Evaluating university research: Same performance indicator, different rankings

Giovanni Abramo [a,*], Ciriaco Andrea D'Angelo [b,1]

[a] Laboratory for Studies of Research and Technology Transfer at the Institute for System Analysis and Computer Science (IASI-CNR), National Research Council of Italy, Roma, Italy
[b] University of Rome "Tor Vergata"—Italy and Laboratory for Studies of Research and Technology Transfer (IASI-CNR), Rome, Italy

ARTICLE INFO

ABSTRACT

Assessing the research performance of multi-disciplinary institutions, where scientists belong to many fields, requires that the evaluators plan how to aggregate the performance measures of the various fields. Two methods of aggregation are possible. These are based on: (a) the performance of the individual scientists or (b) the performance of the scientific fields present in the institution. The appropriate choice depends on the evaluation context and the objectives for the particular measure. The two methods bring about differences in both the performance scores and rankings. We quantify these differences through observation of the 2008–2012 scientific production of the entire research staff employed in the hard sciences in Italian universities (over 35,000 professors). Evaluators preparing an exercise must comprehend the differences illustrated, in order to correctly select the methodologies that will achieve the evaluation objectives.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The Humboldtian philosophy of "education through research" has forged the higher education systems of many nations. Although questioned in the past, in the face of the massification of higher education, with the growing of more vocationally oriented higher education institutions (Ash, 1999), in the current knowledge-based economy the Humboldtian legacy has regained new attention among policy makers, in terms of the "research–teaching nexus". According to a European report (Commission of the European Communities, 2004), competency-ased higher education that is focused on employability in the knowledge society is in need of "education through research". This is because research competencies are useful for professionals in a knowledge society, and because higher education is only able to deliver these competencies if its education is related to research.

The evaluation of university research performance then is becoming ever more common in many nations. The issue of evaluation, as well as the university rankings that are readily available, now attracts the attention of the popular media and a vast and varied public. The *SCImago Institutions Rankings* (SCIMAGO, 2014) and the *CWTS Leiden Rankings* (CWTS, 2014) are the most read and accredited examples of the world rankings carried out by bibliometricians. Other yearly world

* Corresponding author. Present address: Istituto di Analisi dei Sistemi e Informatica, Consiglio Nazionale delle Ricerche, Via dei Taurini 19, 00185 Roma, Italy. Tel.: +39 06 7716417; fax: +39 06 7716461.
E-mail addresses: giovanni.abramo@uniroma2.it (G. Abramo), dangelo@dii.uniroma2.it (C.A. D'Angelo).
[1] Present address: Dipartimento di Ingegneria dell'Impresa, Università degli Studi di Roma "Tor Vergata", Via del Politecnico 1, 00133 Roma, Italy. Tel.: +39 06 72597362; fax: +39 06 72597362.

university rankings attract much greater media and public attention (THE—Times Higher Education, 2014; QS—Quacquarelli Symonds, 2014; SJTU—Shanghai Jiao Tong University, 2014), however most bibliometricians agree in strongly criticizing their methodological weaknesses (Billaut, Bouyssou, & Vincke, 2010; Dehon, McCathie, & Verardi, 2010; Sauder & Espeland, 2009; Liu & Cheng, 2005; van Raan, 2005).

A growing number of countries now also conduct their own periodical evaluation exercises of research quality in the national university systems. The objectives are multiple. In many nations the results of the evaluations serve in the allocation of public resources (Hicks, 2009, 2012; OECD, 2010). The evaluations in fact can have a significant impact on the individual and collective behavior of the actors in the research system (Vanecek, 2014; Himanen, Auranen, Puuska, & Nieminen, 2009; Smart, 2009). Also, where the rankings of the university organizational levels are made public there is a reduction in information asymmetry between the suppliers and the seekers of "new" knowledge, with gains for the efficiency of markets in knowledge and education. Further, students can make informed choices in selecting the institutions for their studies. Private companies can efficiently select partners for joint research, as well as recruit new personnel on the basis of the performance of the universities that provided the candidate training. Universities themselves want to know the strengths and weaknesses of their own organizational units, for purposes of strategic planning. Given such varying motivations, every stakeholder would clearly adopt a unique evaluation perspective, assigning different weights to the dimensions of a performance evaluation as these are linked to the objectives of their concern.

For the state, the typical rationale in allocating more or less state funds to the differently performing universities is to maximize the rate of return on research spending, in the form of yields in scientific and technological advancement. Also, in the light of the above noted research-teaching nexus, increasing the funds to universities that are better in research should translate into overall improvement in the educational offer, and so in the quality of the future labor force. On the basis of the published rankings, students can make informed choices about where to apply for their education. Universities are stimulated to improve performance and rise in the rankings, to attract the best students. The overall process can be one of a virtuous circle, leading to broader economic and social progress.

The objective of this work is to unveil the ambivalence inherent in constructing university research performance ranking lists commissioned by government for efficient funding selection. There are in fact two possible perspectives in facing the problem of maximizing the rate of return on research spending. One perspective is more research-oriented, the other more teaching-oriented. The two perspectives imply different approaches in constructing ranking lists, which presumably lead to different results.

To better exemplify our thinking, consider two universities A and B, identical in size, in fields of research and degree programs, but different in research performance, which is greater in A. Then suppose that the better performance of A is essentially due to 10% of its research staff, which is exceptionally good, even though the remaining 90% have below-median performance in their research fields. Even though all the professors of B have research performance above median, it ranks below A. Now we put ourselves in the shoes of the policy maker that wants to maximize the returns on financing for research. He or she sees these universities as black boxes, a bit like the investor choosing between different stock portfolios. Under parity of conditions, the investor will choose the portfolio offering maximum return, independent of the performance of the individual stocks. Form the research-oriented perspective, the policy maker is also interested in overall returns, rather than the distribution of performance by the individual scientists. The optimal choice will be to allocate resources to university A. From the teaching-oriented perspective, the optimizing choices are not so obvious: is it better to have 10% of classes taught by the "greats" in the subject, and 90% by mediocre professors, or is it better to have fair to good professors for every field of study? The second option could be safer, for a good overall education. Thus from a teaching-oriented perspective, a more penetrating type of ranking seems appropriate and the optimal choice would be to allocate resources to university B.

The extreme institutional characteristics in the above illustration could seem unrealistic, particularly for strongly competitive higher-education systems. However, they could be quite close to reality for systems where competition is largely lacking, as in a number of European nations (Auranen & Nieminen, 2010). For example, a study by Abramo, Cicero, and D'Angelo (2012a) has shown that in the Italian case there is a huge dispersion of research performance within the individual universities, compared to what is seen between them. The methods and indicators of evaluation must take account of such potential realities, and be conceived to achieve the objectives of the given evaluation exercise, whether it be for the use of students, in prioritizing funding, or other.

These considerations lead us to examine two of the different methods for measuring the universities' research performance: one that controls for the impact of outliers on the aggregate performance of the institution, and another that does not provide such control. The former would seem to serve better to inform students in the selection of their universities; the second probably results as better to inform decisions aimed at maximizing overall scientific advancements. To rank universities in research performance for a particular field the bibliometrician can proceed either way. The first, based on the performance of individual professors, interprets the performance of the organizational unit as the average of the individual performances, meaning that the emphasis is on the individual. The other method interprets the field as a black box. It normalizes the output of the all the scientists in a field by the labor input, meaning that emphasis is on the overall product of the scientists in that field, independent of the variability of the individual contributions. The two methods, equally legitimate in an operational sense, give rise to performance scores and rankings that are correlated, but still different. The appropriate choice of method then depends on the aims of the evaluation.