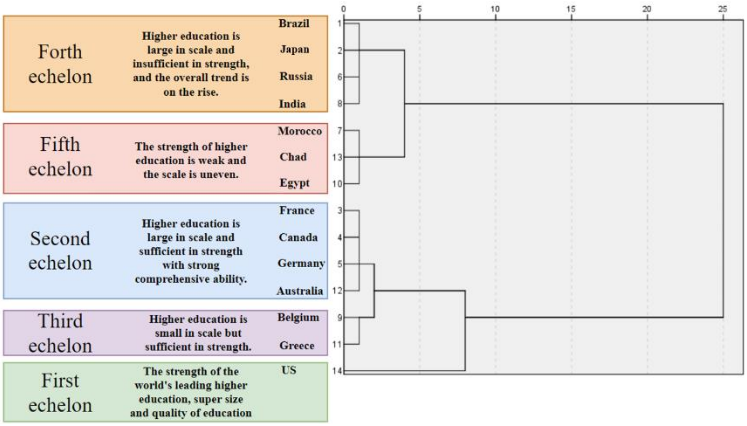

Fig. 2. Country classification and strength description

3 Multiple Linear Regression Model

We select Belgium, which is of moderate health, from the 14 countries mentioned above as our research object. In the influencing factors of the health of the national H-edu-system, we eliminate the indexes with a weight of 0, the remaining indexes are used as independent variables, the evaluation value L is used as the dependent variable, and then the Multiple Linear Regression Model is established. The data we collect and evaluate for Belgium from 2009 to 2018 are shown in the Table 2. We use Stata software to perform standardized multiple linear regression on the data due to the data dimension problem.

Table 2. Regression result

	Coefficients	Standard error	t Stat	P-value
Intercept	-4.24	2.28	-1.86	0.16
X1	2.33	0.0	0.7	0.6
X2	0.0043	0.01	0.32	0.77
X3	3.24	2.68	1.21	0.31
X4	1.9E-05	5.1E-05	3.7E-01	7.3E-01
X5	-5.8E-06	0.00	-0.86	0.45
X6	1.71	0.029	0.060	0.956

The formula is as follows:

$$Y = -4.24 + 2.33x_1 + 0.0043x_2 + 3.24x_3 + 1.9 \times 10^{-5}x_4 - 5.8 \times 10^{-6}x_5 + 1.71x_6 \quad (5)$$

The model's P value is $0.0471 < 0.05$. So we reject the hypothesis (H_0: $\beta_1 = \beta_2 = \cdots = \beta_6 = 0$), which shows that the model makes sense. Adj R-squared is 0.9282.

The good ness of fit is right. It can be clearly understood from the formula that the factors that significantly influence the health of the national H-edu-system are P, GP, and UF.

4 Conclusion

In this study, we conducted a detailed analysis of the health of national higher education systems (H-edu-systems), with a particular focus on identifying areas for improvement and enhancement. Using the United States—one of the most advanced H-edu-systems globally—as a benchmark, we developed a fuzzy comprehensive evaluation model to assess and compare various systems. This model enabled us to pinpoint Belgium as a nation with considerable potential for enhancing the effectiveness and overall health of its higher education system.

For colleges and universities, a strategic approach is essential. Institutions need to articulate their unique characteristics and strengths clearly. This entails the promotion of specialized educational programs that capitalize on these unique qualities, ensuring that they align with the institution's core competencies. Moreover, it is crucial for educational institutions to concentrate on their top-performing departments and majors. By doing so, they can guarantee that these areas not only continue to meet the current educational demands but are also well-prepared to adapt to future challenges and innovations in the field.

Overall, our research highlights the importance of continuous evaluation and adaptation in the development of higher education systems. By embracing a model that allows for detailed assessment and comparison, countries like Belgium can implement targeted improvements that will significantly advance the health and effectiveness of their higher education systems, ultimately contributing to a more robust educational foundation on a national scale.

References

1. Y. Chentukov, V. Omelchenko, O. Zakharova, and T. Nikolenko, "Assessing the impact of higher education competitiveness on the level of socio-economic development of a country," *Problems and Perspectives in Management,* vol. 19, pp. 370-383, 2021.
2. I. E. Allen and J. Seaman, *Changing course: Ten years of tracking online education in the United States*: ERIC, 2013.
3. Z. Tian, "A case study of the internationalisation of higher education in China: meaning, implementation and evaluation," University of Lincoln, 2015.
4. O. Kuznetsova, S. Kuznetsova, E. Yumaev, V. Kuznetsov, and I. Plotnikova, "Enhancing the Role of Educational Services of Higher Education System in the Competitive Specialists Training for Industry," in *E3S Web of Conferences*, 2017, vol. 15, p. 04018.
5. S. Marginson, "Dynamics of national and global competition in higher education," *Higher education,* vol. 52, pp. 1-39, 2006.
6. H. Y. Keser, "EFFECTS OF HIGHER EDUCATION ON GLOBAL COMPETITIVENESS: REVIEWS IN RELATION WITH EUROPEAN COUNTRIES

AND THE MIDDLE EAST COUNTRIES," *Annals of'Constantin Brancusi'University of Targu-Jiu. Economy Series,* vol. 1, 2015.
7. M. Krstić, J. A. Filipe, and J. Chavaglia, "Higher Education as a Determinant of the Competitiveness and Sustainable Development of an Economy," *Sustainability,* vol. 12, p. 6607, 2020.
8. L. M. Portnoi, S. S. Bagley, and V. D. Rust, "Mapping the terrain: The global competition phenomenon in higher education," in *Higher education, policy, and the global competition phenomenon*: Springer, 2010, pp. 1-13.
9. H. Liu, R. Chen, S. Cao, and H. Lv, "Evaluation of college English teaching quality based on grey clustering analysis," *International Journal of Emerging Technologies in Learning (iJET),* vol. 16, pp. 173-187, 2021.
10. O. Yim and K. T. Ramdeen, "Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data," *The quantitative methods for psychology,* vol. 11, pp. 8-21, 2015.
11. https://www.worldbank.org.
12. Gapminder. https://www.gapminder.org/data/.