Journal of Computers Vol. 35 No. 4, August 2024, pp. 123-132
doi: 10.53106/199115992024083504009

Based on Three-Stage Data Envelopment Analysis of The Efficiency Evaluation Model of Specialty Construction

Hong-Yu Zhai*, Quan-Yong Li, Tian-Li Wen, Qing-Xin Ma, and Luo-Dan Ma

Changchun University of Science and Technology, College of Computer Science and Technology, 130022, China

7604888@qq.com

Received 1 July 2023; Revised 11 July 2023; Accepted 11 July 2023

Abstract. Professional construction is the foundation and prerequisite for the development of higher education institutions, and has extremely important significance for the reform and development of universities. The Ministry of Education invests a large amount of funds every year for the construction of majors in universities. How to evaluate the output effectiveness of these investment funds is a problem worth paying attention to. In order to effectively evaluate the effectiveness of college specialty construction and make its evaluation more accurate, based on the traditional Data envelopment analysis model, this paper proposes a three-stage Data envelopment analysis based evaluation model for college specialty construction effectiveness, and takes the data of computer science and technology specialty in the comprehensive evaluation of undergraduate specialty in Jilin Province as an example to analyze the effectiveness of specialty construction.

Keywords: professional construction in universities, data envelopment analysis, evaluation model, investment, output effectiveness comprehensive

1 Introduction

Professional construction in Colleges and universities is an important way to improve the country's independent innovation ability. How to evaluate the efficiency of professional construction in universities has become an important issue faced by domestic universities today. The existing evaluation system usually involves a comprehensive evaluation process that takes into account various factors, often placing more emphasis on the comprehensive evaluation of various professional ability indicators, lacking analysis of professional input-output efficiency. In order to improve the future teaching quality assurance system and constantly improve the level of specialty construction and the quality of talent training, this paper will explore how to use the three-stage Data envelopment analysis method to evaluate the construction efficiency of a specific specialty in colleges and universities based on the comprehensive evaluation data of the specialty, and give targeted improvement plans.

Data envelopment analysis (DEA) is a nonparametric linear technique [1] used to measure the relative efficiency of a group of production decision units. DEA models [2-3] is to determine the relatively effective production frontier with the help of mathematical programming and statistical data by keeping the input or output of the decision-making unit unchanged, project each decision-making unit onto the production frontier of DEA [4], and evaluate their relative effectiveness by comparing the degree of deviation of the decision-making unit from the DEA frontier. It has absolute advantages in the field of multi input and output effectiveness comprehensive evaluation. In order to improve the accuracy of professional construction efficiency evaluation in Colleges and universities, based on the traditional data envelopment analysis model, this paper propose the efficiency evaluation model of professional construction in Colleges and Universities Based on three-stage data envelopment analysis, and takes the data of computer science and technology in the comprehensive evaluation of regional undergraduate majors as an example to analyze the efficiency of professional construction. The model can separate the impact of external environmental factors and random errors on efficiency measurement results, so that all decision-making units are in the same environment and random factors before comparison, so as to provide the dimension of efficiency evaluation, deepen the integration of teaching resources in Colleges and universities, improve the efficiency evaluation system of professional construction, and make leadership energy, teachers Resource allocation and fund arrangement can be used more reasonably.

* Corresponding Author

2 Dimensionality Reduction and Standardization

The efficiency evaluation model of specialty construction in Colleges and universities is a multi input and output effectiveness comprehensive evaluation model. There are 26 quantitative evaluation indexes for computer science and Technology Specialty in the region. These quantitative evaluation indicators can be divided into 14 input variables and 12 output variables according to input and output factors. As shown in Table 1 below:

Table 1. Input/output variables distinction

Input variables	Output variables
Average entrance score X1	The situation of academic papers published by teachers Y1
Average entrance score in overall school enrollment X2	The situation of teachers received scientific research awards at or above the provincial and ministerial level Y2
Professional teacher-student ratio X3	The situation of Research projects led by teachers Y3
Percentage of teacher with doctoral degrees X4	Number of faculty publications on teaching and research Y4
The situation of High-level teacher X5	The situation of teachers leading the development of teaching materials for their own specialties Y5
Professor teaching rate for undergraduate students X6	The situation of teachers preside over teaching and research projects at provincial level or above Y6
Proportion of professional courses undertaken by senior title faculty X7	Undergraduate teaching engineering projects above provincial level in the past years Y7
Proportion of professional teachers with industry experience X8	Teaching achievement awards above provincial level in the past years Y8
Proportion of young and middle-aged teachers trained in practical skills X9	Employment rate in the past four years Y9
Average number of teaching laboratory apparatus per student X10	Proportion of the number of student entrepreneurial science researchers Y10
Average number of new instrumentation per student in the last four years X11	The situation of students won awards in various competitions above the provincial level Y11
Off-campus internship practice situation X12	Students' published papers and patents Y12
Average number of paper and e-book materials per student X13	
Number of existing professional e-book sources X14	

Among ordinary undergraduate colleges and universities in the region, 17 colleges and universities have computer science and technology majors, that is, 17 decision-making units. In order to meet the needs of data envelopment analysis, when constructing the efficiency analysis model of college specialty construction based on three-stage data envelopment analysis, it is necessary to reduce the dimension of the existing input and output variables [5]. According to the correlation between input and output variables, input variables are divided into student source variables F_X1, teacher variables F_X2 and hardware variables F_X3 according to categories, and output variables are divided into teacher output variables F_Y1 and student output variables F_Y2 according to categories, as shown in Table 2 below (the weight of each index is formulated by the Education Instruction Committee of computer science and technology specialty when formulating the index system).

Table 2. Input/output variables dimensionality reduction

Variables		Indicators	Weights
Input variables	Student source F_X1	Average entrance score X1	0.06
		Average entrance score in overall school enrollment X2	0.04
	Teacher resources F_X2	Professional teacher-student ratio X3	0.02625
		Percentage of teacher with doctoral degrees X4	0.01575
		The situation of High-level teacher X5	0.01575
		Professor teaching rate for undergraduate students X6	0.0063
		Proportion of professional courses undertaken by senior title faculty X7	0.00945
		Proportion of professional teachers with industry experience X8	0.01575
		Proportion of young and middle-aged teachers trained in practical skills X9	0.01575
	Hardware F_X3	Average number of teaching laboratory apparatus per student X10	0.009
		Average number of new instrumentation per student in the last four years X11	0.009
		Off-campus internship practice situation X12	0.012
		Average number of paper and e-book materials per student X13	0.009
		Number of existing professional e-book sources X14	0.006
Output variables	Teacher F_Y1	The situation of academic papers published by teachers Y1	0.03
		The situation of teachers received scientific research awards at or above the provincial and ministerial level Y2	0.0225
		The situation of Research projects led by teachers Y3	0.0225
		Number of faculty publications on teaching and research Y4	0.0225
		The situation of teachers leading the development of teaching materials for their own specialties Y5	0.0225
		The situation of teachers preside over teaching and research projects at provincial level or above Y6	0.03
		Undergraduate teaching engineering projects above provincial level in the past years Y7	0.075
		Teaching achievement awards above provincial level in the past years Y8	0.075
	Student F_Y2	Employment rate in the past four years Y9	0.05
		Proportion of the number of student entrepreneurial science researchers Y10	0.03
		The situation of students won awards in various competitions above the provincial level Y11	0.03
		Students’ published papers and patents Y12	0.015

In order to get more accurate efficiency evaluation, the scores of these index items are dimensionless processed according to the maximum of the original score, that is $x' = \frac{x}{\max x} * 100$, each index score is standardized, and then the product of the index scores and the weight under each variable is added to obtain the score of each variable, namely:

$$F_X 1 = 0.06 * X1 + 0.04 * X2. \quad (1)$$

$$F_X 2 = 0.02625 * X3 + 0.01575 * X4 + 0.01575 * X5 + 0.0063 * X6 + 0.00945 * X7 + 0.01575 * X8 + 0.01575 * X9. \quad (2)$$

$$F_X 3 = 0.009 * X10 + 0.009 * X11 + 0.012 * X12 + 0.009 * X13 + 0.006 * X14. \quad (3)$$

$$F_Y 1 = 0.03 * Y1 + 0.0225 * Y2 + 0.0225 * Y3 + 0.0225 * Y4 + 0.0225 * Y5 + 0.03 * Y6 + 0.075 * Y7 + 0.075 * Y8. \quad (4)$$

$$F_Y 2 = 0.05 * Y9 + 0.03 * Y10 + 0.03 * Y11 + 0.015 * Y12. \quad (5)$$

After the above treatment, the results are shown in Fig. 1 below:

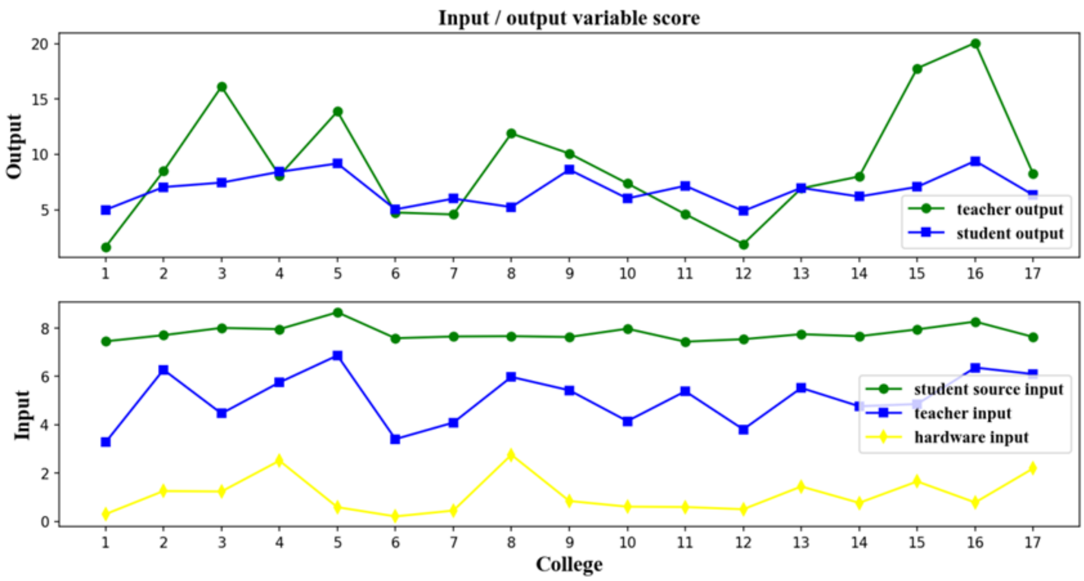

Fig. 1. Input / output variable score

3 Efficiency Analysis Model Based on Three-stage DEA

3.1 Stage 1: Traditional DEA Model Calculates Relaxation Variables

DEA models can be divided into input-oriented and output-oriented models. Since output is not easy to control, this paper will use input-oriented DEA model for analysis. First, establish DEA model:

$$\begin{cases} \min[\theta - \varepsilon(s_1^- + s_2^- + s_3^- + s_1^+ + s_2^+)] \\ s.t. \quad \Sigma_{j=1}^{17}\lambda_j x_j + s_1^- = \theta x_0 \\ \qquad \Sigma_{j=1}^{17}\lambda_j x_j + s_2^- = \theta x_0 \\ \qquad \Sigma_{j=1}^{17}\lambda_j x_j + s_3^- = \theta x_0 \\ \qquad \Sigma_{j=1}^{17}\lambda_j y_j - s_1^+ = y_0 \\ \qquad \Sigma_{j=1}^{17}\lambda_j y_j - s_2^+ = y_0 \\ \lambda_j \ge 0, j = 1,2,...,17 \\ s_1^-, s_2^-, s_3^- \ge 0 \quad s_1^+ + s_2^+ \ge 0 \end{cases} \tag{6}$$

Among them, θ is the valid value of the decision unit, λ_j is the weight, x_j , y_j represents the input and output of the j-th decision-making unit, s_1^-, s_2^-, s_3^- is the input relaxation variable, s_1^+, s_2^+ is the output relaxation variable, ε is archimedes infinitesimal.

In the initial DEA Analysis of the first stage, the main purpose is to obtain the relaxation variables, so as to use the stochastic frontier (SFA) model [6-8] to regress the relaxation variables in the second stage. Since the input-oriented model [9] is adopted in this paper, only the input relaxation variables will be decomposed by SFA regression in the future, and the scores in Fig. 1 will be brought into the model to obtain the input relaxation variables s of each decision-making unit s_1^-, s_2^-, s_3^- , as shown in Table 3:

Table 3. Input relaxation variable

Decision-making unit	Student resource input s_1^-	Teacher resource input s_2^-	Hardware input s_3^-
University 1	0.000	0.000	0.000
University 2	0.140	0.728	0.023
University 3	0.000	0.000	0.000
University 4	0.352	0.332	1.703
University 5	0.000	0.000	0.000
University 6	0.000	0.000	0.000
University 7	0.133	0.071	0.008
University 8	0.000	0.000	0.000
University 9	0.000	0.000	0.000
University 10	0.340	0.177	0.026
University 11	0.000	0.000	0.000
University 12	0.092	0.047	0.134
University 13	0.225	0.160	0.747
University 14	0.066	0.041	0.036
University 15	0.000	0.000	0.000
University 16	0.000	0.000	0.000
University 17	0.081	0.435	0.580

3.2 Stage 2: Construct SFA Regression

Construct SFA regression as follows:

$$S_{ni} = f(Z_i;\beta_n) + v_{ni} + \mu_{ni}; i = 1,2,...,17; m = 1,2,3. \quad \textbf{(7)}$$

Where S_{ni} is the input relaxation value of the n-item of decision-making unit i; Z_i is the environment variable, β_n is the coefficient of environmental variable; $v_{ni} + \mu_{ni}$ is the mixed error, v_{ni} stands for random interference, μ_{ni} indicates inefficient management. Where $v \sim N(0, \sigma_v^2)$ is a random error, which indicates the influence of random interference factors on input relaxation variables; μ is management inefficiency, which indicates the influence of management factors on input relaxation variables. It is assumed that it follows the normal distribution of zero truncation, $\mu \sim N^+(0, \sigma_\mu^2)$.

The purpose of SFA regression is to eliminate the influence of environmental factors and random factors on efficiency measurement, so as to adjust all decision-making units to the same external environment. The adjustable formula is as follows:

$$X_{ni}^A = X_{ni} + [\max(f^j(z_j;\hat{\beta}_n)) - f(z_j;\hat{\beta}_n)] + [\max(v_{ni}) - v_{ni}]. \quad \textbf{(8)}$$

Where X_{ni}^A is the adjusted input; X_{ni} is the input before adjustment; $\max(f^j(z_j;\hat{\beta}_n)) - f(z_j;\hat{\beta}_n)$ is to adjust external environmental factors; $[\max(v_{ni}) - v_{ni}]$ is to put all decision-making units under the same luck level.

The main purpose of this paper in the second stage is to separate the influence of the scale factor of the university where the major is located on the evaluation of professional efficiency. Therefore, the number of undergraduate students in the university where the major is located is introduced z_1, number of teaching staff z_2 and undergraduate special teaching funds z_3 three data are taken as external environmental factors, as shown in Table 4 below:

Table 4. External environmental data

Decision-making unit	The number of undergraduate students in the university where the major is located z_1	The number of teaching staff z_2	Undergraduate special teaching funds z_3
University 1	10369	1050	2457.94
University 2	23337	2825	5153.33
University 3	15947	1816	5376.33
University 4	10960	1086	993.99
University 5	41344	6633	34019
University 6	9230	983	399.72
University 7	13312	954	6484
University 8	14533	1354	5020.22
University 9	15350	2056	6157
University 10	16948	2066	653.67
University 11	10987	745	2651
University 12	18875	3018	7670
University 13	14580	1481	1700
University 14	13650	1568	785.23
University 15	15118	1846	1501.4
University 16	16076	2400	4457
University 17	17186	2289	1780

Use the data in Table 4 to separate s_1^-, s_2^-, s_3^- the three input relaxation variables are decomposed by SFA regression respectively to obtain the input variables of each decision-making unit, and finally the adjusted student input $F_{X_1}^A$, teacher input $F_{X_2}^A$ and hardware input $F_{X_3}^A$, as shown in Table 5 below:

Table 5. Adjusted input variables after SFA decomposition

Decision-making unit	Student input $F_{X_1}^A$	Teacher input $F_{X_2}^A$	Hardware input $F_{X_3}^A$
University 1	7.63	3.71	0.35
University 2	7.71	6.28	1.26
University 3	8.14	4.76	1.27
University 4	8.10	6.13	2.56
University 5	8.81	7.14	0.61
University 6	7.76	3.86	0.26
University 7	7.80	4.41	0.50
University 8	7.80	6.26	2.80
University 9	7.79	5.83	0.89
University 10	8.03	4.32	0.62
University 11	7.59	5.73	0.64
University 12	7.70	4.21	0.53
University 13	7.84	5.75	1.47
University 14	7.77	5.05	0.80
University 15	8.06	5.12	1.68
University 16	8.42	6.75	0.82
University 17	7.72	6.32	2.21

3.3 Stage 3: Secondary Analysis of DEA Model

After the adjustment of the second stage, the input variables have eliminated the influence of environment and random factors. After replacing the input data, the DEA test is carried out again, and the real efficiency value is obtained by using the DEA model of the first stage again.

4 Efficiency Analysis Results and Mining

Through the processing of the above three-stage data envelopment analysis model, the DEA efficiency of the output of 17 colleges and universities in the region is obtained, as shown in Fig. 2.

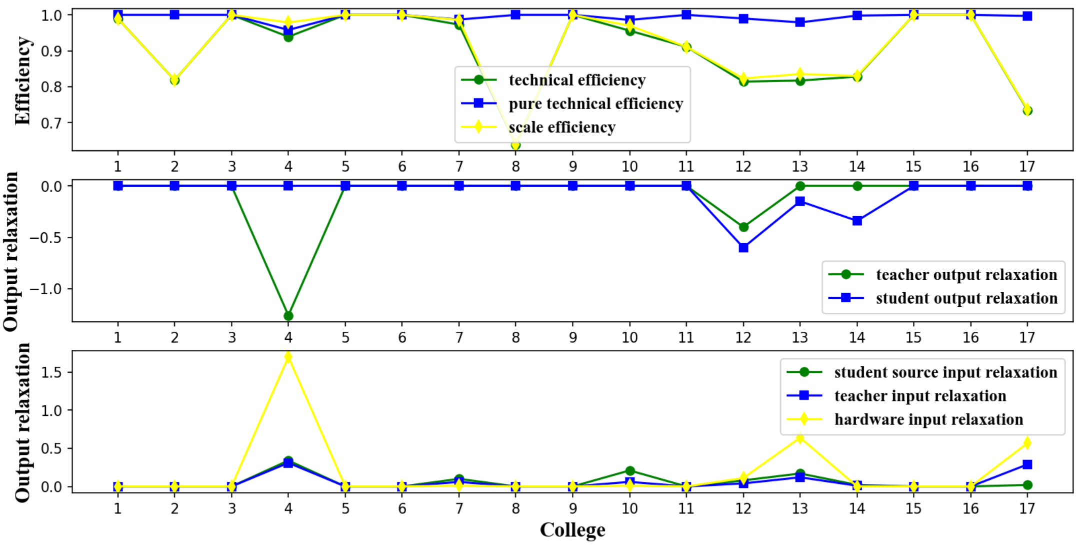

Fig. 2. DEA efficiency of input and output in 17 universities

It can be seen from the calculation results of the model that the technical efficiency of six universities 3, 5, 6, 9, 15 and 16 is 1, and the corresponding relaxation variables are 0, meets the effectiveness of DEA, which means that their input-output is are comprehensive and effective, that is, technology is effective and scale is effective at the same time, the return to scale remains unchanged, the efficiency of professional construction is high, while the other 11 universities are non DEA effective, the comprehensive input-output efficiency still does not reach the best state. Among the 11 non DEA effective universities, the pure technical efficiency of universities 1, 2, 8 and 11 has reached 1, which means that their input resources are used efficiently at the current technical level. The fundamental reason for their failure to achieve comprehensive effectiveness lies in their ineffective scale. Therefore, the focus on their reform should be on how to give better play to their economies of scale. For the remaining seven non DEA effective universities, in addition to improving the economies of scale, they also need to pay attention to the use efficiency of input resources. For example, among the six universities, the university 4 with the lowest pure technical efficiency has slack investment in student source investment, teacher investment and hardware investment, and each slack investment is the highest in the province, this shows that it has huge problems in terms of input. For example, the relaxation of teachers' input and hardware input in university 10 is relatively not high, but the relaxation in student input ranks second in the province, which shows that the main problem in input efficiency lies in student input, which is inconsistent with its output and low efficiency.

During the experiment, when three input variables and two output variables are set, six colleges and universities finally achieve DEA effectiveness. If you want to get a higher differentiated ranking, you can take further dimensionality reduction methods. For example, you can combine the student input and teacher input into software input according to the previously used dimensionality reduction method, which will become two input variables and two output variables, and then conduct three-stage DEA analysis. At this time, the colleges and universities that can achieve DEA effectiveness (technical efficiency is 1) will become five; Further combine the two output variables into comprehensive output. When there are two input variables and one output variable, only two colleges and universities can achieve DEA effectiveness. Finally, if only single input and single output are set, only one college and university can achieve DEA effectiveness. This shows that there is a certain relationship between input and output variables and efficiency ranking in the model. Fewer variables are not conducive to the targeted and specific analysis of the reasons, but during efficiency ranking, the dimension of input and output variables can be further reduced as needed to obtain the desired results. The DEA efficiency calculation results under different variables are shown in Fig. 3 below:

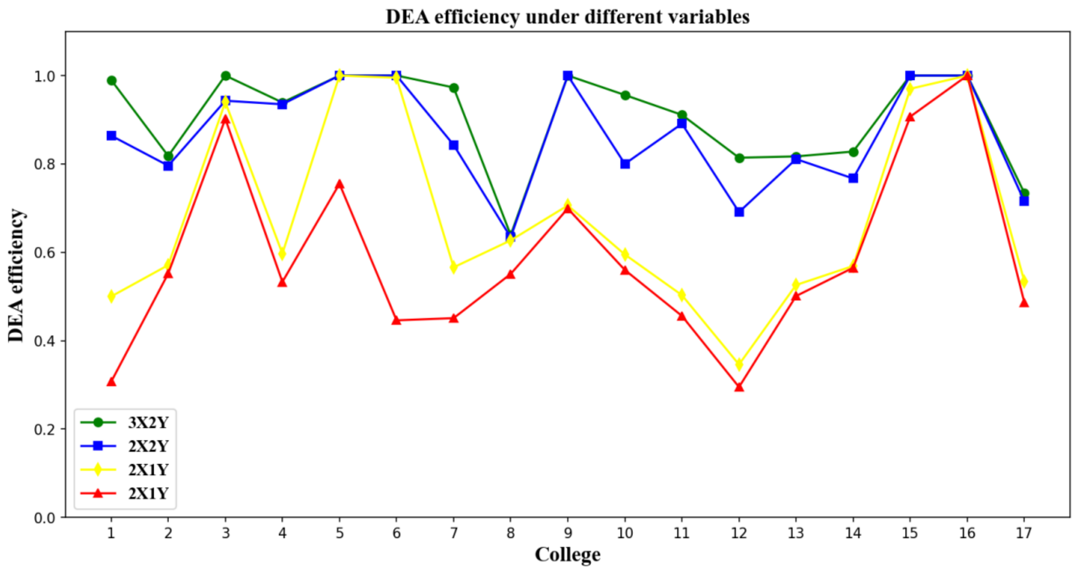

Fig. 3. DEA efficiency under different variables

Another important significance of the comprehensive evaluation of undergraduate majors is to supervise and urge the lower ranked universities and make a more detailed and targeted diagnosis. Still take the university 4 with the lowest pure technical efficiency in the province as an example, further analyze the hardware input variable with the largest relaxation among the three input variables of university 4, the indicators contained in the hardware input variable are the average value of teaching experimental instrument and equipment students, the average value of new instrument and equipment students in recent four years, the practice of off campus practice and the number of existing professional e-book sources. These five indicators are set as input variables A1 to A5 respectively, so as to obtain more meaningful results, the output variable is transformed into a comprehensive output variable y through the dimension reduction method, and the relaxation variables of the five hardware input variables of University 4 are obtained through the three-stage DEA analysis method. As shown in Table 6, it can be seen that the relaxation variable of the number of professional e-book sources is only 0.932, which has the smallest impact on the overall efficiency and can be almost ignored, the greatest impact on efficiency is the number of copies of paper and e-books per students, followed by the average number of new instrument and equipment students in recent four years, the average number of teaching and experimental instrument and equipment students and the practice of off campus practice, which can be further reformed with emphasis according to the degree of importance.

Table 6. University 4 hardware input relaxation variable

Input variable	Relaxation
Average number of teaching laboratory apparatus per student A1	33.607
Average number of new instrumentation per student in the last four years A2	44.287
Off-campus internship practice situation A3	27.645
Average number of paper and e-book materials per student A4	81.091
Number of existing professional e-book sources A5	0.932

5 Conclusion

The fundamental task of colleges and universities is to cultivate talents, and the comprehensive evaluation of undergraduate majors, as an important component of the evaluation of professional construction in Chinese colleges and universities, has begun to guide universities to take cultivating talents as the most fundamental task. It plays a very positive role in striving to improve the level of specialty construction and the quality of talent training. This paper is based on the efficiency evaluation method and professional comprehensive evaluation data of the three-stage DEA model, aiming to provide efficiency evaluation dimensions for the professional construction of various colleges and universities, make more rational use of leadership energy, teachers, resource allocation and fund arrangement, further strengthen professional construction and reform, improve the evaluation system and model of professional construction efficiency in colleges and universities, guide colleges and universities in the province to develop characteristics and level, and provide strong talent and intellectual support for economic and social development has important practical application value for the future development of colleges and universities.

References

[1] N. Neykov, S. Krišťáková, I. Hajdúchová, M. Sedliačiková, P. Antov, B. Giertliová, Economic efficiency of forest enterprises—Empirical study based on data envelopment analysis Forests 12(4)(2021) 462.
[2] B. Prevolšek, M.-B. Gačnik, Č. Rozman, Applying Integrated Data Envelopment Analysis and Analytic Hierarchy Process to Measuring the Efficiency of Tourist Farms, The Case of Slovenia, Sustainability 15(5)(2023) 4314.
[3] K.F. Čiković, J. Lozić, Application of Data Envelopment Analysis (DEA) in information and communication technologies, Tehnički glasnik 16(1)(2022) 129-134.
[4] Y. Ersoy, Performance evaluation in distance education by using data envelopment analysis (DEA) and TOPSIS methods, Arabian Journal for Science and Engineering 46(2)(2021) 1803-1817.
[5] S.H.R. Hajiagha, H.A. Mahdiraji, S.-S. Hashemi, J.-A. Garza-Reyes, R. Joshi, Public hospitals performance measurement through a three-staged data envelopment analysis approach: evidence from an emerging economy, Cybernetics and Systems 54(1)(2023) 1-26.
[6] F.-G. Dong, X.-W. An, C.-Y. Li, Performance evaluation of wind power industry chain based on three-stage DEA, Journal of Renewable and Sustainable Energy 13(3)(2021) 033313.
[7] B. Lin, W.-D. Lu, Port Efficiency Evaluation Based on Three-Stage DEA Model, ICIC express letters. Part B, Applications: an international journal of research and surveys 12(10)(2021) 925-931.
[8] E. Alorzuke, S.-L. Kannika, S. Nirote, W. Anupong, The Studies on Efficiency Analysis in Applied Economics Using a Three-Stage Data Envelopment Analysis, Central European Management Journal 30(3)(2022) 300-311.
[9] S. Nazari, M. Rostamy-Malkhalifeh, A. Hamzehee, Congestion Calculation only by Solving a Linear Programming Model Through Fuzzy Data, International Journal of Data Envelopment Analysis 10(1)(2022) 47-58.