
Competitiveness of Top 100 U.S. Universities: A Benchmark Study Using Data Envelopment Analysis (DEA) and Information Visualization

Gürdal Ertek[1], Bengi Tokdil[1], İbrahim Günaydın[1], Aytaç Göğüş[2]
[1] *Faculty of Engineering and Natural Sciences, Sabanci University, Tuzla, Istanbul, 34956, Turkey*
[2] *Center for Individual and Academic Development, Sabancı University, Tuzla, Istanbul 34956, Turkey*

ABSTRACT
This study presents a comprehensive benchmarking study of the top 100 U.S. Universities. The methodologies used to come up with insights into the domain are Data Envelopment Analysis (DEA) and information visualization. Various approaches to evaluating academic institutions have appeared in the literature, including a DEA literature dealing with the ranking of universities. Our study contributes to this literature by the extensive incorporation of information visualization and subsequently the discovery of new insights.

The main purpose of the study is creating an objective basis of assessment for the candidate students to use for university preferences. Meanwhile, the actionable insights obtained for the domain can guide university managers, as well as candidate students.

Keywords: Data Envelopment Analysis (DEA); Information Visualization; Education Policy.

INTRODUCTION

University education is not only about learning theoretical or technical information on a particular profession, but it also comprises of gaining a different perspective on life and leads to human development. In today's world, university education is not a privilege but rather almost a standard expectation for a successful professional career. Choosing a university and planning one's future upon this decision is thus a significant decision. This chapter presents a detailed benchmarking analysis of the top 100 US universities, as would be viewed from a high school student's perspective. The goal of our study is two-folds: Assisting students for their university selection as well as assisting university managers in improving their universities. Whichever the target audience, the results obtained in the study and the analysis performed can be packaged as an interactive decision support system (DSS) for the target audience.

Our study aims at offering an objective approach to assist prospective students in the in the complicated choice of a higher education institution and assist policy makers for their decision on institutional priorities. Prospective students face a large variety of institutional characteristics: acceptance rate, faculty member per student ratio, percentage of smaller classes, the average freshman retention rate, student evaluations of faculty, the average SAT score, public vs. private ownership, religious affiliations, high admission standards, minimal admission standards, Nobel Prize winning faculty, commuter school with adjunct faculty, single gender, coeducational, urban, suburban, pastoral campus, major sports powers, and many others (Black & Smith, 2004; Eff, Klein, & Kyle, 2010). In addition, each institution exhibits not only materialistic characters such as value of buildings, dorm capacity, average faculty salary, technology expenses, and expenses for other facilities, but also the student applicant's individual characteristics such as SAT score, GPA at high school, academic credentials, family income, student body characteristics, and other qualities (Black & Smith, 2004; Eff et al., 2010). Institutions select their outputs as quality attributes that best meet their mission as perceived by the administration and governing body. Higher education institutions can be modeled as competing, differentiated product producers as profit maximizers (Rosen, 1974), but this is problematic for non-profit higher education institutions, since non-profit institutions are assumed to maximize a value function over a vector of qualities (Eff et al., 2010).

There exist a multitude of rankings for colleges and universities on the bases of many different criteria. Rankings by U.S. news media include those by Consumer Digest (Consumer Digest), Forbes (Forbes), US News & World Review (US News), Washington Monthly (Washington Monthly), and Princeton Review (Princeton Review). International rankings include The Academic Rankings of World Universities by Shanghai Jiao Tong University (Shangai Ranking), Webometrics (Webometrics), and the Good University Guide by The Times of London (Good University Guide). However, all of the mentioned rankings are based on a weighted sum calculation.

Our study compares U.S. academic institutions with the Data Envelopment Analysis (DEA) methodology using 2010-2011 data for the top 100 four-year institutions of higher education as ranked by US News (US News). While we respect the US News ranking in selecting the group of universities to benchmark, we compute and present a new ranking based on the DEA methodology. The input is tuition. Outputs are acceptance rate, instructor per student ratio, and numbers of small classes (with fewer than 20 students). The DEA efficiency score provides an objective means of ranking institutions, not being biased with subjective weights used in other rankings.

The next section gives a background on the study, as well as a review of the related literature. The motivation for the study is presented from an educational perspective is presented. Later, analysis and results are presented, with a discussion of the insights gained into the domain. Finally, the chapter concludes with a summary of findings and prospects for future research.

BACKGROUND

Benchmarking

Benchmarking is the process of comparing an organization's business processes and performance with other organizations of its kind, to identify and implement improvements (Andersen & Jordan, 1998). Robert Camp (1989) developed a 12-stage approach for benchmarking, which consists of the following:

1. Select subject
2. Define the process

3. Identify potential partners
4. Identify data sources
5. Collect data and select partners
6. Determine the gap
7. Establish process differences
8. Target future performance
9. Communicate
10. Adjust goal
11. Implement
12. Review and recalibrate

There are two main reasons for the popularity of benchmarking in education: 1) Meeting the requirement for increased effectiveness and international competitiveness; 2) Increased interest in enhancing quality and the consequent growth of the quality (CHEMS, 1998). Most institutions of higher education desire to learn from each other and share aspects of good practice (CHEMS, 1998). In addition, benchmarking in higher education helps institutions to identify their comparative strengths and weaknesses and learn how to improve, as well as guide them in adopting best practices (Fielden 1997). The benchmarking methodology used in our study is Data Envelopment Analysis (DEA), which is summarized next.

Data Envelopment Analysis (DEA)

DEA is a "data oriented" analytical approach for evaluating and comparing the performance of a set of peer entities, referred to as Decision Making Units (DMUs), which convert multiple inputs into multiple outputs." (Cooper, Seiford, & Zhu, 2011). Besides generating benchmark results, DEA can also be employed to supply new insights into the domain of DMUs (Ulus, Kose, Ertek, & Sen, 2006; Ertek, Can, & Ulus, 2007; Ertek, Tunc, Kurtaraner, & Kebude, 2012).

DEA uses an optimization-based algorithm to determine three types of benchmark results for each of the entities within a group (Cooper, Seiford, & Tone, 2006): For each DMU in the group, the efficiency score value between 0 and 1 represents the relative performance of that entity (DMU) compared to the other entities in the group. The efficiency of a DMU increases with the generation of higher values of its outputs given lower values of its inputs. By definition, an efficient DMU has an efficiency score equal to 1, whereas an inefficient DMU has its score less than 1. The reference set of a DMU refers to the set of efficient DMUs that a DMU should benchmark itself against and take as example. The reference set for an efficient DMU consists of itself, whereas the reference set of an inefficient DMU consists of two or more efficient DMUs. The third and final result generated by DEA is the set of projections, which tells how much of each input a DMU should decrease and/or each output the DMU should increase such that it can become an efficient DMU.

There exist a multitude of DEA models, which differ in the underlying optimization models. Our study employs the BCC (Banker-Charnes-Cooper) Model by Banker, Charnes & Cooper (1984). The BCC model, similar to other DEA models, evaluates the efficiency of each DMU by solving a linear program. However, as an advantage, the BCC model can accommodate variable-returns-to-scale, meaning that model results do not change when the inputs or outputs are multiplied by constants. An output-oriented BCC model is constructed, meaning that it is assumed one has more control on the outputs for a given input value.

Information visualization

Data visualization refers to the visualization of data for understanding its content and discovering hidden patterns and insights. Data visualization helps analysts "understand the context and the detail together" and provides a "powerful way to reason about large data sets" (Myatt & Johnson, 2009). Visualization is a convenient method for knowledge discovery, since it does not require prior knowledge of any algorithm or sophisticated method by the analyst.

In 1990's, with developments in computer science such as increased computing patterns and display size (number of pixels), the graphical methods of 1980's (such as histograms, box plots, and scatter plots; Chambers, Cleveland, Kleiner, & Tukey, 1983) have elevated into a new level, being referred to as information visualization (Spence, 2001; Keim, 2002). The goal in information visualization is same as that of data visualization, but information visualization is particularly applicable and advantageous when the data is large-scale and complex. The new era of information visualization is built on ideas from data mining, statistics, and computer graphics, and expands and improves the visualization methods of 1980's. There are many more visualization schemes in information visualization compared to the mentioned earlier fields, enabling the analysis of data with different structures in diverse domains.

In this chapter, the combination and comparison of the DEA results with information visualization schemes brings fresh and insightful discoveries regarding the domain of higher education.

Earlier Work

Various approaches to evaluating and benchmarking academic institutions have appeared in the literature (Billaut, Bouyssou, & Vincke, 2009; Eff et al., 2010; Ehrenburg, 2003; Liu & Cheng, 2005; van Raan, 2005; Turner, 2008; Archibald & Feldman, 2008). Some of these studies specifically apply DEA for benchmarking: Archibald & Feldman (2008) compare institutions via the use of DEA to construct an efficient frontier for 187 institutions based on graduation rates, SAT scores, high school grades, percent full time faculty, and expenditures per undergraduate. Eff et al. (2010) compare 1,188 institutions for the 2000-2001 academic year by the use of DEA. Eff et al. (2010) choose the inputs as net price or tuition, fees, room, and board less per student financial aid and outputs as SAT score, athletic expenditures, and instructional expenditures, value of buildings, dorm capacity, and student body characteristics. The study of Eff et al. (2010) demonstrates the effectiveness of DEA analysis for evaluating the price-quality relationships offered by institutions of higher education in the U.S.

Once the DEA model is constructed and its results are obtained, information visualization methods can be applied for the analysis of these results. Ulus et al. (2006) analyze the transportation companies traded in NYSE (New York Stock Exchange) by visualizing the efficiency scores against various attributes of the companies. Ertek et al. (2007) analyze the apparel retail industry in Turkey. Ertek et al. (2012) introduce the graph visualization of the reference sets when comparing commercial wind turbines. Other studies that employ visualization for understanding and explaining DEA results are reviewed in Ertek et al. (2012).

The visual insights obtained through the visualization of DEA results enable not only knowledge discovery, but also hypotheses generation. The generated hypothesis can be tested using statistical methods. In this chapter, Mann-Whitney test has been applied for the comparison of group means to test a hypothesis which is suggested by visualization. The Mann-Whitney test uses rank data to compute the test statistics and does not require the data to come from a particular distribution (Conover, 1998). When the distribution of any of the groups does not follow normal distribution, the non-parametric Mann-Whitney test is advantageous over its parametric counterpart t-test.

METHODOLOGY

Data Envelopment Analysis (DEA)

The dataset for the study comes from the US News (US News). Firstly, the data was cleaned and brought to a convenient format to enable DEA and the subsequent data analysis. In cleaning the data, the taxonomy of dirty data in Kim et al. (2003) has been employed.

According to the DEA methodology, DMUs, inputs and outputs were clearly identified and defined in the DEA model within the DEA solver. The created DEA model includes one input and three outputs. Universities were taken as DMUs; tuition was taken as input; acceptance rate, instructor per student ratio, and the percentage of small classes (those with fewer than 20 students) were taken as outputs. The BCC input-oriented DEA model (BCC-I) was selected and run. One can decrease its input selecting a university while the outputs remain the same, which helps candidate students to choose the most suitable universities for themselves, given their budgets. Hence, the input-orientation is followed.
DEA computations were carried out using the Smart DEA Solver software (Akcay, Ertek, & Buyukozkan, 2012). The software generates the DEA results in a tabular structure, making it easy to analyze the results using readily available data mining and information visualization software.

Subsequent to DEA, the Orange data mining software (Orange; Curk et al., 2005) was used for knowledge discovery through information visualization.

RESULTS

In this section, the results obtained through the above analysis methodology are presented. We will present visualizations, discuss these visualizations, propose hypotheses based on these visualizations, and report the results of statistical hypothesis tests regarding the proposed hypothesis. The efficiency scores obtained through DEA are presented in the Appendix, and form the base of the first six analysis (Figures 1-6) and the final analysis (Figure 8). Figure 7 is based on the reference sets obtained through DEA.

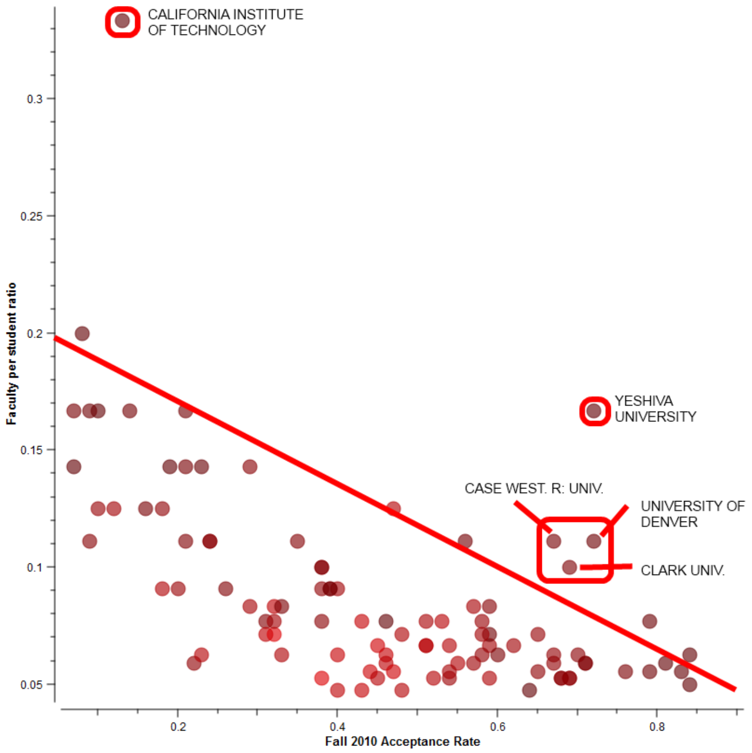


Figure 1. Faculty per student ratio vs. Fall 2010 acceptance rate

Figure 1 illustrates the relationship between acceptance rates and faculty per student ratio. The color of the data points denotes efficiency scores, where darker points represent more efficient DMUs (universities). As seen from the figure there are certain outlier universities. However the most obvious one is California Institute of Technology, which has the highest faculty per student ratio with a low acceptance rate to the university. Therefore, for a successful candidate student who can easily be accepted by many universities, California Institute of Technology can be suggested. Also Yeshiva University is another outlier, which has both relatively high acceptance rates and faculty per student ratio. Therefore for a moderate student, Yeshiva University can be suggested.

Clark University, Case Western Reverse University, and University of Denver are other outliers. These can be all suggested for relatively moderate or unsuccessful candidate students because of their higher faculty per student ratio and higher acceptance rates. Other than the identification of the outlier universities, there is another insight from this figure: There is a linear boundary, which envelops the remaining data points.

The conclusion from the boundary is that, except five outlier universities, faculty per student ratio has a natural linear boundary, which decreases with increasing acceptance ratios. In other words, given the acceptance ratio of a university, one can compute an upper bound on the faculty per student ratio, following a simple linear function. As a university accepts more students, the maximum value for faculty per student decreases. This linear boundary follows the function y=0.133-0.106x, where x is fall acceptance rate and y is faculty per student ratio. Furthermore, while the boundary is linear, the points below follow a nonlinear pattern.

In Figure 1, faculty per student ratio decreases drastically until acceptance ratio takes the value of 0.4, but then varies around a fixed value. This shows that for universities that have acceptance ratio greater than 0.4, faculty per student ratio will not degrade (decrease) systematically. Students, who give the most importance to faculty per student ratio can apply to universities with acceptance ratios greater than 0.4.
For managers, this nonlinear pattern suggests that, if possible, they should plan their student numbers such that faculty per student ratio is more than 0.1, and acceptance rate is less than 0.4.

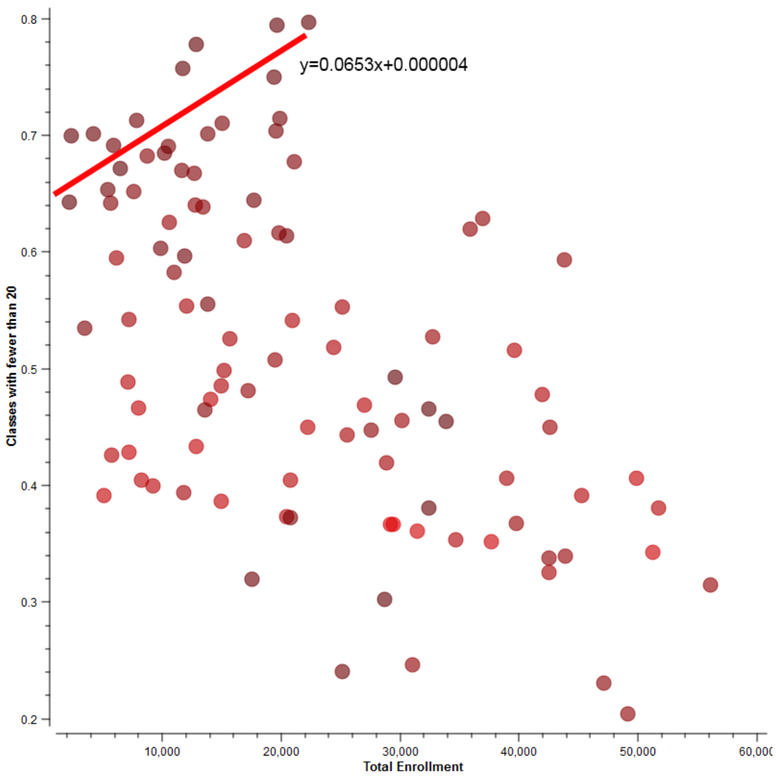


Figure 2. Classes with fewer than 20 students vs. Total enrollment (darker point colors denote higher efficiency scores)

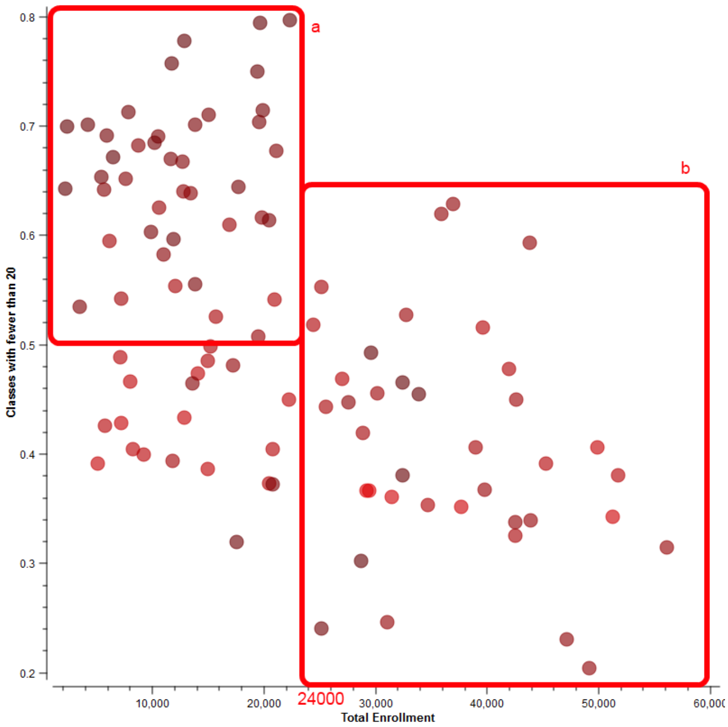


Figure 3. Clustering of efficient vs. inefficient universities

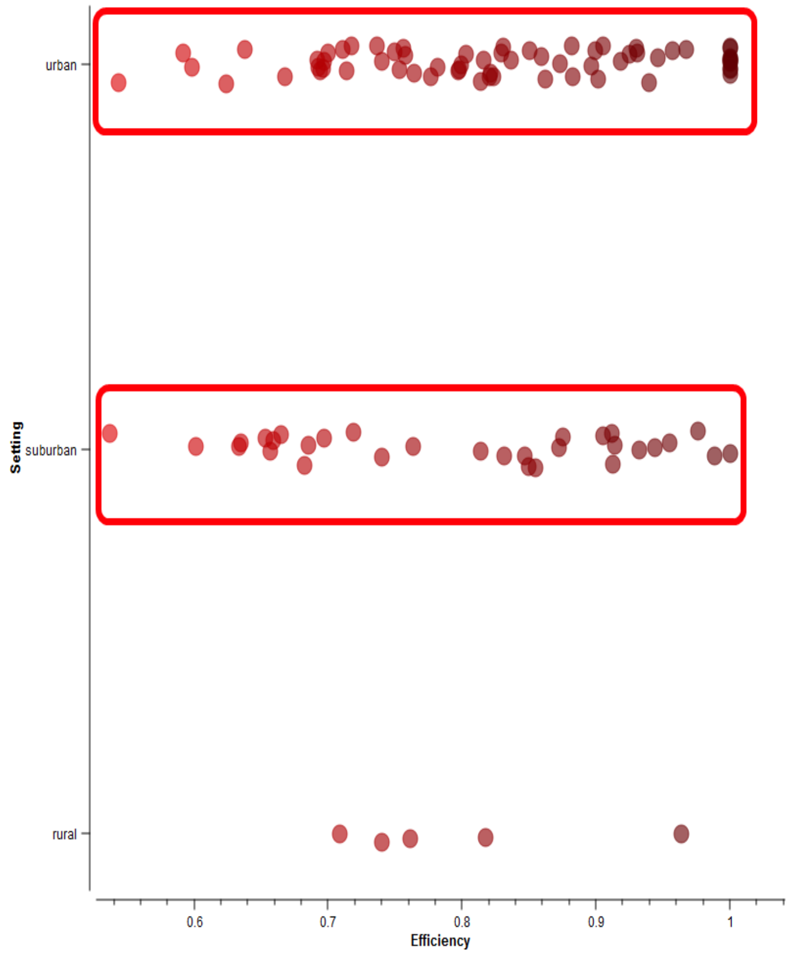
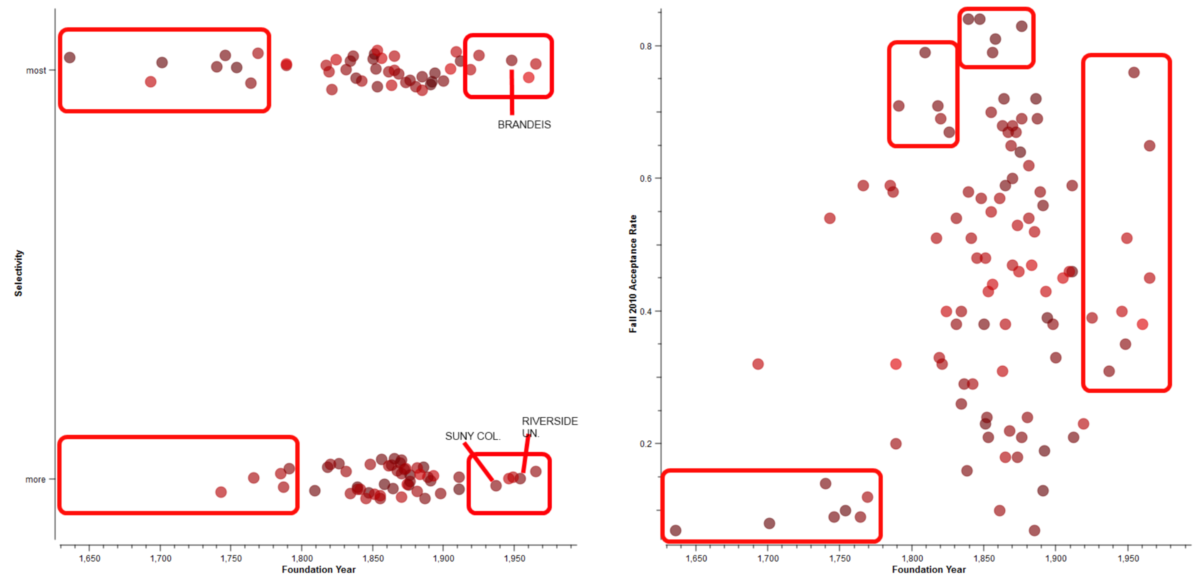


Figure 4. Comparison of efficiency scores of urban and suburban universities



(a) (b)

Figure 5. Efficiency scores with respect to Foundation year and quality metrics (a) Selectivity and (b) Fall 2010 Acceptance Ratio

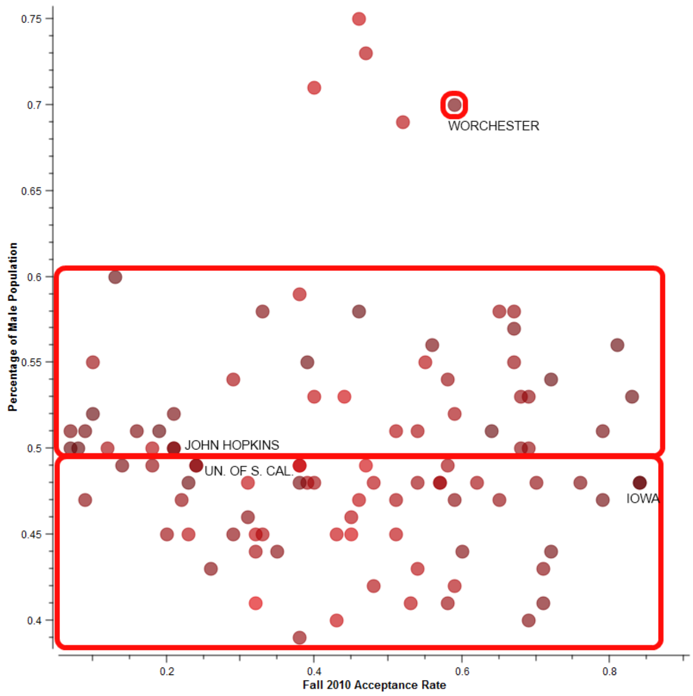


Figure 6. Percentage of Male Population vs. Fall 2010 Acceptance Rate (darker point colors denote higher efficiency scores)

The success of a university can also be measured the balance between enrollment and number of students in a class. In Figure 2, this relation is investigated through a scatter plot illustrating the efficiency of universities with regards to total enrollment and percentage of small classes. According to Figure 2, there is a linear pattern at left top for efficient universities (dark colored points). According to the regression analysis, efficient universities have established a relation between enrollment and percentage of small classes, which can be summarized with the linear equation $y=0.0653x+0.000004$, where x denotes total enrollment and y denotes the optimal number for classes with fewer than 20 students.

Other than the linear pattern for efficient universities, the relation of total enrollment and classes with fewer than 20 students has one more insight which can be shown in Figure 3. There are two distinct set of universities with distinguishing characteristics: In set a, the efficiency of universities seems to be higher. When the Mann-Whitney test is implemented for comparing the efficiency scores of two groups, p for the test comes up to be $p=0.000039$, very strongly suggesting that the efficiencies in set a are higher than that in set b.

The analysis based on Figure 3 provides an additional insight: There is a virtual boundary between the two clusters. This virtual boundary, at the vertical line where total enrollment is 24000, separates the two sets of universities. The cluster set of universities with less than 24000 students include the efficient set of universities.

The next analysis, shown in Figure 4, investigates whether there is a difference in efficiency of urban and suburban universities. Figure 4 displays a colored scatter plot (with jittering) that maps the two groups of universities against their efficiency scores. The Mann-Whitney test for the comparison of the efficiency scores in the two groups yields $p=0.17$, suggesting no statistically significant difference. This result states that, from the perspective of a candidate student, who wishes to maximize the outputs for the tuition s/he pays (input), there should not be a preference based on the urban/suburban setting of the universities.

Figure 5 demonstrates the effect of foundation year upon quality of education (as indicated by Selectivity and Fall 2010 Acceptance Ratio) and efficiency scores of universities. In Figure 5a, it is seen that among the universities which are founded before 1800s and are most selective, almost all have their efficiency scores are equal to 1. Similarly, in Figure 5b, the acceptance rate of universities which are founded before 1800s are very low. Moreover, it is seen that the efficiency scores of the universities founded after 1950s are not high, as indicated with light colors. In Figure 5a, it is seen that Brandeis University, University of

Riverside and Sunny College are universities founded after 1900s that have higher efficiency scores. Furthermore, Figure 5b illustrates that universities that are founded around 1850s are the most efficient ones when compared to those in other clusters. Those ones are even more efficient than the universities which ones were founded around 1800s. We did not apply sample comparison tests regarding these patters, since the number of universities in the groups were too few.

Figure 6 illustrates the relation between efficiency and percentage of male population. %50 can be taken as boundary and balance situation in terms of male-female population. There are obvious outlier universities which have percentage of male population between %65 and %75. In this cluster of particular outlier universities, Worchester University seems to be the most efficient university comparing to the others. Other than Worchester, there is not an obvious efficient outlier. However, Iowa University is the most efficient university whose female population is higher than male population. Meanwhile, John Hopkins and University of Southern California are comparatively efficient universities, which have balance in the distribution of male and female populations.

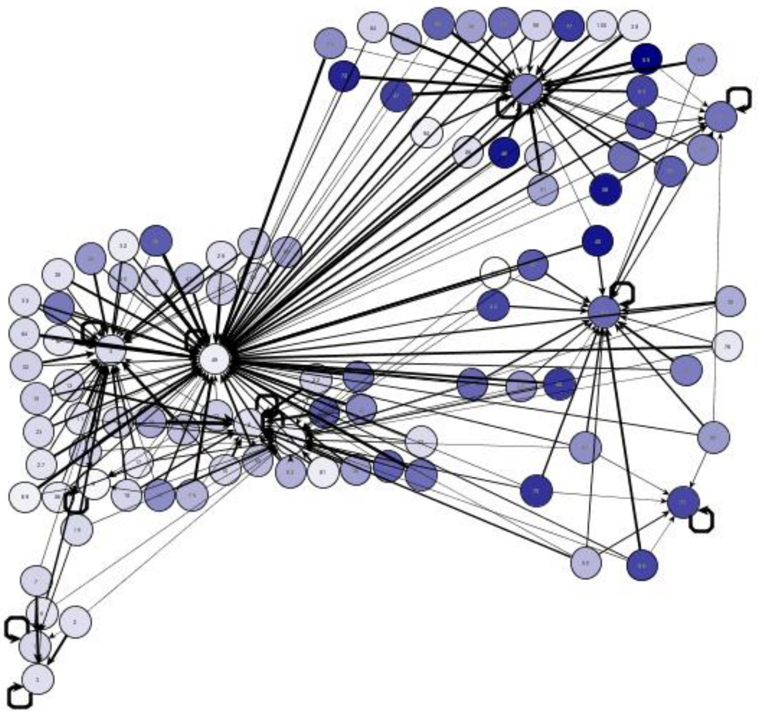


Figure 7. Visualization of reference sets (darker node colors denote larger values of Total Enrollment)

In Figure 7, the effect of number of students upon efficiencies of universities is demonstrated. Each node in the graph denotes a university. Arcs are drawn from a node to another one if the university denoted by the former node has the one denoted by the latter in its reference set.

The arc thickness shows the weight. The visualization has been constructed in yEd software (yWorks) using organic layout algorithm, which tries to minimize arc crossings and cluster related nodes. The circles with darker blue show higher values of Total Enrollment while the lighter ones shows the universities with less enrollment. The universities with similar features are clustered closer to each other. Also the arc points out the university that should be taken as reference. According to the Figure 9, it can be observed that the crowded universities should decrease their student body size by taking as reference the universities with similar features but fewer students. As a result of decreasing total enrollment, the universities can increase their education quality.

A final question that deserves investigation is the following: "How much are the results obtained through DEA similar to the rankings of US News?" To this end, a scatter plot is constructed with the US News rank of the universities on the x axis and the DEA ranks on the y axis. The plot is given in Figure 8, and includes only the inefficient DMUs. The plot is not following a linear line, suggesting that the ranks are different. A formal statistical test, namely the Spearman correlation test (Conover, 1998), gives the p value of 0.024, strongly suggesting that the ranks are indeed different. So, DEA has given significantly different rankings than US News. The difference is due to the different ways the ranks are constructed: The US News ranking does not consider input/output relation and does not measure quality as a function of tuition, whereas DEA gives great importance to the return on tuition. Therefore, in DEA, it is possible for a mediocre university to have a high ranking just because it gives a high return on the input (tuition) for even one of the outputs.

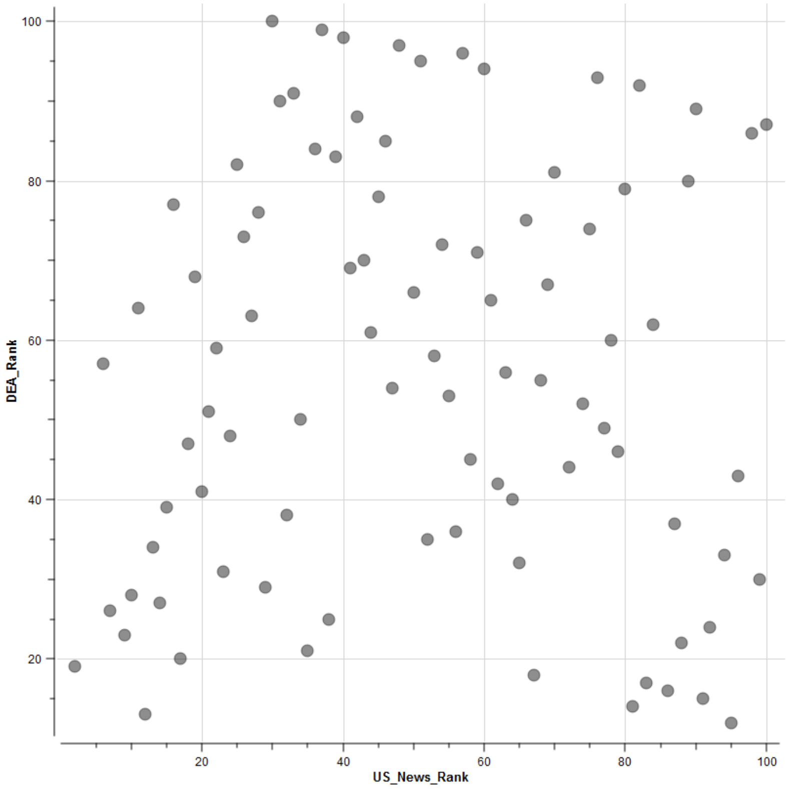


Figure 8. Comparison of ranks obtained with DEA (DEA_Rank) with those in US News.

CONCLUSIONS

The results on the relationship between acceptance rates and faculty per student ratio indicate that colleges should tend to decrease their acceptance rates for the sake of the quality of education. The quality of education at the university can be measured the balance between enrollment and number of students in a class. Findings suggest that efficiency-enhancing policies can be implemented. Legislators and education officials can address four strategies to improve the efficiency at higher education. These are 1) Increasing faculty per student ratio 2) Preferring and keeping small classes for each course 3) Preferring and keeping total enrollment below 24000 students, and 4) Achieving somewhat more balanced ratios for male and female students. Findings can influence the shape of higher education organizations, e.g. deciding acceptance rates, determining numbers instructor per students, and arrange class sizes. Figure 9 summarizes the strategies to improve efficiency at higher education. The university rankings have influences on national policy and are shaping institutional decision making and behavior. Therefore, this study also provides suggestions for policy makers. Figure 9 suggests some steps to follow before deciding and implementing the strategies for enhancing efficiency at higher education. Policy does matter and determined strategies for making policy and regulations are important.

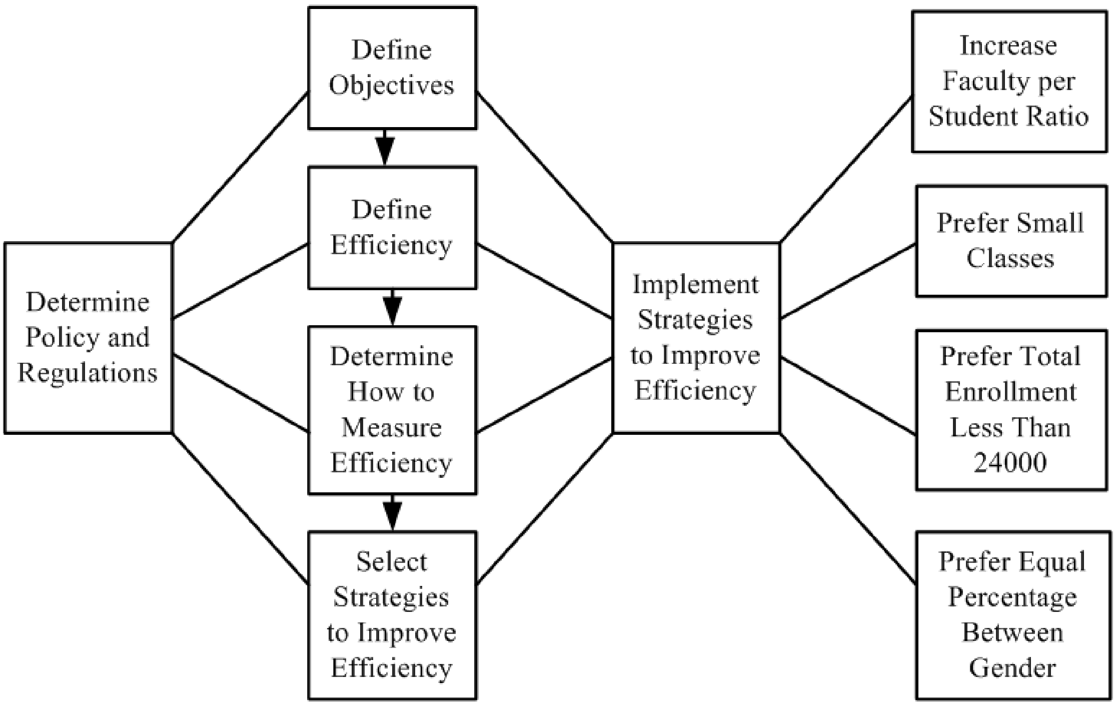


Figure 9. Suggested Steps and Strategies to Improve Efficiency at Higher Education

The DEA scores are helpful both to prospective students and to policy makers. The DEA scores aid policymakers' attempts to measure the cost effectiveness of academic institutions. The DEA scores also suggest a means of identifying peer institutions to take as reference, through the reference sets. The results also show that DEA rankings can be greatly different than those in the popular media.

Over the past decade, colleges and universities have used more resources to educate each graduating student to improve the quality of learning. The tuitions are increased every year but paying more tuition is not a grantee for the quality of students graduating from institutions. While finance and quality of higher education are major concern of governments, students, public policy makers, and higher education leaders, an efficiency-based perspective is not the common approach. It should be noted that the efficiency-based perspective provided in our study encompasses quality, as well as the cost of education. DEA scores can help policymakers' attempts to measure the cost effectiveness of diverse institutions in relation with quality criteria appropriate to their missions and within their resources. Since DEA takes into account quality characteristics, this information could also lead to better informed choices by prospective students. Further research can compare the institutions with the use of DEA methodology by considering other inputs such as average faculty salary and considering other outputs such as the average freshman retention rate and scores from the student evaluations of faculty as indicators of quality criteria.

REFERENCES

Akcay, A.E., Ertek, G., & Buyukozkan G., (2012). Analyzing the solutions of DEA through information visualization and data mining techniques: SmartDEA framework. *Expert Systems with Applications, 39,* 7763–7775.

Andersen, B., & Jordan, P. (1998). Setting up a performance benchmarking network. *Journal of Production Planning & Control: The Management of Operations, 9*(1), 13-19.

Archibald, R.B. & Feldman, D.H. (2008). Graduation rates and accountability: regressions versus production frontiers. *Research in Higher Education. 49*(1): 80-100.

Banker, R.D., Charnes, A. & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Journal of Management Science, 30,* 1078–1092.

Billaut, J., Bouyssou, D., & Vincke, P. (2009). Should You Believe in the Shanghai Ranking? An MCDM View. Retrieved August 6, 2012 from http://smai.emath.fr/IMG/pdf_RI296_Shanghai.pdf

Black, D.A. & Smith. J.A. (2004). How Robust is the Evidence on the Effects of College Quality? Evidence from Matching. *Journal of Econometrics 121*(1), 99–124.

Camp, R. (1989). *Benchmarking: The search for industry best practices that lead to superior performance.* ASQC Quality Press.

Chambers, J.M., Cleveland, W.S., Kleiner, B., & Tukey, P.A. (1983). *Graphical methodsfor data analysis*. Pacific Grove, CA: Wadsworth & Brooks/Cole Publishing Company.

Conover, W.J. (1998), *Practical nonparametric statistics*. John Wiley & Sons.

CHEMS (Commonwealth Higher Education Management) (1998). *Benchmarking in Higher Education: An International Review.* Twente: CHEMS.

Consumer Digest. The Hidden Costs of Higher Education; Plus: Consumers Digest's Top 100 College Values. Retrieved August 6, 2012, from http://www.consumersdigest.com/money/the-hidden-costs-of-higher-education.

Cooper, W.W., Seiford, L.M., & Tone, K. (2006). *Introduction to Data Envelopment Analysis and its Uses.* New York, NY: Springer.

Cooper, W.W., Seiford, M.L., Zhu, J. (2011) Data Envelopment Analysis: History, Models, and Interpretations. *International Series in Operations Research & Management Science. 164*, 1-39.

Curk, T., Demšar, J., Xu, Q., Leban, G., Petrovič, U., Bratko, I., Shaulsky, G., & Zupan, B.

(2005). Microarray data mining with visual programming. *Bioinformatics*, *21*(3): 396-8.
Eff, E.A., Klein, C.C., & Kyle, R. (2012). "Identifying the Best Buys in U.S. Higher Education." *Research in Higher Education* (February 22). doi:10.1007/s11162-012-9259-2
Ehrenberg, R. G. (2003). *Tuition Rising: Why College Costs So Much*. Cambridge: Harvard University Press.
Ertek, G., Can, M. A., & Ulus, F. (2007). Benchmarking the Turkish apparel retail industry through data envelopment analysis (DEA) and data visualization. *EUROMA 2007, Ankara, Turkey.*
Ertek, G., Tunç, M.M., Kurtaraner, E., & Kebude, D. (2012). Insights into the Efficiencies of On-Shore Wind Turbines: A Data-Centric Analysis. *INISTA 2012 Conference. July 2-4, 2012, Trabzon, Turkey.*
Fielden, J. (1997). *Benchmarking University Performance,* CHEMS monograph.
Good University Guide. http://www.thetimes.co.uk/tto/public/gug/.
Forbes. America's Top Colleges. http://www.forbes.com/top-colleges/list/ .
Keim, D.A. (2002). Information visualization and data mining. *IEEE Transactions on Visualization and Computer Graphics, 8*(1), 1-8.
Kim, W., Choi, B.J., Hong, E.K., Kim, S.K., & Lee, D. (2003). A taxonomy of dirty data. *Data Mining and Knowledge Discovery, 7*(1):81–99.
Liu, N.C. & Cheng, Y. (2005). The Academic Ranking Of World Universities. *Higher Education in Europe*, *30*(2):127-136.
Myatt, G.J., & Johnson, W.P. (2009) *Making Sense of Data II: A Practical Guide to Data Visualization, Advanced Data Mining Methods, and Applications.* Wiley-Blackwell.
Princeton Review. College Rankings. http://www.princetonreview.com/college-rankings.aspx
Orange. http://orange.biolab.si/.
Rosen, S., 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy, *82*(1), 34-55.
Shangai Ranking. http://www.arwu.org.
Spence, R. (2001). *Information visualization*. ACM Press, 2001.
Turner, D. (2008). World university rankings. *International Perspectives on Education and Society 9*, 27-61.
Ulus, F., Kose, O., Ertek, G., & Sen, S. (2006). Financial benchmarking of transportation companies in the New York Stock Exchange (NYSE) through data envelopment analysis (DEA) and visualization. *4th International Logistics and Supply Chain Congress*, Izmir, Turkey.
US News. http://colleges.usnews.rankingsandreviews.com/best-colleges.
van Raan, A.F.J. (2005). Fatal Attraction: Ranking of Universities by Bibliometric

Methods. *Scientometrics*, 62, 133-145.
Washington Monthly. College Guide. http://www.washingtonmonthly.com/college_guide/rankings_2011/national_university_rank.php.
Webometrics. http://www.webometrics.info.
yWorks. http://www.yworks.com.

KEY TERMS & DEFINITIONS

Data Envelopment Analysis (DEA): A non-parametric analytical method for benchmarking a group of entities.
Information Visualization: The field of computer science that works with the visualization of large-scale complex data for discovering new useful knowledge.

APPENDIX A. Efficiency Scores for the Top 100 Universities Listed by US News

Efficiency	Name	State
1.00	Harvard University	MA
1.00	Yale University	CT
1.00	Columbia University	NY
1.00	California Institute of Technology	CA
1.00	University of Chicago	IL
1.00	Yeshiva University	NY
1.00	Brigham Young University--Provo	UT
1.00	University of Iowa	IA
1.00	SUNY College of Environmental Science and Forestry	NY
1.00	University of Missouri	MO
1.00	Iowa State University	IA
0.99	University of Colorado	CO
0.98	Northwestern University	IL
0.97	University of Tulsa	OK
0.96	Miami University--Oxford	OH
0.96	University of Denver	CO
0.96	Auburn University	AL
0.95	Worcester Polytechnic Institute	MA
0.94	Princeton University	NJ
0.94	Rice University	TX
0.93	University of Rochester	NY
0.93	Drexel University	PA
0.93	University of Pennsylvania	PA
0.93	St. Louis University	MO
0.92	Case Western Reserve University	OH
0.91	Stanford University	CA
0.91	Washington University	MO
0.91	Duke University	NC
0.90	Tufts University	MA
0.90	University of California--Riverside	CA

0.90	Carnegie Mellon University	PA
0.90	Syracuse University	NY
0.90	Clark University	MA
0.88	Johns Hopkins University	MD
0.88	Tulane University	LA
0.87	Pepperdine University	CA
0.87	University of Vermont	VT
0.87	Brandeis University	MA
0.86	Brown University	RI

Efficiency	**Name**	**State**
0.86	Southern Methodist University	TX
0.85	Emory University	GA
0.85	Northeastern University	MA
0.85	University of Massachusetts—Amherst	MA
0.85	Michigan State University	MI
0.84	Texas A&M University--College Station	TX
0.83	University of California--Santa Cruz	CA
0.83	Vanderbilt University	TN
0.83	University of Southern California	CA
0.82	Indiana University--Bloomington	IN
0.82	New York University	NY
0.82	University of California--Berkeley	CA
0.82	Virginia Tech	VA
0.82	Ohio State University--Columbus	OH
0.81	University of Illinois--Urbana-Champaign	IL
0.81	Clemson University	SC
0.80	Purdue University--West Lafayette	IN
0.80	Massachusetts Institute of Technology	MA
0.80	Boston University	MA
0.80	Georgetown University	DC
0.78	University of Alabama	AL

0.78	University of Wisconsin--Madison	WI
0.76	Marquette University	WI
0.76	Wake Forest University	NC
0.76	Dartmouth College	NH
0.76	University of Pittsburgh	PA
0.76	George Washington University	DC
0.75	Rutgers, the State University of New Jersey--New Brunswick	NJ
0.75	University of Notre Dame	IN
0.74	University of Miami	FL
0.74	University of Washington	WA
0.74	University of Connecticut	CT
0.74	Fordham University	NY
0.72	University of Virginia	VA
0.72	Baylor University	TX
0.71	University of Georgia	GA
0.71	University of Michigan--Ann Arbor	MI
0.71	Cornell University	NY
0.70	Pennsylvania State University--University Park	PA

Efficiency	**Name**	**State**
0.70	University of Delaware	DE
0.70	Stevens Institute of Technology	NJ
0.70	University of Minnesota--Twin Cities	MN
0.69	University of California--Los Angeles	CA
0.69	Lehigh University	PA
0.69	Georgia Institute of Technology	GA
0.69	University of California--Irvine	CA
0.68	Texas Christian University	TX
0.67	University of San Diego	CA
0.66	University of California--Santa Barbara	CA

0.66	Binghamton University--SUNY	NY
0.66	Boston College	MA
0.65	College of William and Mary	VA
0.64	American University	DC
0.63	Colorado School of Mines	CO
0.63	University of Florida	FL
0.62	Rensselaer Polytechnic Institute	NY
0.60	University of Maryland--College Park	MD
0.60	University of Texas--Austin	TX
0.59	University of California--Davis	CA
0.54	University of California--San Diego	CA
0.54	University of North Carolina--Chapel Hill	NC