2009-09-22

Pros and Cons of Research Assessment: Lessons from Rankings

Ellen Hazelkorn
Dublin Institute of Technology, ellen.hazelkorn@dit.ie

Directorate of Research and Enterprise

Reports

Dublin Institute of Technology *Year* 2009

Pros and Cons of Research Assessment: Lessons From Rankings.

Ellen Hazelkorn
Dublin Institute of Technology, ellen.hazelkorn@dit.ie

— Use Licence —

Attribution-NonCommercial-ShareAlike 1.0

You are free:

- to copy, distribute, display, and perform the work
- to make derivative works

Under the following conditions:

- Attribution.
  You must give the original author credit.
- Non-Commercial.
  You may not use this work for commercial purposes.
- Share Alike.
  If you alter, transform, or build upon this work, you may distribute the resulting work only under a license identical to this one.

For any reuse or distribution, you must make clear to others the license terms of this work. Any of these conditions can be waived if you get permission from the author.

Your fair use and other rights are in no way affected by the above.

Pros and Cons of Research Assessment: Lessons from Rankings

World Social Science Report

Ellen Hazelkorn
Director of Research and Enterprise, and Dean of the Graduate Research School
Director, Higher Education Policy Research Unit (HEPRU)
Dublin Institute of Technology
http://www.cser.ie/hepru.htm
http://www.oecd.org/edu/imhe/rankings

Why Assess Research?

The assessment of university-based research is now an integral part of the economic and the research environment. This is linked to higher education's central role as economic driver and to public demand for greater accountability and return-on-investment – especially in the current economic context. While research assessment has been conducted at the national level for quite some time, cross-national or global comparisons are more recent and more influential. The emergence of global rankings, whose purpose is to measure 'world-class excellence', has drawn attention to a country's capacity to participate in world science and hence the global economy. In many instances, there has been a gap between national perception and ranking position.

These factors have contributed to the growing importance and sense of urgency now attached to research assessment by the EU, national governments and funding agencies, research organisations and higher education institutions (HEIs), academics, industry and investors. Although a growing amount of research is being conducted outside academic institutions, by NGOs, think tanks and consultancy firms, the focus is on publicly-funded research.

In response, a ranking and research assessment 'industry' has effectively developed. A variety of organisations and groups have been established to examine, propose and discuss the most appropriate indicators for particular sets of disciplines, for teaching and learning, and for third-mission activities. The EU has established an Expect Group on the Assessment of University-based Research, and a consortium to develop a Multi-dimensional European Ranking System in response to the popularity of the Shanghai Jiao Tong *Academic Ranking of World Universities*.

Rankings and research assessment now form a permanent and necessary part of higher education and publicly-funded research. Research assessment is an important mechanism, at both the national and institutional level, in order to boost research performance and quality, to optimize resource allocation, to differentiate mission or institutional profiling, to facilitate international benchmarking and to identify peers for networking and strategic alliances. It also serves as a tool to increase public awareness and understanding and hence participation in broader discussions about higher education (IHEP, 2009: 1-2). Because research assessment requires improved data collection, it can be beneficial for strategic planning and management, and institutional autonomy.

International evidence shows that ranking and assessment processes can have perverse effects especially when indicators are considered in isolation and simple correlations are made. The evidence also shows that a number of governments, HEIs and researchers are

making decisions and realigning their priorities in order to match indicators. This includes over-concentrating research in a few (elite) HEIs, focusing on particular disciplines (primarily the sciences), and neglecting local or regional issues in order to publish in high impact international journals. Throughout the world, governments and HEIs have rewritten strategies and priorities, and have made significant changes at both the system and institutional level in order to improve their position in global rankings (Hazelkorn, 2008).

As indicators are not value-free, the chosen methodology and interpretation of the results can have significant implications and carry numerous risks. Throughout the following pages, we shall discuss the limitations of some frequently used indicators, and offer some possible alternatives for a 'good practice' model.

Limitations and Unintended Consequences

Research assessment and rankings share a number of characteristics. They both seek to benchmark higher education performance on the basis of selected and, sometimes weighted, indicators. Rankings rely heavily on traditional research outputs captured in international bibliometric and citation databases such as Thompson-Reuters *Web of Science* or Elsevier-Scopus. The scores are aggregated into a final descending rank. Rankings are essentially one-dimensional, since each indicator is considered as independent from the others. Their popularity is largely related to their simplicity; like restaurants, televisions or hotels, rankings provide, at least in appearance, an easy guide to quality.

In contrast, research assessment is often a multifaceted review of performance, conducted by public agencies, using qualitative and quantitative indicators. The UK's RAE offers a good example of this. Organised every 5 years since 1986, it is based on institutional submissions in subject areas or units of assessment which are ranked by a panel of subject specialist peer reviewers. The results determine the level of resource allocation. This is in sharp contrast to other systems that focus mainly on quality assurance (such as in the Netherlands). In recent years, concern about the financial cost, the human resources and time needed, the level of bureaucracy and allegations of 'gaming' have led to the adoption of a more metrics- or indicator-based system. Like the UK, Australia has abandoned its Research Quality Framework (RQF) in favour of the Excellence in Research for Australia Initiative (ERA).

While the results of research assessment are rarely ordered in a hierarchical manner, publication of their results by the media or other organisations has often led to the production of a 'league table' of HEIs. This practice has facilitated the restructuring of the higher education system, and has arguably led to a growing convergence between assessment and rankings.

Bibliometric and citation databases seek to identify the core literature by selecting journals which publish the overwhelming majority of peer-reviewed articles (around 9000 in *Web of Science* and 18,000 in Scopus). While there are efforts to extend coverage to the arts, humanities as well as social science journals, the main beneficiaries of this methodology have been the physical, life and medical sciences. This is due to the fact that these disciplines publish frequently with multiple authors. In contrast, the social sciences and humanities are likely to have single authors and to publish in a wide range of formats (monographs, policy reports, translations, etc.), whereas the arts produce major art works, compositions and media productions, and engineering focuses on conference proceedings and prototypes.

Since, as Thomson Reuters says 'English is the universal language of science at this time in history', international databases have tended to favour English language publications. This

disadvantages the social sciences and humanities which often consider issues that are primarily of national relevance, and publishing them in the national language. It can also benefit countries where English is the native language, and countries which publish the largest number of English-language journals.

This disparity is further reflected in citation practices. Citations aim to measure the impact of research on academic knowledge. The system however has natural limitations and is open to gaming. Authors are most likely to reference other authors whom they know; given an intrinsic tendency to reference national colleagues and/or English-language publications, the reputational or halo factor implies that certain authors are more likely to be quoted than others. This may occur because of the significance of their work but it may also be due to informal networks. Self-citation, by which authors reference their own work, can also have a knock-on positive affect.

Bibliometric and citation databases capture past performance which is usually interpreted as an indicator of future potential. As a result, new research fields and inter-disciplinary research tends to be squeezed out; papers that challenge orthodoxy sometimes find it hard to get published and/or are less likely to be published in high-impact journals.

> Because articles published in new journals remain invisible to most citation indices, they also remain invisible to almost all ranking systems. Such invisibility dramatically skews scholarship,...implicitly encourag[ing] conservatism... (Adler and Harzing, 2009,:78)

There is also an assumption that journal quality is a proxy for article quality.

By measuring impact in terms of papers read by academic peers, citation and bibliometric indices can ignore research which affects policy, legislation or regulatory regimes, technological or social interventions, business creation and employment, etc. This is a key omission – not just because it advantages certain disciplines over others but because it projects a narrow image of research.

Research has traditionally been divided up into two categories: basic and applied. Over time, these boundaries have tended to blur as research and researchers engage in all aspects of the knowledge triangle. Knowledge has also become more democratized as an increasing number of people become aware of the issues and actively contribute to the application of knowledge. Yet, collaborative research and its social impact or economic benefits do not usually form a central feature of assessment. Admittedly, social impact or economic benefits can be difficult to measure, but its value, to paraphrase Einstein, derives from the ability to measure what counts rather than that which can be easily measured.

Peer review represents a cornerstone for research assessment. Assessing research quality requires a detailed understanding of the field and its contribution to knowledge. But peer review also has its limitations. Evaluators often evaluate research in terms of what they know; therefore, novel and challenging ideas are often marginalised as noted above. Marginson notes, ‘Not all path-breaking innovations gain early peer recognition and some are sidelined precisely because they challenge established ideas’ (2008: 17). Peers often conform to conventionally accepted patterns of belief, and may be influenced by a researcher’s reputation rather than his or her actual contribution to knowledge.

Finally, the results of the research assessment process are usually publicised as institutional results. Because research is increasingly conducted by teams, individual performance data is aggregated using the research field, discipline or department as the unit of assessment (Individual performance usually serves for promotional or similar purposes.) While this

method offers the best opportunities for comparison, both within and between HEIs, comparisons at the department level can be problematic because they are often historical constructs. Nevertheless, it is best to assess research at the sub-institutional level in order to overcome natural distortions that arise when results are aggragated to the institutional level. This is because large HEIs, especially those with medical schools, do best in systems which simply quantify total output, such as global rankings. Most HEIs are excellent in certain domains and in need of improvement in others; whole-HEI comparisons brand everything according to the majority. Differences in disciplinary practice or new or emerging fields of investigation can be undermined by this method.

Research Assessment 'Good Practice'

In order to overcome many of these limitations, careful attention must be paid to the purpose of research assessment. This, of course, depends on the end user, e.g. policymakers and government agencies, HEIs, public or private research organisations (PROs), potential researchers or graduate research students, employers, civil society and the media. Each group uses information differently to satisfy a diverse and often conflicting set of objectives. The experience of rankings suggests that the number of users and uses is increasing and that it is not possible to control the ways in which people use or interpret the data once it has been published.

The choice of indicators is therefore vital; the results can impact on individual, institutional and national reputation and status, students' choices and opportunities, and our own understanding of knowledge and knowledge production (Hazelkorn, 2009). Thus, indicators should be ready-to-use, appropriate and verifiable, and the process must be transparent and replicable. It should enable decision-making by internal and external users, and facilitate comparisons over time and across different types of HEIs. Indicators should not be affected by any bias and they should instil trust. In other words, those being assessed must believe in the indicators' appropriateness and truthfulness. Having too few indicators can lead to distortion. Too many can make the exercise complicated and costly. Ultimately, the choice and weight of indicators should seek to strike a balance between fairness and feasibility (European Commission, 2006; Cañibano et al. 2002).

'Good practice' suggests that research assessment should

1) *Combine indicator-based quantitative data with qualitative information*, for example information based on expert peer and/or end-user assessment. This enables the quantitative information to be tested and validated within the context and purpose of the assessment.
2) *Recognise important differences across research disciplines*. Peer-reviewed journal articles represent the primary publication channel for practically all academic disciplines. However, the complexity of knowledge has led to a diverse set of output formats: audio-visual recordings, computer software and databases, technical drawings, designs or working models, major works in production or exhibition and/or award-winning design, patents or plant breeding rights, major art works, policy documents or briefs, research or technical reports, legal cases, maps, translations or editing of major works within academic standards, etc..
3) *Include impact and benefit assessment*. Assessment should include indicators capable of capturing and recognising the fact that research does not exist in isolation. This largely differs along disciplinary lines. It may include indicators such as graduate employment, the number of companies established and employees hired, changes to policy, legislation and regulatory

regimes, waste and pollution reductions or improvements in health care (see RQF, 2007). Stakeholder esteem indicators point to how research is viewed by the wider community. Among such indicators, one finds: keynote addresses; prestigious national and international awards and prizes; international visiting research appointments; appointments to advisory committees in national or international organizations.

4) *Involve self-evaluation* as a means of pro-actively including the research community in the assessment of its own contribution. It also represents a way of placing the research process – which includes the organization, management, and developments over time – into context and of making sure that it stays in line with the institution's mission (Spaagen et al, 2007).

Conclusion

The European Council's 2006 Communication, *Delivering on the modernisation agenda for universities: education, research and innovation*, illustrates the ways in which the legacy of rankings has become embedded in higher education policy:

> Universities should be funded more for what they do than for what they are, by focusing funding on relevant outputs rather than inputs,…Competitive funding should be based on institutional evaluation systems and on diversified performance indicators with clearly defined targets and indicators supported by international benchmarking.

This has implications not just for research assessment processes but on academic behaviour as well. There has been a clear shift from self-declaration to external verification of quality. Greater attention is being given to the issue of knowledge access. Open science, open source and institutional repositories are just some of the many existing alternatives that are being explored and adopted. In some cases, national agencies are pressing for these changes in order to maximise its visibility, accessibility and scientific impact for society and the economy, .

An important obstruction to a more inclusive research assessment process lies within the academy itself. Because research has the "capacity to shape academic careers at the point of hiring and promotion" (Marginson, 2008:17), it has become vital to identify indicators and methodologies that measure, assess and reward the full spectrum of research activity – across all disciplines, including interdisciplinary work, and discipline outlets. This will help to incentivise the academy, increase investor-confidence and inform the public. It is also vital because a major handicap for resarchers engaging in new forms of knowledge production is that recruitment, tenure, promotion and prestige still rewards traditional, disciplinary Mode 1 outputs.

While governments and national agencies may wish to set up simple processes, there is no single set of value-free indicators. Thus, the choice of indicators, the methodology used and the weightings assigned to them are vital. Greater attention needs to be given to all these factors in order to ensure that the process is fit-for-purpose and avoids producing unintended consequences.

References

Adler, N.J. and Harzing, A.W. 2009. When Knowledge Wins: Transcending the Sense and Nonsense of Academic Rankings. *Academy of Management Learning and Education*. Vol. 8, No. 1, pp. 72-95.

Cañibano, L., Sánchez, M.P., García-Ayuso, M., and Chaminade, C. 2002. *Guidelines for managing and reporting on intangibles. Intellectual Capital Report.* Madrid, Vodafone Foundation.

Australian Research Council. 2009. *The Excellence in Research for Australia (ERA) Initiative*. Canberra. http://www.arc.gov.au/era/default.htm (Accessed 22 August 2009.)

European Commission. 2006. *RICARDIS: Encourage Corporate Measuring Reporting on Research and Other Forms of Intellectual Capital.* Brussels. http://ec.europa.eu/invest-in-research/pdf/download_en/2006-2977_web1.pdf (Accessed 22 August 2009.)

Hazelkorn, E. 2008. Learning to live with league tables and ranking: the experience of institutional leaders. *Higher Education Policy*. Vol. 21, No. 2, pp. 195-215.

Hazelkorn, E. 2009. The impact of global rankings on higher education research and the production of knowledge. *UNESCO Forum on Higher Education, Research and Knowledge Occasional Paper*, No. 15..

Institute of Higher Education Policy. 2009. *Impact of College Rankings on Institutional Decision Making: Four Country Case Studies*. Washingtonm DC. http://www.ihep.org/assets/files/publications/g-l/ImpactofCollegeRankings.pdf (Accessed 22 July 2009.)

Marginson, S. 2008. The Knowledge Economy and the Potentials of the Global Public Sphere. Melbourne, Centre for the Study of Higher Education University of Melbourne. http://www.cshe.unimelb.edu.au/people/staff_pages/Marginson/Beijing%20Forum%202008%20Simon%20Marginson.pdf (Accessed 1 February 2009.)

Higher Education Funding Council of England. 2008. *Research Assessment Exercise.* http://www.rae.ac.uk/ (Accessed 26 July 2009.)

Development Advisory Group for the RQF. 2006. *Research Quality Framework: Assessing the Quality and Impact of Research in Australia*. http://www.unisa.edu.au/rqie/rqfhistory/docs/TheRecommendedRQFwebsiterelease14November2007.pdf (Accessed 22 July 2009.)

Elsevier B.V. *Scopus Overview: What is it?*. http://www.info.scopus.com/overview/what/ Accessed 22 July 2009)

Spaapen, J.H., Dijstelbloem, F., and Wamelink. 2007. *Evaluating Research in Context. A Method for Comprehensive Assessment*, 2nd edn.The Hague, Consultative Committee of Sector Councils for Research and Development.

Thompson Reuters. 2009. *The Thomson Reuters Journal Selection Process*. http://thomsonreuters.com/products_services/science/free/essays/journal_selection_process/ (Accessed 22 July 2009.)