

The Educational Efficiency Evaluation Framework: By Using DEA Model and CA Method

Yi Yuan and Mingfeng Shan

Abstract—Regarding the PISA result from Shanghai China, the perspective reason concerning Chinese educational financial budget and regime are demonstrated. The DEA and Cluster Analysis approach are used to analyze the Educational Efficiency for Shanghai 17 Districts. Those Districts are grouped as rich or poor resource input, the efficiency results show that the suburb districts are more efficient comparing with the rich source urban district. The evaluation framework could be helpful for government to make the planning of resource allocation.

Index Terms—PISA, cluster analysis, DEA model, efficiency.

I. INTRODUCTION

In a global economy background, the countries' success and achievement is no longer measured against national standards alone, but against the best-performing and most rapidly improving education systems. More and more countries are looking beyond their own borders for evidence of the most successful and efficient policies and practices.

Program for International Student Assessment (PISA) [1] is a triennial international survey which aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. Since 2000 to 2012, students representing more than 70 economies have participated in the assessment. Over the past decade, PISA has become the world's premier yardstick for evaluating the quality, equity and efficiency of school systems. But the evidence base that PISA has produced goes well beyond statistical benchmarking. By identifying the characteristics of high-performing education systems PISA allows governments and educators to identify effective policies that they can then adapt to their local contexts.

In 2009, Shang Hai students first appeared for the PISA tests and made a stunning debut on the assessment of three subjects with a score of 577 while the OECD average was 497. Three years later, in 2012, Shang Hai student keep the rank top in the examination. The average score of three subjects is 588 against the OECD average score 497.

The purpose of this paper is to demonstrate the pivot view for Shang Hai Basic Education system which involving the primary and secondary educational level. The government monetary input regime and data are overviewed. The educational system efficiency evaluation framework is defined and modeled. In the end, The Data Envelopment Analysis and Cluster Analysis approaches are used to

evaluate the Shang Hai Basic Education system production efficiency. It will dedicate the policy and decision maker understand the Shang Sai basic educational system running and educational input/out efficiency evaluation and assessment.

We suggest that the design of the evaluation framework must include two sets of functionalities: externally-oriented, and internally-oriented. The externally-oriented functionality of the Model is directed towards evaluating the external competitive environment, as well as identifying the differences between the current state of the unit and the states of its competitors. The internally-oriented functionality, on the other hand, is directed towards the optimization of the level of productivity of the unit, as well as towards an identification of the factor impacting the efficiency of the input–output process. We suggest that such a Model could be implemented using a combination of parametric and non-parametric data analytic and data mining techniques including Data Envelopment Analysis(DEA), Cluster Analysis(CA), and Multivariate Regression(MR).

II. FRAMEWORK STRUCTURAL COMPONENT OVERVIEW

A. Data Envelopment Analysis

Data envelopment analysis (DEA) is a technique for measuring the relative efficiency of a set of decision making units (DMUs) which apply multiple inputs to produce multiple outputs in a period of time. Since the pioneering working by A. Charnes [2]. DEA was first introduced by Assuming a constant returns-to-scale (CRS), this original DEA model (referred to as the CCR model) is capable of determining an overall relative or technical efficiency (TE) for each DMU with respect to the efficient production frontier described by the data. Later modified by Banker, Charnes, and Cooper [3], the resulting BCC model allowed for variable returns-to-scale (VRS). As a result, the BCC model is able to measure technical efficiency and inefficiency and, when used with the CCR model, scale inefficiency. The adaptability and versatility of this methodology in measuring the relative efficiency of a decision making unit (DMU) is practically limitless. Indeed, DEA has been successfully applied for many years to a variety of disciplines, industries, and problems.

Since enhanced efficiency or productivity often translates to savings with respect to time and money, DEA has been used quite extensively to evaluate the relative efficiency of container terminals throughout the world. For example, Cullinane [4] evaluated the relative efficiency of 26 container ports using DEA from two perspectives—deterministically

Manuscript received May 14, 2015; revised August 11, 2015.

The authors are with the Shanghai University, China (e-mail: environ@shu.edu.cn, smf2002@vip.sina.com).

(using an output-oriented DEA model under both VRS and CRS) and stochastically (based on a log-linear Cobb-Douglas production function and under the error assumption of half-normal, truncated normal, and gamma distributions). Cullinane *et al.* found that while there were differences in the efficiency ranks of the two methods, those rankings were highly correlated. In addition, Cullinane *et al.* recognized the importance of these efficiency measures in that they could influence a port's decision-making process and governmental policy-making. Pergelova [5] used both DEA and stochastic frontier analysis (SFA) to evaluate the efficiency of marketing automobiles in Spain with respect to offline and online delivery methods. Khodabakhshi [6] used a chance constrained DEA model to develop a stochastic super-efficiency model for the purpose of evaluating the efficiency of 17 electrical distribution companies in Iran. Khodabakhshi, Asgharian, and Gregoriou [7] next applied their stochastic super-efficiency DEA model to the efficiency assessment of United States public banks and thrifts CEOs.

Specific for educational domain, DEA approach is generally used for efficiency evaluation. Some recent references include Avkiran [8] and Abbott and Doucouliagos [9] for Australia, Salerno [10] and Calhoun [11] for the United States, Afonso and Santos [12] for Portugal, Warning [13] for Germany, Johnes [14] for England, Jongbloed and Salerno [15] and Cherchye and Abeele [16] for the Netherlands, and Castano and Cabanda [17] for the Philippines. Joumady and Ris [18] represent an exception as they work with a set of countries (Austria, Finland, France, Germany, Italy, Netherlands, Spain and the United Kingdom). They also innovate by using the competence gained during the undergraduate years and the competence required by their current jobs as output measures. Ferrari and Laureti [19] model the human capital production process in the university system, as one in which a production unit produces itself to estimate the output-efficiency of human capital formation in the University of Florence. They use however the DEA approach. Laureti [20] use a hetero-frontier model in order to consider the effect of students' individual characteristics and the influences of the resources and organization of the specific degree course on efficiency.

B. Cluster Analysis

Clustering is a popular non-directed learning data mining technique for partitioning a dataset into a useful set of mutually exclusive clusters such that the similarity between the observations within each cluster is high, while the similarity between the observations from the different clusters is low [21]. There are different reasons for doing clustering, and one of them is to find a set of natural groups (i.e., segmentation), and the corresponding description of each group. This is relevant if there is the belief that there are natural groupings in the data. Jain [22] noted that there are three approaches for assessing cluster validity: 1) external assessment which involves comparing the generated segmentation (i.e., set of clusters) with an apriority structure, typically provided by some domain experts; 2) internal assessment which attempts to determine if the generated set of clusters is "intrinsically appropriate" for the data; and 3) relative assessment which involves comparing two

segmentations (i.e. two sets of clusters) based on some performance measures and measure their relative performance. Our use of cluster analysis is based on the assumption that there are natural groupings in the data, and will involve the use of external assessment to assess cluster validity. There are numerous algorithms available for doing clustering. They may be categorized in various ways such as: hierarchical [23] or partitioned [24], deterministic or probabilistic [25], hard or fuzzy [26]. A hybrid partitioned /hierarchical approach provided is used to generate the clusters.

III. SHANG HAI EDUCATION SYSTEM INTRODUCTION

In this paper, Shang Hai Basic Educational system is defined including the primary and secondary period, covering the teenage from 7 years old to 15 years old students. In Chinese National financial system, there are three levels, such as National, provincial and local financial executive. The local government is the mainly responsible for the regional educational financial budget implementation (local payment). The national financial supply the fix budget (national payment) for the educational development and the provincial government's payment (transfer payment and education supplementary payment) is generally for decreasing the differences among the regional. The regime is showed in Fig. 1.

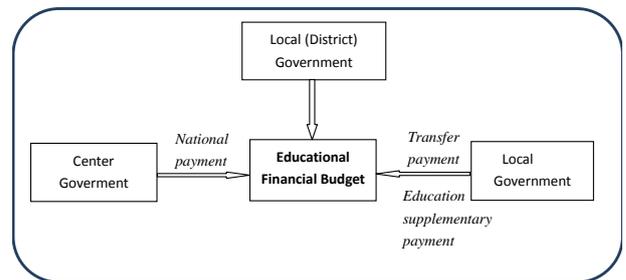


Fig. 1. The educational financial budget executive regime scheme in Shang Hai city.

IV. DATA AND MEASUREMENTS

TABLE I: SIX MAJOR INDICATORS DEFINITION FOR EDUCATIONAL BUDGET EXECUTIVE EVALUATION

No.	Indicators	Code	Definition
1	Total budget per capita	(TBC)	Total budge /student number in the district
2	Equipment budget per capita	(EBC)	Total equipment budge/student number in the district
3	The ratio of student/teacher	(RST)	Student number/teacher number in the district
4	Quota per class	(QC)	Student number/class number in the district
5	Quota per school	(QS)	Class number/school number in the district
6	Student Density per Km ²	(SD)	Student number/district square

According the public data which is announced by Shang Hai Municipal Education Commission (SMEC), Shang Hai Municipal Finance Bureau (SMFB) and Shang Hai Municipal Statistics Bureau (SMSB), for the period 2008-2012, Shang

Hai district educational budget executive indicators data are calculated. There are six major indicators considerable shown in Table I and Table II.

TABLE II: THE CALCULATION RESULT FOR MAJOR INDICATORS OF BUDGET EXECUTIVE EVALUATION

No.	District	TBC	EBC	RST	QC	QS	SD
1	Huangpu	9754	627	15	22	28	1923
2	Xuhui	4412	675	38	23	34	999
3	Changning	4235	614	6	26	30	909
4	Jingan	7481	562	6	24	30	2791
5	Putuo	6015	483	27	23	33	912
6	Zhabei	4226	720	9	20	32	1320
7	Hongkou	3993	457	9	18	31	1502
8	Yangpu	4845	380	11	21	28	773
9	Minhang	3156	317	15	26	36	228
10	Baoshan	3204	464	9	22	37	307
11	Jiading	3343	429	7	28	37	110
12	Pudong	2777	379	56	28	38	203
13	Jinshan	2940	360	2	26	39	79
14	Songjiang	2415	178	15	33	42	94
15	Qingpu	2524	157	9	27	38	68
16	Fengxian	2696	201	14	28	42	95
17	Chongming	4149	493	8	18	33	33
	Mean	4245	441	15	24	35	726

V. EMPIRICAL PROCEDURE

A. Cluster Analysis Approach

In order to make the cluster analysis by six indicators, the data standardization process was compulsory. The mean value of indicators was used as constrains. The result of standardization is shown as Table III.

TABLE III: THE STANDARDIZATION RESULT FOR MAJOR INDICATORS OF BUDGET EXECUTIVE EVALUATION

No	District	TBC	EBC	RST	QC	QS	SD
1	Huangpu	2.30	1.42	0.99	0.92	0.81	2.65
2	Xuhui	1.04	1.53	2.56	0.96	0.98	1.38
3	Changning	1.00	1.39	0.42	1.07	0.86	1.25
4	Jingan	1.76	1.27	0.43	0.98	0.86	3.84
5	Putuo	1.42	1.09	1.82	0.96	0.94	1.26
6	Zhabei	1.00	1.63	0.62	0.83	0.92	1.82
7	Hongkou	0.94	1.04	0.63	0.74	0.89	2.07
8	Yangpu	1.14	0.86	0.74	0.86	0.81	1.07
9	Minhang	0.74	0.72	1.00	1.07	1.02	0.31
10	Baoshan	0.75	1.05	0.63	0.92	1.06	0.42
11	Jiading	0.79	0.97	0.47	1.17	1.06	0.15
12	Pudong	0.65	0.86	3.73	1.16	1.07	0.28
13	Jinshan	0.69	0.82	0.14	1.09	1.10	0.11
14	Songjiang	0.57	0.40	0.99	1.36	1.19	0.13
15	Qingpu	0.59	0.36	0.57	1.14	1.09	0.09
16	Fengxian	0.64	0.46	0.93	1.17	1.19	0.13
17	Chongming	0.98	1.12	0.52	0.74	0.94	0.04

Through the IBM SPSS Statistics software, the K-mean cluster and hierarchical cluster methods were used for processing. The results are shown in following context.

According to the cluster analysis, there are two cluster are generated. The group I are defined as the rich resource area, The group II are defined as the poor resource area.

Group I: Huangpu, Jingan, Xuhui, Putuo, Changning, Zhabei, Hongkou, Yangpu, Chongming

Group II: Baoshan, Minhang, Jinshan, Jiading, Qingpu, Fengxian, Songjian, Pudong

B. Data Envelopment Analysis Approach

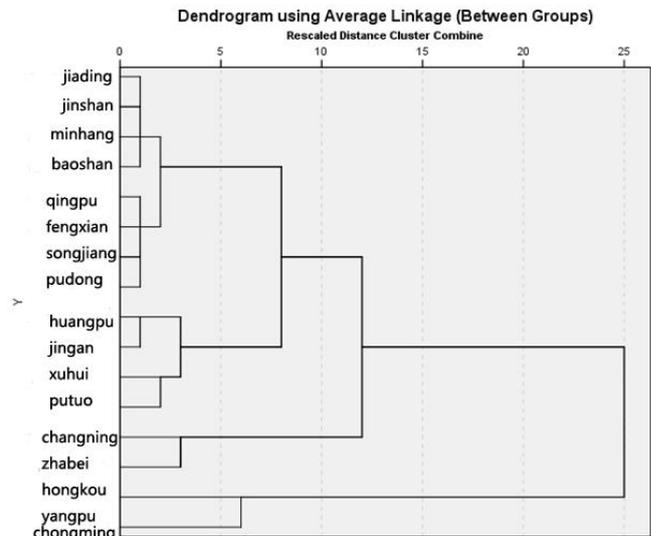


Fig. 2. The K-mean cluster and hierarchical cluster analysis results.

Base the cluster analysis results. The DEA approach is perform in order to assess the district resource allocation efficiency. The input/output indicators are defined as following:

TABLE IV: THE DEA MODEL INPUT OUTPUT INDICATORS AND DEFINITIONS

No.	Indicators	Code	Definitions
1	Total budget per capita	(TBC)	Input Indicator
2	Equipment budget per capita	(EBC)	Input Indicator
3	The ratio of student/teacher	(RST)	Input Indicator
4	Quota per class	(QC)	Output Indicator
5	Quota per school	(QS)	Output Indicator
6	Student Density per Km ²	(SD)	Output Indicator

By using the software DEA Solver, the Experiment methods are shown as following:

TABLE V: THE DEA MODEL METHOD ASSUMPTION

No.	Scale assumption	Orientation	Slacks calculating method
Strategy 1	CRS	Input	multi-stage
Strategy 2	CRS	Output	multi-stage
Strategy 3	VRS	Input	multi-stage
Strategy 4	VRS	Output	multi-stage

The results for efficiency evaluation are shown as following:

TABLE VI: THE DEA MODEL RESULTS

No.	District Name	Technical Efficiency			
		Strategy 1	Strategy 2	Strategy 3	Strategy 4
1	Huangpu	0.663	0.663	0.69	0.862
2	Xuhui	0.772	0.772	0.777	0.906
3	Changning	0.973	0.973	0.974	0.974
4	Jingan	1	1	1	1
5	Putuo	0.592	0.592	0.629	0.874
6	Zhabei	0.933	0.933	0.933	0.962
7	Hongkou	1	1	1	1
8	Yangpu	0.657	0.657	0.748	0.744
9	Minhang	0.753	0.753	0.827	0.873
10	Baoshan	0.867	0.867	0.883	0.929
11	Jiading	0.844	0.844	0.844	0.977
12	Pudong	0.852	0.852	0.914	0.917
13	Jinshan	1	1	1	1
14	Songjiang	1	1	1	1
15	Qingpu	1	1	1	1
16	Fengxian	0.944	0.944	1	1
17	Chongming	0.56	0.56	0.661	0.815
	Mean Efficiency	0.848	0.848	0.875	0.931

According the Technical Efficiency result, there are six districts named Jingan, Hongkou, Jinshan, Songjiang, Qingpu and Fengxian, those districts are considerable as 100% efficiency regarding the input and output indicator. The rest districts have the space to increase the input/output efficiency.

VI. DISCUSSION AND CONCLUSIONS

By the CA approach, the districts were classified by educational resource into rich or poor resource group. The results showed the rich resource group member was urban district except Chongming District. The poor resource group member was suburb district. This phenomenon implicit the urban sprawl zone resource is lack, the equilibrium should be considerable.

By the DEA approach, five districts were considered as 100% efficiency, there are two urban districts and the rest are suburb district. Comparing with the mean efficiency value, most of district's efficiency was higher than mean values. Meanwhile, the urban and suburb districts were mixed. It approve that the efficiency wasn't relating the educational source rich or poor.

With this evaluation framework, the government could make the planning of resource allocation for next year, for each indicator, the DEA model demonstrate the result of resource input slacks quantity, the slack value could be criteria for government.

REFERENCES

[1] OECD. Program for International Student Assessment introduction. [Online]. Available: <http://www.oecd.org>

[2] A. Charnes and W. W. Cooper, "The non archimedean CCR ratio for efficiency analysis," *European Journal of Operational Research*, vol. 15, pp. 333-334, May 1984.

[3] W. W. Cooper, Seiford, and K. Tone, "Introduction to data envelopment analysis and its uses: With DEA solver software and references," New York: Springer, 2006, ch. 3, pp. 37-41.

[4] K. Cullinane, T. Wang, D. Song, and P. Ji, "The technical efficiency of container ports: Comparing data envelopment analysis and stochastic frontier analysis," *Transportation Research Part A: Policy & Practice*, vol. 40, no. 4, pp. 354-374, Mar. 2008.

[5] A. Pergelova, Prior, and J. Rialp, "Marketing communication efficiency in the Spanish automobile sector: Analyzing the role of online advertising through DEA and stochastic frontiers," *Academia: Revista Latino americana de Administración*, vol. 41, pp. 91-108, Jan. 2009.

[6] Khodabakhshi, "An output-oriented super-efficiency measure in stochastic data envelopment analysis: Considering Iranian electricity distribution companies," *Computers & Industrial Engineering*, vol. 58, no. 4, pp. 663-671, Sep. 2010.

[7] Khodabakhshi, "An input-oriented super-efficiency measure in stochastic data envelopment analysis: Evaluating chief executive officers of US public banks and thrifts," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2092-2097, Nov. 2010.

[8] N. K. Avkiran, "Investigating technical and scale efficiencies of Australian Universities through data envelopment analysis," *Socio-Economic Planning Sciences*, vol. 35, no. 1, pp. 57-80, Mar. 2010.

[9] M. Abbott and Doucouliagos, "The efficiency of Australian universities: A data envelopment analysis," *Economics of Education Review*, vol. 22, pp. 89-97, Feb. 2010.

[10] C. S. Salerno, "On the technical and allocate efficiency of research-intensive higher education institutions," Doctoral dissertation, The Pennsylvania State University: University Park, pp. 259-261, Jun. 2002.

[11] J. Calhoun, "Data envelopment analysis of relative efficiencies of institutions of higher learning," Association for the Study of Higher Education, Portland, Jul. 2008.

[12] A. Afonso and M. Santos, "Public tertiary education expenditure in Portugal: a non parametric efficiency analysis," ISEG-UTL Economics Working Paper, n.05/DE/CISEP, Feb. 2004.

[13] S. Warning, "Performance differences in German higher education: empirical analysis of strategic groups," *Review of Industrial Organization*, vol. 24, pp. 393-408, Jan. 2004.

[14] J. Johnes, "Data envelopment analysis and its application to the measurement of efficiency in higher education," *Economics of Education Review*, vol. 25, pp. 273-288, Apr. 2006.

[15] B. Jongbloed and C. Salerno, "De bekostiging van het universitaire onderwijs en onderzoek in nederland: Modellen," *Thema's en Trends*, Jun. 2003.

[16] L. Cherchye and Abeele, "On research efficiency: A micro-analysis of Dutch university research in economics and business management," *Research Policy*, vol. 34, pp. 495-516, Jul. 2005.

[17] M. Castano and E. Cabanda, "Performance evaluation of the efficiency of Philippine Private Higher Educational Institutions: Application of frontier approaches," *International Transactions in Operational Research*, vol. 14, pp. 431-444, Jan. 2007.

[18] O. Joumady and C. Ris, "Performance in European higher education: A non-parametric production frontier approach," *Education Economics*, vol. 13, no. 2, pp. 189-205, Aug. 2005.

[19] G. Ferrari and T. Laureti, "Evaluating technical efficiency of human capital formation in the Italian university: evidence from Florence," *Statistical Methods and Applications*, vol. 14, no. 2, pp. 243-270, Jul. 2005.

[20] T. Laureti, "Modeling exogenous variables in human capital formation through a heteroscedastic stochastic frontier," *International Advances in Economic Research*, vol. 14, no. 1, pp. 76-89, Feb. 2008.

[21] M. Osei-BrysonK, "A hybrid clustering algorithm," *Computers and Operations Research*, vol. 34, no. 11, pp. 55-69, Feb. 2007.

[22] A. Jain, M. Murty, and P. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264-323, Oct. 2009.

[23] F. Murtagh, "A survey of recent advances in hierarchical clustering algorithms which use cluster centers," *Computer Journal*, vol. 58, pp. 236-244, Sep. 1983.

[24] J. McQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Jan. 1997, pp. 281-297.

[25] H. Bock, "Probability models in partitioned cluster analysis," *Computational Statistics and Data Analysis*, vol. 23, pp. 5-28, 2009.

[26] H. Eilat, B. Golany, and A. Shtub, "The project evaluation: An integrated DEA and balanced scorecard approach," *Omega*, vol. 36, no. 5, pp. 895-912, Aug. 2008.



Yuan Yi was born in Liaoning province, China, on Oct. 17, 1978. At 2000, he achieved the bachelor degree of chemistry engineering in Liaoning Technology University, Jinzhou City, China. At 2006, he got the master degree of ecology marine management at Vrije University Brussels, Belgium. The model algorithm is the major field. Until now, he is in the process for PhD study in management science, Shanghai University, Shanghai, China.

Yuan worked as a PhD student at Shanghai University Higher Educational Research Institute. His research interests include the educational management modeling application and developing.

Dr. Yuan Yi participated some projects which were funded by Shanghai government. The achievement is considering by related government bureaus when they publish or revise some relative policy.