

Rankings: How Should Universities Respond?
Second Leiden University International Symposium on Ranking

Leiden University
Kamerlingh Onnes Building
Friday February 2, 2007

Opening Address by

Professor Douwe Breimer
Rector Magnificus & President of Leiden University

Ladies and Gentlemen,

Welcome to Leiden University and the city of Leiden. We have about 200 participants from more than 30 countries, within Europe including Russia, Ukraine, Lithuania and Estonia, and from the United States, Mexico, Taiwan, Hong Kong, Japan, Nigeria, South Africa, and Australia.

Our University is the oldest of the Netherlands -432 years old- and during all these centuries up to now this university has been one of Europe's major centers of outstanding research and teaching. This building, where we are today, is one of the world's Holy Grounds of science. It is a place of discoveries that changed the world. Of the eleven Nobel Prize winners affiliated to Leiden University, eight were student, did their PhD-work, or have been professor in this building.

In the last few years rankings of universities, though controversial, have become increasingly popular. Two rankings widely attracted the attention of policy makers, the scientific world and the public media: the rankings published by the Jiao Tong University in Shanghai from 2003, and the rankings published by Times Higher Education Supplement from 2004.

Rankings suggest a similar simplicity for the evaluation of scientific performance as in the case of a football league. The immediate observation that the well-known US top-universities take the lead, reinforces these suggestions. Universities respond enthusiastically if they feel their position is worth publishing. Although things are not so simple and the various methodologies used in these ranking still have to be discussed thoroughly, the influence of these rankings is striking. Rankings have become unavoidable; they are now part of academic life. General ideas about international reputation of universities will be influenced considerably by rankings, high ranking universities will 'advertise' their position on the rankings as final truths about their research quality, and so these ranking may considerably guide the choice of young scientists and research students. Ranking lists have changed the worldwide academic landscape.

In Germany, the Centre for the Development of Universities (Centrum für Hochschulentwicklung, CHE) applies a broad range of criteria to judge the standing of German universities, with a focus on specific disciplines, not only research, but also teaching. CHE does not provide an 'overall ranking', but applies a more

differentiated approach: which universities are in the top for chemistry, for engineering, and so on. More and more 'ranking producers' follow this field-specific differentiation.

Rankings have strong 'de-equalizing' effects. They directly confront scientists, university board members, policy-makers, politicians, journalists, and the interested man-in-the-street with inequality. Rankings strengthen the idea of an academic elite, and institutions use the outcomes of rankings, no matter how large the methodological problems are, in their rivalry with other institutions.

Rankings reinforce the 'evaluation culture'. Evaluations are necessary to account for the received financial support by society. Academic researchers cannot withdraw anymore from these responsibilities as in earlier times, simply because nowadays the academic enterprise has grown to such an extent that it consumes a considerable part of the public means. Moreover, governments and national research councils are more and more inclined to use the outcomes of evaluations for the distribution of finances over the institutions, and university boards are willing to use the same outcomes for re-distribution of money allocation within their institutions.

There is an interesting difference between the concepts of 'reputation' and 'contemporaneous performance'. On the one hand, any serious peer review committee responsible for the evaluation of university research would not even think of using long-time-ago Nobel Prize Winners for the judgment of the (nowadays) quality of a university. If they would do, they would disqualify themselves as evaluators. But on the other hand, the history of a university with 'grand old men' does count considerably in reputation. Universities publish in their websites honorary lists of Nobel Prize Winners (if they have) and other famous scientists. Thus, the historical tradition of scientific strengths of particularly the older, classic universities is a strong asset in their present-day reputation.

The major problem is: what do we want to measure, and how does 'established reputation' relate to 'contemporaneous performance'. For the moment we leave this question to further research, but we remark that established reputation is not necessarily the same as 'past glory'. Often we see that institutions with an established reputation are remarkably strong in maintaining their position. They simply have more power to attract the best people, and this mechanism provides these renowned institutions with a cumulative advantage to further reinforce their research performance.

Judgment by knowledgeable colleague-scientists, known as peer review, is the principal procedure of assessing research performance, notwithstanding its shortcomings and disadvantages. In most cases, peer review is applied on a relatively small scale, ranging from the review of a submitted paper or a research proposal by two or three referees, the review of the record of candidates for a professorship by, say, five experts in the field, and the assessment of research groups and research programs within a specific discipline by between five and ten peers.

This implies two important things. First, peers can be regarded as experts with respect to the quality of the object. Second, the object to be evaluated has a 'size' that is comparable with the usual working environment of the peer, namely a research group or a research program and, thus, surveyable for individual peer

judgment. In rankings, however, scientists have to judge much larger entities, even complete universities, and so the 'cognitive distance' to the object to be evaluated increases considerably. Therefore, it is questionable whether all the individual academics involved in such large-scale surveys can be regarded as knowledgeable experts in all those parts of the evaluated entities that really matter. In such cases, 'experts' will tend to judge on the more general basis of established reputation, instead of their own actual knowledge (if they have!) of recent past performance.

This awareness, however, is precisely what a peer must have. It is also this recognition of recent past performance that forms the strength of bibliometric analysis. Indeed, bibliometric indicators can be seen as the aggregate of typical peer review. Well-informed colleague-scientists play their role as a member of an 'invisible peer review college' by referring in their own work to earlier work of other scientists. And as this happens for all publications of a university in many disciplines, the outcomes of a bibliometric analysis on the level of a university will be statistically very significant.

Dependence of the outcomes on the choice of experts is one of the major problems. This dependence may cause biases in fields of expertise. After publication of the first THES ranking, The Sydney Morning Herald gloried the presence of not less than six Australian universities in the top 50, while the German news agency DPA bemoaned the absence of the country in the top positions. Six universities in the top-50 universities count for 12%, whereas Australia contributes for about 2.5 % to the worldwide scientific impact. Germany contributes around 8 % to worldwide scientific impact. Most of these Australian 'top' universities score low to very low in citations. These strange discrepancies between the results of an expert survey and bibliometric findings suggest biases.

Bibliometric assessment of research performance is based on one central assumption: scientists, who have to say something important, do publish their findings vigorously in the open, international journal literature. This assumption introduces unavoidably a 'bibliometrically limited view of a complex reality'. Journal articles are not in all fields the main carrier of scientific knowledge. However, the daily practice of scientific research shows that inspired scientists in most cases and particularly in the natural sciences and medical research fields, 'go' for publication in the better and -if possible- the best journals.

But this is less the case in engineering research, social and behavioral sciences, and certainly for the humanities. Thus, the strength of a university in engineering, in the social and the behavioral sciences or in the humanities may contribute little -or even hardly- to the position of that university in a ranking based on bibliometric data. Smaller universities, and particularly those with an emphasis on social sciences and humanities, will have a better chance by the peer review element in the THES ranking as compared to the more bibliometrically-oriented Shanghai study. A striking example is the difference in position of the London School of Economics, a top-position in the THES ranking and a low position in the Shanghai ranking.

The increasing use of bibliometric data in evaluation procedures and particularly in rankings underlines the vital importance of a clear, coherent and effective presentation to the outside world of universities in their publications. For instance, King's College, University of London (KCL), introduced a code of practice to ensure that all publications are properly attributed to the College. This is in light of recent evidence that up to 25% of citations from KCL academics in recent years were

missed due to failure to use 'King's College London' in the address mentioned in the publication heading.

We observe a quite surprising phenomenon: it appears clearly that the group of outstanding universities will not be much larger than around 200 members. Most of the top-universities are large, broad research universities. They have attracted on the basis of their reputation for already a long time the best students and scientists. These universities are the 'natural attractors' in the world of science, and, apparently, around 200 of these institutions are able to acquire and hold on the vast majority of top-scientists. After ranking position 200 or so, there will certainly be smaller universities with excellent research in specific field of science. There is, however, not much room anymore for further 'power houses of science' because no more excellent scientists are available on this planet.

A more concrete bias is related to publication language. Recent work shows that the utmost care must be taken in interpreting bibliometric data in a comparative evaluation of national research systems. The measured impact value depends upon whether one includes or excludes publications written in languages other than English, particularly French and German. Generally the impact of non-English publications is very low. Thus, these publications count on the output side, but they contribute very little, if any, on the impact side. Therefore, such non-English publications considerably 'dilute' the measured impact of a university. These findings clearly illustrate again that indicators need to be interpreted against the background of their inherent limitations such as, in this case, effects of publication language.

On the basis of the same data and the same technical and methodological starting points, still different types of impact-indicators can be constructed, for instance one focusing entirely on impact, and another in which also scale (size of the institution) is taken in to account. Rankings based on these different indicators are not the same, although they originate from exactly the same data.

Re-iteration of the expert-survey with renowned scientists from the first-round 200 top-universities is a first, necessary step in the direction of judgment by highly qualified reviewers. It would be very interesting to see what differences will emerge from such a first-iteration expert round as compared to the original ranking and, by that, how robust the first ranking is. We expect that already in the first re-iteration the expert judgments will converge to the bibliometric measures and that after, say, two re-iterations, there will be a high correlation between expert-based and bibliometric outcomes.

In the political debates on evaluations and rankings often the basic question is: what are the characteristics of successful or, 'world-class' universities? There must be 'sufficiently enough' quality, but what does that mean, and how can it be measured? Successful universities are those universities that succeed to perform significantly above the international average in more than half of their first, say, 20 largest fields.

This symposium addresses pitfalls and challenges in the worldwide ranking of universities. It provides you with a review of recent developments, including new rankings and changes within the rankings already in place. It also presents discussions on the application of rankings by and their impact on universities and governments, particularly their strategies to improve research performance.

There cannot be discussion on the general statement that performance *must* be reflected in observable evidence. But we need to discuss the conditions under which it is possible to evaluate in a reliable and valid way the research strengths of universities with highly automated procedures, within a reasonable time perspective and with reasonable costs. Here the 'Berlin principles' are important guidelines for the construction and the use of rankings. These Berlin principles refer to important issues as for instance recognition of the diversity of universities, transparency of the methodology, and taking input data into account.

Leiden, as a founding member of the European League of Research Universities, has taken the initiative to organize on an annual or bi-annual basis a two-day conference devoted to present and discuss the latest developments in university quality assessment and ranking. This symposium is the second of this successful initiative.

I wish you an exciting symposium!