ISSN 2663-8169 Vol. 3, Issue 2: 44-49, DOI: 10.25236/IJNDE.2021.030210

Touch the Pulse of Higher Education System: Using Health Diagnosis

Qinyi Zhang[1], Yizhuo Qin[2], Lingyu Qiu[1]

[1]*School of Statistics and Information, University of International Business and Economics, Shanghai 201600*
[2]*School of Financial Management, Shanghai University of International Business and Economics, Shanghai 201600*

Abstract: *In this paper, we propose a health diagnostician model to synthetically score national higher education systems. Firstly, we construct a Health Diagnostician Model for Higher Education, using Entropy Weight Method and Factor Analysis to screen 57 indicators and retain 7 representative indicators. We also introduce utility functions U and risk coefficients R to assess the health status quo of higher education system, where the input variables of the model are the "Health Portraits" composed of 57 indicators from each country, and output variables are health scores. Secondly, we extracted the "Health Portraits" from the World Bank Education Database for 31 countries for the last 10 years, and used the utility score and ranking to verify the validity of the model, we obtaining an average score of 0.166, The United Kingdom, United States Canada ranked in the top three. Mexico, with a score of 0.123, was selected for further study.*

Keywords: *Higher education, Health Diagnostician, Utility, Entropy method, Factor Analysis*

1. Introduction

Higher education often refers to the post-secondary schooling like colleges or universities. The tertiary education level is usually an epitome of hard power and soft power for a nation. It is commonly believed that the stronger a country is, the more high-quality tertiary institutions it owns. As for third-level education system. It is a kind of organizational structure, which calls for the coordination of fiscal expenditure, etc. The re-framing of this structure asks for the long-lasting efforts at all levels of a nation. In 2020, about 190 countries worldwide suspended offline courses due to the epidemic. While some schools have attempted to continue offering classes through online education, only half of the students have been able to complete some or all of their classes online. Thus, how to measure the health of the system and how to adjust accordingly become urgent issues.

In order to solve the problems, we have decided to take the following measures to make contributions in this article: firstly, we analyze factors affecting a healthy higher education system and develop a "diagnostic" model to assess the health of each countries higher education system. Then, we test multiple countries' education systems with the "diagnostic" model to verify the validity of the model. The diagnostic model is used to calculate the utility scores of current state and target state, in order to compare and analyze the health and sustainability of the system.

2. Health Diagnostician Model for Higher Education

2.1 Modeling Preparation

The higher education system of a country provides educated and trained citizens to the society. During the pandemic, higher education systems around the world have adjusted through online learning. Remote education has shed light on some "sub-health" problems in higher education systems. In order to assess the health and suitability of the higher education system, a customized health diagnostic is developed. And Table 1 shows the main notations and symbols we use in the paper. Other will be defined when using.

Table 1: Descriptions of notations and signs

Symbols	Unit	Description
x_{ij}	-	Indicator j for country i
p_{ij}	%	Proportion of indicator j for country i
e_j	-	Entropy of indicator j
g_j	-	Coefficient of variation of indicator j
a_j	-	Indicator weight
$E(x_{in})$	-	Benefit score
$E(x_{out})$	-	Expenditure score
R	-	Correction coefficient
U	-	Utility
θ^T	-	Characteristic parameter
S	-	Buffering parameter
U_S	-	Modified utility facing pandemic
U_G	-	Modified utility facing technology change

2.2 Input variables: "health portrait"

Initially, we extracted the data of 31 countries in ten years from World Bank's education statistics database. [1] Data includes 57 indicators from 10 series. Then, we create a "health portrait" of 31 countries, which was obtained from publicly available sources. The 31 countries are selected from all continents of the world to avoid diagnostic errors caused by geographical or cultural differences as much as possible. Figure1 below show the distribution of selected countries.Further, we deal with the missing values and outliers of the data. We make the negative indicator positive and carry out the standardized processing.

Figure 1: The selected countries evenly distributed all over the world

2.2.1 Make negative indicators positive

The negative indicator is opposite to the positive one. Its value is inversely proportional to the result, that is, the larger the value, the worse situation. We need to reverse some indicators to ensure that the score calculated is correct. The formula to make negative indicators positive is as follows:

$$x_i' = maxx_i - x_j, x_0 \leqslant x_j \leqslant maxx_i \tag{1}$$

ISSN 2663-8169 Vol. 3, Issue 2: 44-49, DOI: 10.25236/IJNDE.2021.030210

2.2.2 Standardized processing

If the units of the various indicators are not unified, we should carry out standardization process to unify the indicators as dimensionless. The formula we use is as follows:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{2}$$

In this formula, $\bar{x}_j$ represents the average value of indicator j under each research object; s_j represents the standard deviation of the indicator j.

2.3 Weight analysis: Entropy Weight Method

In order to avoid the impact of similarity between indicators on the diagnosis effect, we use the entropy weight method to extract the indicator with the biggest entropy value from each category as the representative. This measure can simplify the diagnosis process and optimize the final effect.

2.3.1 Entropy Weight Method

The concept of entropy is first originated from thermodynamics. Entropy can measure the degree of disorder of system information. The more information, the smaller the uncertainty and the smaller the entropy. According to the characteristics of entropy, the entropy value e can be used to judge the degree of dispersion of an indicator. The bigger the entropy value, the greater the degree of dispersion of the indicator while the influence of the index on the comprehensive evaluation is greater too. Therefore, the entropy weight method can be used to calculate the entropy and weights of each indicator to select the representative indicators.

2.3.2 Steps of Entropy method

(1) Calculate the proportion P_j, which represents the uncertainty of the indicator j. The calculation formula is

$$p_{ij} = \frac{x_{ij}}{\sum_{j=1}^{7} x_{ij}} \tag{3}$$

(2) Calculate the entropy value e_j which represents the uncertainty of the entire indicator j:

$$e_j = -k \sum_{j=1}^{31} p_{ij} \ln p_{ij} \tag{4}$$

Where k is a positive constant, $k = \frac{1}{\ln 31}, 0 \leqslant e_j \leqslant 1$. When the probability of each information occurrence is equal and the contribution of each program tends to be the same, e_j tends to equal 1. When $p_j = \frac{1}{m}$, the contribution of each scheme is completely equal. When the entropy e_j value is 1, the weight of the indicator is 0, and ignore this indicator in decision-making.

(3) Calculate the coefficient of variation.

The coefficient of variation refers to the difference in the contribution of each program within the indicator. The bigger the value, the greater the effect of the indicator on the research object. The coefficient of variance for the indicator j is

$$g_j = 1 - e_j \tag{5}$$

(4) Assign weights to the indicators. The weight of the indicator represents the role of each indicator in the comprehensive scoring. The weight value is the proportion of the difference coefficient of the indicator to the sum of the difference coefficients of all indicators.

$$a_j = \frac{g_j}{\sum_{j=1}^{n} g_j} \tag{6}$$

2.4 Determine analysis items: Factor Analysis

In order to further simplify the model, we perform dimensionality reduction on the extracted 10 indicators by factor analysis. The main idea of factor analysis is to find a few variables to describe the correlation between the original variables through the internal connection of the correlation coefficient matrix. According to the size of the correlation between variables, the variables with higher correlation are divided into the same group. So the correlation between variables in different groups is low. Each group is called a common factor. The model of factor analysis is as follows:

ISSN 2663-8169 Vol. 3, Issue 2: 44-49, DOI: 10.25236/IJNDE.2021.030210

$$X = BF + \varepsilon \tag{7}$$

It should satisfy:

(a) $m \leq p$

(b) $\operatorname{cov}(F, \varepsilon) = 0$, that means F and ε are unrelated

(c) $D(F) = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} = I_m$, $F_1, \cdots, F_m$ are uncorrelated, variances are 1

(d) $D(\varepsilon) = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_p^2 \end{bmatrix}$, $\varepsilon_1, \cdots, \varepsilon_p$ irrelevant and variances are different

Where $B = (b_{ij})_{p\times m}$ is the factor loading matrix, which represents the load of X_i on F_i . $F = (F_1, F_2, \cdots, F_m)^{\mathrm{T}}$ is a common factor, $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p)^{\mathrm{T}}$ is a special factor or error.

2.5 Output score: Evaluate the health of higher education system

Initially, we divide the results of factor analysis into two categories: benefit-oriented indicators and cost-oriented indicators. After getting benefit-oriented indicator values and respective weights, we can calculate the weighted value of benefits $E(X_{in})$, which is recorded as:

$$E(X_{in}) = \sum_{i=1}^{n(in)} a_j x'_{ij} \tag{8}$$

The cost-oriented index value is also recorded as

$$E(X_{out}) = \sum_{i=1}^{n(out)} a_j x'_{ij} \tag{9}$$

Then, we introduce the idea of utility in economics to construct the output score of the diagnostic device. Utility measures a person's satisfaction about one item. Higher utility means persons feel more satisfied. A rational individual prefers receiving more return but paying less. So, Benefit-oriented indicator like high enrollment, good national quality will increase the utility while cost-oriented indicator will decrease the utility. How to connect the benefit and cost to utility. Bodie, a famous economist, gives a utility formula in his book *Investment*(2021) [2], that is:

$$Utility = E - \frac{1}{2}A\delta^2 \tag{10}$$

E is the expected return of a portfolio, positively related to utility, A is to measure the investor's level of risk aversion. δ represents risks, negatively related to utility We can also simulate this formula since benefit-oriented indicator is positively related while cost-oriented indicator negatively. The more a nation dislike cost in higher education, the bigger R in our model will be. Underdeveloped countries are more cost averse because of weak economy. We introduce R as the correction coefficient. R is a constant and the value is the reciprocal of Government expenditure on tertiary education as percent. The utility function expression is as follows:

$$U = E(X_{in}) - R * E(X_{out}) \tag{11}$$

The value range of U is [-1,1]. The closer the value is to 1, the better the health of higher education system, and the closer the value to -1, the worse the health of the higher education system.

3. Health Diagnosis of Higher Education System

3.1 For 31 countries

Since our Health Diagnostician Model is a generalized one, the parameters of the model cannot be determined before selecting the target state. We tested the higher education systems in 31 countries to verify the effectiveness of the model. Firstly, we calculate the average value of each indicator in 31 countries. Then, we use the entropy method to obtain the weight of each indicator. After factor analysis, we retain 7 indicators from 57 indicators to form the Health Diagnostician Model. The 7 indicators are

ISSN 2663-8169 Vol. 3, Issue 2: 44-49, DOI: 10.25236/IJNDE.2021.030210

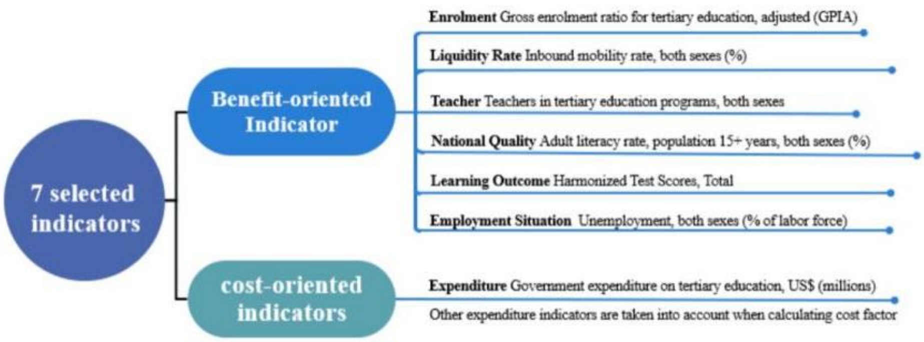

Figure 2: Selected indicators and definition of the indicators

Next, we calculate entropy value, for example, indicator Enrollment's e value equals 0.984, very high, indicating this indicator is more scattered and the contribution to the whole is higher. In addition, we calculate the coefficient of variation to describe the contribution of each indicator to the diagnosis result. For the indicator Enrollment, its g value equals 0.016, meaning this indicator explains 1.6% of the comprehensive score. Other indicators values are as follows:

Table 2: Value of e and g of every indicator

Indicator	e	g
Enrollment	0.984	0.016
Liquidity Rate	0.861	0.139
Teacher	0.772	0.228
Expenditure	0.686	0.314
National Quality	0.979	0.021
Learning Outcome	0.954	0.046
Employment Situation	0.898	0.102

Then, we calculate the weight of the benefit-oriented scores. Take the indicator Enrollment as an example again, its weight is 0.029, meaning the weight of Enrollment is 2.9%. The weight of other indicator are given as table below shows

Table 3: Weight of every indicator in E(xij)

Indicator	Weight of $E(X_{in})$	Indicator	Weight of $E(X_{in})$
Teacher	0.413	Employment Situation	0.185
Liquidity Rate	0.252	Learning Outcome	0.083
National Quality	0.039	Enrollment	0.029

Among them, the "teacher" factor contribution rate is 41.3%, indicating that the quality of teaching can determine the strength of a country's higher education level to a large extent; Liquidity Rate, which is the percentage indicator contribution of foreign students, is 25.2%, indicating that the acceptance of foreign students by the higher education system affects the education level. Meanwhile, we calculate the weight of cost-oriented scores, the weights for every indicator are shown as below. For example, the weight of current expenditure equals 0.113, it means that this indicator accounts for 11.3% of the change of $E(X_{out})$.

Table 4: Weight of every indicator in E(xout)

Indicator	Weight of $E(X_{out})$
Current expenditure as % of total expenditure (%)	0.113
Teaching staff compensation as a percentage of total expenditure in tertiary public institutions (%)	0.244
Government expenditure as % of GDP (%)	0.097
Initial government funding per tertiary student as a percentage of GDP per capital	0.137
Non-teaching staff compensation as a percentage of total expenditure in tertiary public institutions (%)	0.409

Finally, we can get the final utility score by formula (11) after making sure the weights. The ranking of every country and their own utility score are

ISSN 2663-8169 Vol. 3, Issue 2: 44-49, DOI: 10.25236/IJNDE.2021.030210

Table 5: Utility score and ranking of 31 countries

Country Name	Rank	Score	Country Name	Rank	Score
United States	1	0.530	Iceland	17	0.176
United Kingdom	2	0.517	Cambodia	18	0.142
Canada	3	0.473	Korea, Rep.	19	0.131
France	4	0.466	Mexico	20	0.123
Norway	5	0.422	Thailand	21	0.116
Japan	6	0.420	Poland	22	0.076
Italy	7	0.387	Turkey	23	0.051
Switzerland	8	0.385	Brazil	24	0.037
China	9	0.358	South Africa	25	-0.024
Singapore	10	0.317	Greece	26	-0.029
New Zealand	11	0.307	Iran, Islamic Rep.	27	-0.031
Denmark	12	0.278	Afghanistan	28	-0.070
Hong Kong, China	13	0.278	Egypt, Arab Rep.	29	-0.096
Russian Federation	14	0.231	Colombia	30	-0.122
India	15	0.223	Austria	31	-1.109
Malaysia	16	0.184			

As can be seen from the above table, the utility scores of most countries are within a reasonable range. The ranking of the health degree of the national higher education system conforms to our common sense. It proves that the H-D Model can effectively evaluate the health degree of the higher education system. The figure below also demonstrates every country's score.

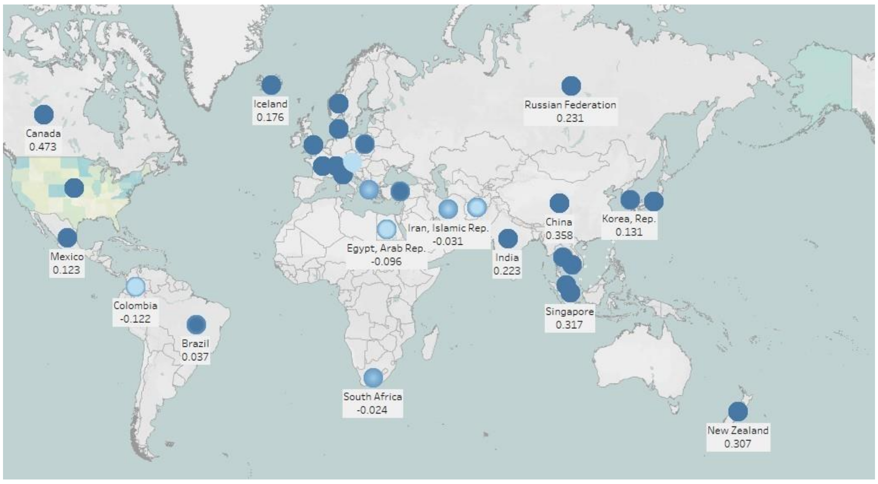

Figure 3: Every selected country's score on maps

4. Conclusion

Borrowing the principle of utility in economics, we establish a model including 7 indicators to assess the healthy of every country's system. The higher the utility score by our model, the healthier and more sustainable this country's system is. In a nutshell, higher education will support a country to go further and perform better on the world stage but improving the quality of it won't happen overnight. It requires the relentless efforts and determinations of every part of the country. Due to the special social and economic conditions, such as the outbreak of war and epidemic, many indicators of higher education system will change greatly. We can set a special situation indicator in the model, and the weight will be affected, and the sustainability of the evaluation system of higher education will also change.

References

[1] World Bank.https://databank.shihang.org/source/education-statistics. Accessed 5. Feb., 2021

[2] Zvi Bodie, Alex Kane and Alan J. Marcus. Investment, 12th ed.; McGraw-Hill Education: 2 Penn Plaza, New York, the U.S., 2021; pp.161.