University Performance Evaluation Approaches: The Case of Ranking Systems

Loukas N. Anninos, Ph.D Candidate in Business Administration

Department of Business Administration, University of Piraeus, Hellas

Holder of Scholarship, Onassis Foundation

lanninos@unipi.gr

Keywords: Performance, Evaluation, Higher Education, Ranking Systems

Category: Literature Review

Introduction and Paper Objectives

During the last years, there has been extensive argumentation regarding university accountability, the evaluation of their performance (in the educational and administrative operation) and the publication of results with a view to more objective decision making (Ewell, 1999; Banta, Borden, 1994; Fuhrman, 1999, 2003; Pounder, 1999; Wakim, Bushnell, 1999; King, 2000; Goertz, Duffy, 2001; Welsh, Dey, 2002; Welsh, Metcalf, 2003; Bolton, 2003; Black, Briggs, Keogh, 2001). Decisions may be taken by individuals (e.g., students) aiming at choosing a university for studies, by the state aiming at a rational base to allocate resources and an imprint of higher education competitiveness or by the institutions themselves aiming at introducing changes and improvement wherever necessary. Moreover, universities do not constitute individualised organizational units but they operate and affect the wider economic and social system in which they belong. They are therefore accountable a) to the academic staff they employ (work in a suitable working climate and have great opportunities for scientific advancement), b) to the state (use of resources productively (efficiently and effectively) and c) the students and the society (comprehensive educational experience, scientific education and professional training to acquire quality of life) (Vidovich and Slee, 2001; Löfström, 2002, Corbett, 1992). Consequently, the evaluation of their performance proves to be a highly significant process for university institutions with many receivers of its results.

Baring in mind, a) university accountability towards the state and all stakeholders (that engage in institutional goal setting and operation and are influenced by their results), b) globalization that encourages the mobility of academic staff and students and hence stresses the need for international comparability of higher education systems, study programmes and degrees, c) the European objectives to create the area of European Higher Education which presupposes evaluation, d) global competition of higher education institutions in order to create attractive educational multi-cultural environments and the trend towards university collaboration and e) the
opportunity to improve and eliminate institutional weaknesses, universities should have a suitable and reliable management system with processes and mechanisms of performance measurement that would allow comparisons and improvement (Wakim, Bushnell, 1999; Pounder, 1999; Al-Turki, Duffuaa, 2003, Diamond, 2002; Meyerson, Massy, 1994; Welsh, Metcalf, 2003).

The effective management of any higher education system presupposes evaluation of results in institutional, departmental and study programme level. So, many evaluation approaches have been developed and successfully implemented globally with similarities and differences. Therefore, the present paper aims to present the different university performance evaluation approaches used internationally and examine the scientific correctness and suitability of the most common ranking systems, based on the presentation of issues regarding ranking systems that are raised in literature.

**Approaches of University Performance Evaluation**

The evaluation of university performance is a basic priority of the state, which means designing the necessary legal framework for university evaluation, establishing independent actors to undertake the evaluation procedure and developing performance evaluation systems (this may be also left to institutions, experts or both).

University performance evaluation is not a new issue. Institutional evaluations were undertaken in American universities almost 100 years ago while in Europe, France was the first country that initiated comprehensive institutional evaluation in 1984 (Hämäläinen et al., 2001). However, the differences in university type and profile, the different external environment that each university operated and the priorities of the higher education systems of each country contributed in the development of different approaches for university performance evaluation. Moreover, the difficulties in the precise definition of certain elements in the discipline of higher education have also contributed in the development of many definitions, processes and systems of evaluation.

According to literature, university performance evaluation is achieved through:

- Typical Evaluation, focusing a) in the quality of a subject in all study programmes that the subject is taught (for example, the subject “total quality management” in a business administration study programme), b) in the study programme, c) in the quality of an institution in every aspect of each operation (for example, educational or administrative) and d) the quality of a specific theme, that is a practice within higher education (for example students’ summer training programme in organizations) (DEI, 2003).

- Accreditation is the procedure by which a private or a state-independent actor evaluates the quality of an institution or a study programme with the view to certify that it meets specific and pre-defined standards (Vlasceanu et al., 2004). The result of the accreditation procedure will provide the awarding of a status, a recognition or a license for operation for a certain period of time. It may include an initial self study and external evaluation by experts. Its main objective is to maintain and improve quality in a higher education institution, study programme or course (Di Nauta et al., 2004).

- Audit: It is the process by which it is examined if the mechanisms and procedures that assure quality within an evaluation unit are present, are functioning properly and are effective. It focuses on the accountability and examines whether the stated objectives are being achieved. The reasons for quality audit include the evaluation of performance of quality assurance
systems and quality monitoring procedures, the assurance that units are responsible for quality, the initiation of improvements in the priority setting procedure and the facilitation of decision making. It also helps towards learning and improvement along with university accountability (Hämäläinen et.al, 2001; Vlasceanu et.al., 2004).

- **Benchmarking:** According to Vlasceanu et.al., (2004), benchmarking is a systematic method to collect and present information regarding the performance of organizational units and allow comparisons with the aim to establishing best practices, identifying performance weaknesses and strong points. Benchmarking is a diagnostic, self-assessment and learning tool at the same time, while on the other hand it constitutes a dynamic process of learning and performance comparison (Epper, 1999). Benchmarking may be internal, external competitive, external collaborative, external cross sectional and implicit and its methodology can be based on an excellence model, be horizontal or vertical or even be based on specific performance indicators sets (Alstete, 1995). Its main idea is to supply the institutional administration with an external reference point or a standard to evaluate quality or the cost of internal activities, practices and procedures (Hämäläinen et.al, 2002).

- **Ranking Systems:** It is an established technique used to present the ranking of a university in comparison with other universities in terms of their performance. They provide information to students, university administration and stakeholders regarding the quality of universities. Even though there are many problems regarding their methodology and the scientific base and validity of the systems, they are still popular and a means of initiating improvements (sometimes only on the surface) within institutions.

- **Data Envelopment Analysis (DEA):** It is a linear programming technique used when there are many inputs and outputs but no clear functional relationship between the two. It is a tool for evaluating relative efficiency (Kocher et.al, 2006). DEA permits the analysis of multiple input and output factors at the same time (Rickards, 2003). There are two types of linear programming techniques used for performance evaluation at university level, which focus on cost efficiency, research productivity or aggregate performance (Abbott and Doucouliagos, 2003; Johnes 1996; Johnes and Johnes 1995; Kao, 1994; Muniz, 2002; Post and Spronk, 1999; Ruggiero et.al, 2002). These researchers use regression analysis and DEA to evaluate performance. There are also studies that compare teaching and research performance of departments (Gander, 1995; Sinuany –Stern et.al, 1994) and studies that use ranking systems as a base for DEA models (Sarrico et.,al 1997; Breu and Raab 1994).

**Ranking Systems: an inadequate university performance evaluation approach**

In every performance evaluation approach, there are some issues critical for reliability and success that need to be addressed. These issues are relevant to the actor that is responsible for the evaluation, the object of evaluation, the orientation and mission of each institution and whether these factors are taken into consideration, the reason for evaluation, the frequency of evaluation and the methodology followed, the values of scientific validity and the dynamic nature of the evaluation system so that it keeps pace with changes and developments regarding higher education.

The appearance of the rankings systems can be traced in 1865 to European studies that aimed to define whether environment or heredity was the determining factor in producing man of genius (Hattendorf, 1996). It was attempted to assess the quality of institutions and affiliated scholars in science and medicine. The results
influenced the thinking of educators regarding quality assessment. During the period between 1925 and 1979, six multidisciplinary reputational rankings of graduate departments in the US were published (Hattendorf, 1996). The Keniston Study in 1959 used departmental ratings to produce institutional ratings. In 1982 the 5 volume Assessment of Research-Graduate Programmes in the US was published. However, it was not until the eighties that a proliferation of educational rankings occurred in the US and Canada. Since then a significant number of rankings systems have been developed, national or international in scope, following different methodologies (based on reputation, citation analysis, faculty productivity, statistics) to measure institutional, departmental or individual quality. Ranking systems like university institutions differ from country to country. Furthermore, the necessary data can be primary or secondary; there can be differences in the indicators used and the statistical analysis entailed.

It’s worth mentioning that educational rankings attract considerable attention from students, employers as well as institutions, even though there exists heavy scepticism regarding the scientific base of the systems (Merisotis, 2002). The following table (Table I) presents the ranking systems that have been developed and used globally:

### Table I: Ranking Systems of the World

<table>
<thead>
<tr>
<th>Country</th>
<th>Ranking System(s)</th>
<th>Country</th>
<th>Ranking System(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>Good University Guide</td>
<td>Canada</td>
<td>Macleans’s University Ranking</td>
</tr>
<tr>
<td></td>
<td>Melbourne Institute</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td>The Times Good University Guide</td>
<td>Poland</td>
<td>Perpektywy-Rzeczpospolit Uniwersytet</td>
</tr>
<tr>
<td></td>
<td>Times Higher Education Supplement</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Times Higher Education Supplement / w</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Guardian University Guide</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Sunday Times System</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States of America</td>
<td>US News and World Report (USNWR)</td>
<td>Japan</td>
<td>Daigaku Ranking</td>
</tr>
<tr>
<td></td>
<td>Washington Monthly</td>
<td></td>
<td>Recruit LTD Diamond</td>
</tr>
<tr>
<td></td>
<td>The Center (University of Florida)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>Shanghai Jiao Tong</td>
<td>Spain</td>
<td>Excellencia Ranking 2005</td>
</tr>
<tr>
<td></td>
<td>NETBIG Competence Limited</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GIMS (Guangdong Institute of Management Science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RCCSE (Research Center for China Science Evaluation-Wuhan University)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CDGDC (China Academic Degrees and Graduate Education Development Center)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CUAA (Chinese Universities Alumni Association)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CIES (Shanghai Institute of Educational Science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russia</td>
<td>Career Journal Reg 631/26-02-2001</td>
<td>Italy</td>
<td>La Republica</td>
</tr>
<tr>
<td>Germany</td>
<td>CHE / DAAD</td>
<td>Asia</td>
<td>Asiaweek (until 2003)</td>
</tr>
</tbody>
</table>

Note: Rankings systems that are used to rank graduate schools or are discipline specific, are excluded from the table.

In bibliography there have been four comparative studies of ranking systems by Provan and Abercromby (2000), Dill and Soo (2004), Van Dyke (2005) and Usher and Savino (2006). More specifically, Provan and Abercromby (2000) compare USNWR, Asiaweek, THES, Macleans and Australia Good University Guide, refer to
the criticism of the systems by academics underlining the fact that many universities participate in the rankings to benefit from publicity and attract students. Critique is exercised on the selection of indicators, the assignment of weights and statistical insignificance of difference between institutions. They go on to say that there is lack of objectivity in the selection of indicators and inconsistent methodologies.

Dill and Soo (2004) criticise rankings systems regarding statistical validity, the selection of indicators that reflect quality and the negative impact on university performance. They concentrate on USNWR, Australian Good University Guide, MacLeans, Times Good University Guide and Guardian University Guide. They examine validity, comprehensiveness, comprehensibility and functionality of the systems and reach the conclusion that the systems can be supplemented with other indicators and reflect the quality of an institution in a better way.

Van Dyke (2005) does a detailed presentation and comparison of ranking systems (Asiaweek, TheCenter, CHE, Good Guides, The Guardian, Macleans, Melbourne Institute, Perspektywy, The Times and USNWR) regarding indicators and attributes the difference of the systems to the variety of objectives, systems, culture and availability of data.

Finally, Usher and Savino (2006) compare 19 ranking systems from Australia, Canada, China, USA, Hong-Kong, Italy, Poland, Germany, Spain and the United Kingdom. They pinpoint the fact that the difference in the content of the systems can be ascribed to the geographical location and culture, they refer to the standardization issue of results. However, there is agreement on the best institutions and category based rankings. International ranking systems can be complemented with indicators that would allow inter-institutional performance comparison.

The above studies present comprehensively the comparison of the systems and refer to some (but certainly not all) weak points. On the other hand, they do not examine the suitability of indicators included, they do not present methods to evaluate the educational processes, the methodology is not adequately compared and they do not answer on the issue of implementation compatibility in a different country that the one in which they have been developed.

There are other researchers that criticize ranking systems, as appears from the table below:
Table II: Criticism on the ranking systems

<table>
<thead>
<tr>
<th>By</th>
<th>Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stuart, 1995</td>
<td>Institutional profile is not taken into account, reputation is used as an indicator of quality, people that evaluate may be prejudiced, publishers think that institutions with strict admission procedures have high quality</td>
</tr>
<tr>
<td>MacGowan, 2000</td>
<td>Validity, statistical correlation between systems / systems and variables</td>
</tr>
<tr>
<td>NORC, 1997</td>
<td>Common acceptance of performance criteria, objectivity of indicators, weights, methodology, suggests the examination of statistical attributes of indicators and need for improvements, suggests ways to develop commonly accepted indicators</td>
</tr>
<tr>
<td>Webster (1992), Webster (2001)</td>
<td>Problems in the use of indicators, methods to alter performance according to objectives Weightings ascribed are different from actual contribution of each indicator to performance due to multi collinearity (USNWR)</td>
</tr>
<tr>
<td>Ehrenberg (2003), Ehrenberg and Monks (1999), Machung, (1998)</td>
<td>Changes in methodology, weights and criteria used A small change in methodology results in better position in the league table which does not represent changes in quality</td>
</tr>
<tr>
<td>Hossler 2000</td>
<td>No information on data collection and calculation of performance, provision of mistaken data by the institution for better performance results</td>
</tr>
<tr>
<td>Dichev 1999</td>
<td>He found out that 10% of fluctuation in performance based on USNWR is ascribed to changes in quality while 90% is ascribed to calculation methods. Insufficient design of systems (Business Week, USNWR)</td>
</tr>
<tr>
<td>Trieschmann 2000</td>
<td>USNWR, Business Week focus only on knowledge utilization and not research (which is not evaluated as it should)</td>
</tr>
<tr>
<td>Gose 1999, Hossler 2000</td>
<td>Incorporation of quality indicators that focus on student learning and engagement in the educational process</td>
</tr>
<tr>
<td>Clarke 2002</td>
<td>The selection of indicators should be based on reliability, validity, comparability and be categorized as input, process, output</td>
</tr>
<tr>
<td>Wallpole 1996</td>
<td>Student satisfaction should be incorporated</td>
</tr>
<tr>
<td>Gioia and Corley 2002, Vaunno 2002</td>
<td>The actual performance improvement based on rankings Few criteria are related to quality, problems in methodology, weights and criteria</td>
</tr>
<tr>
<td>Van Raan 2005</td>
<td>Bibliometric methods to evaluate performance</td>
</tr>
<tr>
<td>Gater 2003</td>
<td>Evaluation should be continuous, reliability, indicators selection, presentation of data</td>
</tr>
<tr>
<td>Lombardi et al 2000, 2001</td>
<td>Reliability issues of indicators due to differences in institutional size, objectives and mission</td>
</tr>
<tr>
<td>Siemens et al 2005</td>
<td>They pose the question if teaching and research should be weighted in a ranking system</td>
</tr>
</tbody>
</table>

Merisotis (2002) suggests that before the development of any system for university performance evaluation, it has to be determined what is the content and the meaning of quality in higher education so that it becomes clear, what it is important and to whom. The comprehensive performance evaluation of a university cannot be based solely on a ranking system. Mainly because as appears from the heavy criticism they attract and their analysis, quality is not sufficiently evaluated. The methodology used by each ranking system will also have to be evaluated under different circumstances. The aggregation of the individual indicators in a single one needs further attention. It is possible that systems which result in institutional categories (like the German system) may be more reliable and useful and provide value added information to all interested parties.

The primary aim of a university performance evaluation system should be institutional improvement through quality assurance in every process and action. Moreover, the provision of performance information to the state and all interested parties should not be underestimated. Ranking systems could supplement the evaluation procedures undertaken by official actors.
Conclusions

It is difficult to develop a universal ranking system that would have the possibility to adjust according the specific conditions of each country and provide reliable and internationally comparable performance data.

Each ranking system should state the objective of rankings and identify precisely the audience it addresses. Institutional profile and mission should be taken into account as well as the context in which a university operates. Results should be adjusted based on the peculiarities of each institution.

The methodology, the collection and analysis of data should be transparent and the selection of indicators must have scientific rudiments, reliability and validity. A ranking system should emphasize on all educational processes (teaching, research, external engagement) and infrastructure and categorize indicators to inputs, processes, outputs and outcomes. The assignment of weights is another issue that needs to be addressed and should result from an extensive analysis of the actual contribution of each indicator in university performance. Rankings systems should be also dynamic, that is, be able to change when circumstances or developments in the higher education sector demand it (for example in the case of interdisciplinary programmes).

Ranking systems, as appears from literature are an inadequate approach to evaluate the performance of an institution. However, if certain changes are incorporated, they can be a useful tool for students and other stakeholders. Only when the above presuppositions are fulfilled can a ranking system become a reliable and complementary tool for university performance evaluation.

REFERENCES


Banta, T., Borden, V., (1994), Performance indicators for accountability and improvement, *New Directions for Institutional Research*, No.82, pp.95-106


Danish Evaluation Institute, (2003), Quality procedures in European higher education, *ENQA*

Di Nauta, P., Omar, P.J., Schade, A., Scheele, J.P., (Eds), (2004), Accreditation models in higher education, *ENQA*

Dill, D and Soo, M.,(2004), Is there a global definition of academic quality? A cross national analysis of university ranking systems, *Public Policy for Academic Quality, Background Paper*

Ehrenberg, R.,(2003), Method or madness? Inside the USNWR college rankings, Paper presented at the Wisconsin Center for the Advancement of Postsecondary Education Forum on the Use and Abuse of College Rankings, ERIC

Epper, R., (1999), Applying benchmarking to higher education, *Change*, pp.24-31


Fuhrman, S.H.,(1999), The new accountability, *CPRE Policy Brief*


Gater, D.,(2003), Using national data in university rankings and comparisons, The Center


Goertz, M.E; Duflly, M.C., (2001), Assessment and accountability across the 50 states, *CPRE Policy Brief*

Gose, B.,(1999), A new survey of “good practices” could be an alternative to rankings, *Chronicle of Higher Education, Vol.49, Issue 9*

Löfström, E., (2002), In search of methods to meet accountability and transparency demands in higher education: experience from benchmarking, Socrates Intensive Programme “Comparative Education Policy Analysis” Lake Bohinj, Slovenia, August pp 21-30
Lombardi, J., Graig., D., Capaldi, E.Gater, D.,(2000), The myth of number one: indicators of research university performance, The Center
Lombardi, J., Graig., D., Capaldi, E.Gater, D.,Mendonca, S.,(2001), Quality engines: the competitive context for research universities, The Center
MacGowan, B.,(2000), Those magazine rankings: lets beg them to stop, Professional School Counselling, Vol.3, Issue 4
Merisotis, J., (2002), Summary report on the invitational roundtable on statistical indicators for the quality assessment of higher / tertiary education institutions: rankings and league table methodologies, Higher Education in Europe, Vol.27, No.4,pp.475-480
Provyn, D., Abercromby, K.,(2000), University league tables and rankings: a critical analysis, CHEMS,
Ridley, D.,Cuevas, M.,Matveev, A.,(2001), Transitions between tiers in US News and World Report rankings of colleges and universities, ERIC
Stuart, D., (1995), Reputational rankings: background and development, New Directions for institutional research, 88, pp.17-19
Thomson, N.,(2000), Cooking the school books (yet again). The US News college rankings gets phonier and phonier, On line
Stuart, D., (1995), Reputational rankings: background and development, New Directions for institutional research, 88, pp.17-19
Van Dyke, N., (2005), Twenty years of university report cards, Higher Education in Europe, Vol.30, No.2,