

7-1-2012

Research Evaluation in Practice: Interview with Linda Butler

Gali Halevi
Elsevier

Follow this and additional works at: <https://www.researchtrends.com/researchtrends>

Recommended Citation

Halevi, Gali (2012) "Research Evaluation in Practice: Interview with Linda Butler," *Research Trends*: Vol. 1 : Iss. 29 , Article 6.

Available at: <https://www.researchtrends.com/researchtrends/vol1/iss29/6>

This Article is brought to you for free and open access by Research Trends. It has been accepted for inclusion in Research Trends by an authorized editor of Research Trends. For more information, please contact r.herbert@elsevier.com.

Section 5: Expert Opinion

Research Evaluation in Practice: Interview with Linda Butler

Gali Halevi, PhD



Linda Butler

Web: <http://www.lindabutler52.com.au/>

Email: linda.butler52@gmail.com

During your career, you have taken part in government-driven research projects using bibliometrics methodologies. Could you give an example or two of the outcomes of these research projects and the way they informed scientific funding?

The most influential body of research I have undertaken relates to analyses of the way Australian academics responded to the introduction of a sector-wide funding scheme that distributes research funding to universities on the basis of a very blunt formula. The formula is based on data on research students, success in obtaining competitive grant income, and the number of research outputs produced. For research outputs, a simple count is used. It does not matter where a publication appeared – the rewards are the same. By looking in detail at the higher education sector, and after eliminating other possible causal factors, I was able to demonstrate that the introduction of the formula led to Australian academics significantly increasing their productivity above long-term trend lines. While the increase was welcome, what was of major concern to policy makers were the findings that the increase in output was particularly high in lower impact journals, and that Australia's relative citation impact had fallen below that of a number of its traditional OECD comparators.

These findings were part, though not all, of the driver for Australia to introduce a new funding system for research. The same blunt formula is still being used, but it is anticipated that much of the funding it distributes will before long be based on the results of the Excellence in Research for Australia (ERA) initiative, the second exercise of which will be conducted in 2014 (the first was held in 2012). The same research has also been influential in Norway and other Scandinavian countries where governments sought to avoid the pitfalls of simple publication counts by introducing a tiered system of outputs, with those in more prestigious journals or from more prestigious publishers receiving a higher weighting and therefore resulting in greater funding.

See also: [Powerful Numbers: Interview with Dr. Diana Hicks](#)

Examining the literature, there appear to be far more research evaluation studies focusing on life and medical sciences. Why, in your opinion, are these not as prevalent in the social sciences?

I believe this is primarily because quantitative indicators are seen as fairly robust in the biomedical disciplines and are therefore, on the whole, reasonably well accepted by

researchers in those fields. This is not the case for the social sciences. There is nothing surprising in this. The biomedical literature is well covered by major bibliometric databases. In addition, sociological studies have given us much evidence on the meaning of citations in the life sciences and this, together with evaluative studies that have been shown to correlate well with peer review, means researchers have some confidence that measures based on the data are reasonably robust – though always with the proviso they are not used as a blunt instrument in isolation from peer or expert interpretation of the results.

The same can't be said for the social sciences (or the humanities and arts). There is some evidence that a citation in these disciplines has a different meaning – their scholarship does not build on past research in the same way that it does in the life sciences. It is also well known that coverage of the social sciences is very poor in many disciplines, and only moderate in the best cases. Evaluative studies that use only the indexed journal literature have sometimes demonstrated poor correlation to peer review assessments, and there is understandably little confidence in the application of the standard measures used in the life sciences.

What can be done to measure arts & humanities as well as social sciences better?

I think the most promising initiatives are those coming out of the European Science Foundation, which has for a number of years been investigating the potential for a citation index specifically constructed to cover these disciplines. The problem is that, as it would need to cover books and many journals not indexed by the major citation databases, it is a huge undertaking. Given the current European financial climate I don't have much confidence that this initiative will progress very far in the short-term. It is also an initiative fraught with problems, as seen in the ESF's first foray into this domain with its journal classification scheme (<http://www.esf.org/research-areas/humanities.html>). Discipline and national interest groups have been very vocal in their criticisms of the initial lists, and a citation index is likely to be just as controversial.

Many scholars in these disciplines pin their hopes on Google Scholar (GS) to provide measures that take account of all their forms of scholarship. The problem with GS is that it is not a static database, but rather a search engine. As GS itself clearly points out, if a website disappears, then all the citations from publications found solely in that website

will also disappear, so over time there can be considerable variability in results, particularly for individual papers or researchers. In addition, it has to date been impossible to obtain data from GS that would enable world benchmarks to be calculated – essential information for any evaluative studies.

Do you think that open access publishing will have an effect on journals' content quality, citations tracking and general impact?

The answers to these questions depend on what "open access publishing" means. If it refers to making articles in the journal literature that are currently only accessible through paid subscription services publicly available, I would expect the journal "gatekeepers" – the editors and reviewers – to continue with the same quality control measures that currently exist. If all (or most) literature becomes open access, then the short-term citation advantage that is said to exist for those currently in open access form will disappear, but general impact could increase as all publications will have the potential to reach a much wider audience than was previously possible.

But if "open access publishing" is interpreted in its broadest sense – the publishing of all research output irrespective of whether or not it undergoes any form of peer review – then there is potential for negative impact on quality. There is so much literature in existence that researchers need some form of assessment to allow them to identify the most appropriate literature and avoid the all too real danger of being swamped by the sheer volume of what is available. Some form of peer validation is absolutely essential. That is not to say that peer validation must take the same form as that used by journals – it may be in the form of online commentary, blogs, or the like – but it is essential in some format.

Any new mode of publication presents its own challenges for citation tracking. On the one hand, open access publishing presents huge possibilities in a much more comprehensive coverage of the literature, and potential efficiencies in harvesting the data. But on the other hand they present problems for constructing benchmarks against which to judge performance – how is the "world" to be defined? Will we be able to continue using existing techniques for delineating fields? Will author or institutional disambiguation become so difficult that few analysts will possess the knowledge and computer power required to do this?

What forms of measurements, other than citations, should be applied when evaluating research quality and output impact in your opinion? (i.e. usage, patents)

It is important to use a suite of indicators that is as multi-dimensional as possible. In addition to citation-based measures, other measures of quality that may be relevant include those based on journal rankings, publisher rankings, journal impact measures (i.e. SNIP, SJR etc.) and success in competitive funding schemes. Any indicator chosen must be valid, must actually relate to the quality of research, must be transparent, and must enable the construction of appropriate field-specific benchmarks. Even then, no single indicator, nor even a diverse suite of indicators, will give a definitive answer on quality – the data still need to be interpreted by experts in the relevant disciplines who understand the nuances of what the data is showing.

Choosing indicators of wider impact is a much more fraught task. Those that are readily available are either limited in their application (e.g. patents are not relevant for all disciplines), or refer merely to engagement rather than demonstrated achievement (e.g. data on giving non-academic presentations, or meetings with end-users attended). And perhaps the biggest hurdle is attribution – which piece (or body) of work led to a particular outcome? For this reason, the current attempts to assess the wider impact of academic research are focussing on a case study approach rather than being limited to quantitative indicators. The assessment of impact in the [UK's Research Excellence Framework](#) is the major example of such an approach currently being undertaken, and much information on this assessment approach can be found on the website of the agency overseeing this process – the [Higher Education Funding Council of England](#). See also: [Research Impact in the broadest sense: REF 14](#)

During your years as a university academic, did you notice a change among university leaders and research managers in the perception and application of bibliometrics?

From a global perspective, the biggest change has occurred since the appearance of university rankings such as the Jiao Tong and THE rankings. Prior to this, few senior administrators had much knowledge of the use of bibliometrics in performance assessments, other than the ubiquitous journal impact factor. The weightings given to citation data in the university rankings now ensure that bibliometrics are at the forefront

of universities' strategic thinking and many universities have signed up to obtain the data that relates to their own university and use it internally for performance assessment.

In Australia, most university research managers had at least a passing knowledge of the use of bibliometrics in evaluation exercises by the 1990s, through the analyses undertaken by the unit I headed at The Australian National University, the Research Evaluation and Policy Project. However their interest increased with the announcement that bibliometrics were to form an integral part of a new performance assessment system for Australian universities – the Research Quality Framework which was ultimately superseded by the ERA framework. This interest was further heightened by the appearance of the institutional rankings mentioned above. While ERA is not currently linked to any substantial funding outcomes, it is expected to have financial implications by the time the results have been published from the second exercise to be held in 2014. Australian universities are now acutely aware of the citation performance of their academics' publications, and many monitor that performance internally through their research offices.

The downside of all this increased interest in, and exposure to, bibliometrics is the proliferation of what some commentators have labelled "amateur bibliometrics" – studies undertaken by those with little knowledge of existing sophisticated techniques, nor any understanding of the strengths and weaknesses of the underlying data. Sometimes the data is seriously misused, particularly in its application to assessing the work of individuals.

What are your thoughts about using social media as a form of indication about scientific trends and researchers' impact?

I have deep reservations about the use of data from social media to construct performance indicators. They relate more to popularity than to the inherent quality of the underpinning research, and at this point in time are incredibly easy to manipulate. They may be able to be used to develop some idea of the outreach of a particular idea, or a set of research outcomes, but are unlikely to provide much indication of any real impact on the broader community. As with many of the new Web 2.0 developments, the biggest challenge is determining the meaning of any data that can be harvested, and judging whether any of it relates to real impact on either the research community, on policy, on practice, or on other end-users of that research.