

Lessons Learned from the Rules on Ranking of Universities [1]*

by

Jean-Jacques DROESBEKE**

KEYWORDS. — Higher education; Evaluation, Quality.

SUMMARY. — It is well known that university assessment and ranking must serve many purposes: they respond to demand for interpretable information on the standing of higher education institutions; they provide some of the rationale for the allocation of funds; they are also used to differentiate between different types of institutions and different programmes and disciplines.

In addition, many public managers would like to use methods based on these tools in order to “measure the quality” of higher education institutions or programs within a particular country or region.

What are these ranking methods? Are they really useful in a process of evaluation? Should we use them for evaluating development research?

We will present here some specific methodologies (Ranking of the Times Higher Education Supplement, the Shanghai ARWU Ranking, the CHE Research Ranking) in trying to reply to the above-mentioned questions. We will also examine a set of principles of quality and good practice in ranking proposed by an International Ranking Expert Group to allow some discussion on these questions.

1. Introduction

Rankings appear in many areas. Competitive sports (football, basket ball...), cinema (Oscar Academy Awards...), music competitions (Queen Elisabeth competition...), gastronomic guides (Michelin, Delta...) are some well known examples. Sometimes, rankings are put together just for fun. However, many of them are taken very seriously.

In the field of higher education, ranking is closely linked to the topic of evaluation which is traditionally carried out through procedures of peer review of scientific publications. However, academic evaluations are increasingly being undertaken in addition to this traditional procedure.

Two types of rankings are regularly quoted in the media. The first one is that of the Shanghai Jiao Tong University — called the *Shanghai* ranking — and the second one is the yearly ranking published in the *Times Higher Education Supplement* — called the *Times* ranking.

But let us begin by considering the basic question: “What is a ranking?”

Ranking is *positioning comparable objects on an ordinal scale based on a (non strict) weak order relation among (statistical) function of, or a combination of functions of measures or scores associated with those objects* (Glänzel & Debackere, 2009). These functions, which are usually based on variables for evaluative purposes, are called indicators. Different indicators X_k representing different aspects of quality, form the components of a *composite*

* The numbers in brackets [] refer to the notes pp. 000.

** Member of the Royal Academy for Overseas Sciences, Prof. Université libre de Bruxelles (LMTD, Solvay Brussels School of Economics and Management), CP 139, av. F.D. Roosevelt, 50, B-1050 Brussels (Belgium).

indicator Y , the basis of the ranking. This composite indicator is usually a linear combination of the X_k s, that is,

$$Y = \sum_k \lambda_k X_k$$

where λ_k ($k=1, 2, \dots, p$) are p pre-defined weights and, without loss of generality, verify the equality $\sum \lambda_k = 1$ — so that Y is a weighted mean of the individual indicators X_k .

This method has been used for a long time and the most problematic issues in applying composite indicators are well known (GLANZEL & DEBACKERE 2009): possible interdependence of components, altering weights resulting in a different ranking, potential obscure and irreproducible results, ignored random errors of statistical functions, reduction of a multi-dimensional space into linearity. Moreover, the interpretation of a ranking is linked to the criteria which are used.

Let us now consider two examples of rankings. We will complete this presentation by an interesting third method.

1.1. THE SHANGHAI RANKING

The Composite Indicator currently aggregates 6 Criteria.

- The first criterion used to measure the quality of education of a university is the number of alumni who were awarded a Nobel Prize (physics, chemistry, medicine, economics) or a Fields Medal (mathematics). An alumni is defined as a person who holds at least one degree from the university being assessed; such a person counts as one unit if the degree was obtained after 1990, as 0.9 if it was obtained between 1980 and 1990, etc.; it counts as 0.1 if the degree was obtained between 1900 and 1910.
- Two criteria measure the quality of the teaching staff of the university: the number of Nobel Prizes and Fields Medals awarded to the Academic staff teaching at the university under review — here too, the weights decrease with seniority and a complex fraction system weights the results if the laureate was active in several universities simultaneously and/or if several laureates shared the same prize — and the number of teachers - researchers who are among the 250 most cited authors for a given period of time within 21 major subject categories.
- The fourth criterion is the number of articles from the university under review published in Nature and Science within the last 5 years, with a weight system for co-authored publications — this criterion is not taken into account if the university is not active in these fields and its weight is then redistributed between the other criteria.
- The fifth criterion is the number of articles from the university under review listed, for a given period, in the *Science Citation Index* and the *Social Science Citation Index* [2].
- A sixth criterion concerns the academic performance with respect to the size of the university under review: the scores for the previous criteria are summed up and divided by the number of teachers – researchers [3].

For each of these criteria the highest-ranked university is given a score of 100; the others are given a score obtained by a rule of 3.

The global score of a university is a weighted sum [4] of the scores obtained for the different criteria.

The reliability of this composite indicator is a real problem which has been discussed in many conferences and publications (see for instance VINCKE 2009).

1.2. THE *TIMES* RANKING

The *Times Higher Education Supplement* ranking is published by a private firm. The methodology is based upon two surveys. The first one concerns several hundred scientists from different countries who list the universities they consider to be the best; the second one is conducted on a number of employers. These two surveys allow extracting a first score for each university which will receive a weight of 50 % in the final indicator obtained — here also — as a weighted sum of scores for different criteria.

Four other criteria are taken into account; the impact, in terms of citations, of university researchers (with a weight of 20 %), the student/teacher ratio (with a weight of 20 %), the percentage of foreign students (with a weight of 5 %) and the percentage of foreign teachers (with a weight of 5 %).

Until 2006, the standardisation to 100 of each scale followed the Shanghai method. But in 2007, the authors of this method decided to replace the normalization to the best performers by the “z-score” method”. For each criterion, the empirical mean and the empirical standard deviation are computed. For each university, a standardized score is obtained by taking the difference between the score and the mean divided by the associated standard deviation [5].

Here also a number of comments can be made about the *Times* ranking and many questions have not received satisfactory replies. As the first criterion is based on recommendations formulated by “experts”, this method seems to be closer to the well known “peer review” than the Shanghai ranking. Unfortunately it is not really the case.

As this meeting also concerns the bibliometric approach, it is significant to note that in 2005, Anthony Van Raan of the University of Leiden calculated that the correlation between the scientists’ replies to the questions of the *Times* and a bibliometric analysis was equivalent to...zero (VAN RAAN 2006)!

1.3. THE CHE APPROACH

The *Centre for Higher Education Development* (CHE) was founded in 1994 by the German Rectors’ Conference and the Bertelsmann Foundation. A *CHE University Ranking* has been published since 1999 and focuses on information for prospective students who want to choose a university. It includes indicators on teaching and learning, resources and facilities, on research activities as well as information that is important for this target group.

But research is not at the centre of this ranking. This is why the CHE decided to set up a *Research Ranking* devoted to an academic target group. It includes indicators on publications, citations, number of PhDs, research grants and patents and also some bivariate and correlation analysis (BERGHOFF & FEDERKEIL 2009) [6].

Three big differences can be underlined between the CHE rankings and the previous ones. At the level of ranking, programmes and fields are considered instead of whole institutions. On the other hand, the method takes into account a multi-dimensional ranking instead of a composite overall score. Finally for each indicator, the CHE ranking classifies universities into only three groups; a top, a middle and a bottom group, instead of providing a league table.

For all of these reasons, it appears that the CHE method seems better than the two previous ones.

2. The International Ranking Expert Group

We have seen that during recent years, a number of widely covered worldwide rankings of universities have been appearing. If correctly understood and interpreted, they could contribute to national accountabilities and quality assurance processes. Given this trend, it is important that the produced rankings hold themselves accountable for quality in their own data collection, methodology, and dissemination.

In view of the above, the *International Ranking Expert Group* (IREG) was founded in 2004 by the UNESCO European Centre for Higher Education (UNESCO-CEPES) in Bucharest and the Institute for Higher Education Policy in Washington, DC. In 2006, IERG's second meeting in Berlin proposed a set of principles of quality and good practice in HEI rankings. Let us detail these recommendations [7].

2.1. RECOMMENDATIONS ABOUT PURPOSES AND GOALS

- *Be one of a number of diverse approaches to the assessment of higher education inputs, processes, and outputs.* Ranking can provide comparative information and improved understanding of higher education, but should not be the main method of assessing what high education is and does. Rankings provide a market-based perspective that can complement the work of government, accrediting authorities, and independent review agencies.
- *Be clear about their purpose and their target groups.* Rankings have to be designed with due regard to their purpose. Indicators designed to meet a particular objective or too inform one target group may not be adequate for different purposes or target groups.
- *Recognize the diversity of institutions and take the different missions and goals of institutions into account.* Quality measures for research-oriented institutions, for example, are quite different from those that are appropriate for institutions that provide broad access to underserved communities. Institutions that are being ranked and the experts that inform the ranking process should be consulted often.
- *Provide clarity about the range of information sources for rankings and the messages each source generates.* The relevance of ranking results depends on the audiences receiving the information and the sources of that information (such as databases, students, professors, employers). Good practice would be to combine the different perspectives provided by those sources in order to get a more complete view of each higher education institution included in the ranking.
- *Specify the linguistic, cultural, economic, and historical contexts of the educational systems being ranked.* International rankings in particular should be aware of possible biases and be precise about their objective. Not all nations or systems share the same values and beliefs about what constitutes “quality” in tertiary institutions, and ranking systems should not be devised to force such comparisons.

2.2. RECOMMENDATIONS ABOUT DESIGN AND WEIGHTING OF INDICATORS

- *Be transparent regarding the methodology used for creating the rankings.* The choice of methods used to prepare rankings should be clear and unambiguous. This transparency should include the calculation of indicators as well as the origin of data.
- *Choose indicators according to their relevance and validity.* The choice of data should be grounded in recognition of the ability of each measure to represent quality and academic and institutional strengths, and not availability of data. Be clear about why measures were included and what they are meant to represent.
- *Measure outcomes in preference to inputs whenever possible.* Data on inputs are relevant as they reflect the general condition of a given establishment and are more frequently available. Measures of outcomes provide a more accurate assessment of the standing and/or quality of a given institution or program, and compilers of rankings should ensure that an appropriate balance is achieved.
- *Make the weights assigned to different indicators (if used) prominent and limit changes to them.* Changes in weights make it difficult for consumers to discern whether an institution's or program's status changed in the rankings due to an inherent difference or due to a methodological change.

2.3. RECOMMENDATIONS ABOUT COLLECTION AND PROCESSING OF DATA

- *Pay due attention to ethical standards and the good practice recommendations articulated in these Principles.* In order to assure the credibility of each ranking, those responsible for collecting and using data and undertaking on-site visits should be as objective and impartial as possible.
- *Use audited and verifiable data whenever possible.* Such data have several advantages, including the fact that they have been accepted by institutions and that they are comparable and compatible across institutions.
- *Include data that are collected with proper procedures for scientific data collection.* Data collected from an unrepresentative or skewed subset of students, faculty, or other parties may not accurately represent an institution or programme and should be excluded.
- *Apply measures of quality assurance to ranking processes themselves.* These processes should take note of the expertise that is being applied to evaluate institutions and use this knowledge to evaluate the ranking itself. Rankings should be learning systems continuously utilizing this expertise to develop methodology.
- *Apply organizational measures that enhance the credibility of rankings.* These measures could include advisory or even supervisory bodies, preferably with some international participation.

2.4. RECOMMENDATIONS ABOUT PRESENTATION OF RANKING RESULTS

- *Provide consumers with a clear understanding of all of the factors used to develop a ranking, and offer them a choice in how rankings are displayed.* This way, the users of rankings would have a better understanding of the indicators that are used to rank institutions or programs. In addition, they should have some opportunity to make their own decisions about how these indicators should be weighted.
- *Be compiled in a way that eliminates or reduces errors in original data, and be organized and published in a way that errors and faults can be corrected.* Institutions and the public should be informed about errors that have occurred.

3. Comments

Taking into account these recommendations, one cannot ignore the fact that most of the current ranking proposals have serious flaws: bias toward “big” universities, favouring institutions from the Anglo-Saxon world, favouring universities which excel in exact and biomedical sciences (rather than in human and social sciences), being too strongly influenced by the values of the weights...Moreover ranking should take into account the diversity of tasks of universities as well as diversities across disciplines (Dehon *et al.* b 2009).

It is not surprising that most university authorities claim that rankings published by the media are disputable. Everybody is convinced that these rankings are not representative of “the true” quality of universities. But it is a fact that all universities integrate ranking criteria in their governances nowadays.

Recent initiatives allow hoping for some improvement in the methodology. In 2006, for instance, the *Programme for Institutional Management in Higher Education* (IMHE) of the Organisation for Economic Co-operation and Development (OECD) and the *International Association of Universities* (IAU) proposed to carry out a study of the positive and negative effects of rankings on strategic and administrative decision making by universities (Hazelkorn, 2008). The OECD has now started an international assessment programme to evaluate the outcomes of higher education policies (*Assessment of Higher Education Learning Outcomes*: AHELO). The aim of this study is to verify whether it is possible to test university performance levels using criteria which are independent of linguistic, cultural and geographic factors.

It is clear that the flaws of the current ranking methods do not allow to use them for evaluating development research. But here also it should be useful to take the different criteria into account qualitatively. The proposals of the German CHE and the scores obtained on each criterion — even if the values of these scores do not have an absolute meaning — may allow seeing the real level of a university, a faculty, a department and even a research centre.

The recommendations proposed by the *International Ranking Expert Group* can certainly help to improve the quality of a research activity, of an education project or of a service to the society.

ACKNOWLEDGEMENT

We thank Marjorie Gassner for her judicious remarks.

NOTES

- [1] This talk results from an international conference organised by Catherine Dehon, Mathias Dewatripont, Jean-Jacques Dreesbeke, Dirk Jacobs and Catherine Vermandele in December 2007 at the Université libre de Bruxelles on the topic of *Ranking and Research Assessment in Higher Education* (DEHON *et al.* a 2009). The conference was organized within the framework of a European PhD programme in Socio-Economic and Statistical Studies regrouping a set of European universities.
- [2] In the first version of the Shanghai ranking, only these 5 criteria were taken into account (reduced to 4 for social science institutions).
- [3] As the authors of this ranking did not have data available for all the universities, this additional criterion was not taken into account for all universities.
- [4] A constant weight of 25 % when there are 4 criteria, 20 % when there are 5; with 6 criteria, the first one and the sixth are worth 10 % each while the other four are each worth 20 %.
- [5] The z-score indicates how far the university deviates from the mean using the standard deviation as the unit.

- [6] See also: www.che.de/downloads/Methoden_Hochschulranking_2008_AP106.pdf.
[7] See www.che.de/downloads/Berlin_Principles_IREG_534.pdf.

REFERENCES

- BERGHOFF, S. & FEDERKEIL, G. 2009. The CHE approach, in *Ranking universities*, — DEHON *et al.* (Eds.): 41-63.
- Centre for Higher Education Development, 2006. Berlin Principles on Ranking of Higher Education Institutions. — www.che.de/downloads/Berlin_Principles_IREG_534.pdf
- Centre for Higher Education Development, 2008. Vorgehensweise und Indikatoren. — www.che.de/downloads/Methoden_Hochschulranking_2008_AP106.pdf.
- DEHON, C., JACOBS, D. & VERMANDELE, C. (Eds.) 2009(a). *Ranking universities*, Bruxelles, Editions de l'Université de Bruxelles.
- DEHON, C., JACOBS, D. & VERMANDELE, C. 2009 (b). Ranking and research assessment in higher education: current and future challenges. — *In: DEHON et al.* (Eds.), *Ranking universities*, 1-10.
- GLANZEL, W. & DEBACKERE, K. 2009. On the “multi-dimensionality” of ranking and the role of bibliometrics in university assessment, in *Ranking universities*, DEHON *et al.* (Eds.), 65-75.
- HAZELKORN, E. 2008. Learning to live with league tables and ranking: the experience of institutional leaders. — *Higher Education Policy*, **21**: 193-215.
- VAN RAAN, A. F. J. 2006. Performance-related differences of bibliometric statistical properties of research groups: Cumulative advantages and hierarchically layered networks. — *Journal of the American Society for Information Science and Technology*, **57**, 14: 1919-1935.
- VINCKE, Ph. 2009. University rankings. — *In: DEHON et al.* (Eds.), *Ranking universities*, 11-25.