Guest editorial: evaluation of research performance

K. Benke
School of Engineering,
University of Melbourne, Australia,
kbenke@unimelb.edu.au

S. Kilic
School of Population and Global Heath,
University of Melbourne, Australia

A perennial question discussed in many research institutions around the world is how best to measure the research performance of an individual. This editorial considers some of the methods used to do so, with an emphasis on the role of journal publication and its limitations, with suggestions for improving current practice.

At a national level, politicians are raising the prospect of linking research funding to innovation and economic impacts. In the private sector, performance is often measured by productivity improvements and patents. But most universities have traditionally generated little in terms of patents and intellectual property (IP) for business interests, even though there are signs that this could change in the future. So, in universities, research performance is gauged mainly by academic publications.

In other words, there remains a continuing schism between universities, where the performance of the scholar is almost invisible without publication, and many large, multinational companies which have a completely opposite attitude towards intellectual property. The attitude in business is that *if you publish your IP in the open literature for free, then there cannot be any value in your work*.

One consequence is that there now exists a very large database of internal, confidential research reports in the Aerospace, Defence and Industrial sectors, with many scientists having publication records that may be very extensive, but classified and so unavailable for the assignment of publication metrics. Such metrics should therefore be used mainly for comparing university researchers.

We will focus here on publication - particularly those in refereed journals even though other measures of research performance can be used, such as patents, economic impacts, community impacts and innovative software (Kilic, 2014). There are contentious aspects, however, associated with journal publication that are not widely apparent and require further elaboration.

Publication metrics are described in several recent reviews, see for example, Jarwal et al (2009), Scoble et al (2010) and Smith et al (2013). These reviews focus on critical analysis of very specific *bibliometric indicators* rather than the broader view that we wish to encourage in this paper.

There are at least three widespread, purposeful activities which are currently generating publication pressure on both academics and scientists:

1. Measurement of individual performance level,
2. Evaluation of competitive grant applications, and
3. Ranking of institutions.

Indeed, in recent years, there appears to have been an intensifying public discussion, about the importance of scientific research publication, within the open literature (see Laurance et al, 2013a, 2013b).

But why have publications in the first place? To answer, we suggest that there are at least ten functions that publications serve:

1. **Validation of research** (by peer review)
2. **Proof of performance** (for project assessment and awards)
3. **Impact of research** (as measured by citations)
4. **Grant applications** (based predominantly on publication record)
5. **Professional recognition** (for scientist and organisation)
6. **Precedence** (established for results and discoveries)
7. **Documentation** (of project work, methodology, data)
8. **Notification to peers** (in the scientific community)
9. **Archival storage** (for reference)
10. **Copyright** (often jointly with the scientific journal)

We consider the first five of these to be the most important. This is because functions 1-5 relate to the validation, peer review, assessment and impact of research, whereas functions 6-10 encapsulate administrative processing, registration, legal and intellectual property considerations.

It is clear that without publication, an academic project will have a tenuous and uncertain existence, both legally and scientifically. In fact, without satisfying points 1-10 above, any effort expended on the project may be completely wasted and not gain recognition, especially after substantial time elapses, such as years or decades. In short, the performance of a scholar will be almost completely invisible without publication (Laurance et al 2013a).

Laurance et al (2013b) take this point further when they write:

> *. . . if you are an academic, your publishing record will have a crucial impact on your career. It can profoundly affect your prospects for employment, for winning research grants, for climbing the academic ladder, for having a teaching load that doesn't absorb all your time, for winning academic prizes and fellowships, and for gaining the respect of your peers . . . It's not called 'publish or perish' for nothing.* (page 1).

Indeed, research has revealed a strong correlation between publication performance as sole or first author and the probability of winning a Nobel Prize - 75% of winners are found in the top 1000 most frequently cited scientists (Garfield & Malin, 1968).

At present, the gold standard of research performance for universities and public research organisations remains the 'peer-reviewed paper in an international journal', as evident from various reviews and online sources, e.g. Garfield & Malin (1968), Rennie et al (1997), Garfield (2010), Laurence et al (2013 a, b), SCIMAGO (2014).

A scientific paper is usually submitted to an academic journal listed in an online scientific database such as Scopus, Thomson Reuters 'Web of Science', or Google Scholar, and most research institutions have rankings for publications and their importance.

Usually, a peer reviewed journal paper outranks a book chapter, conference paper, letter article and internal report. Textbooks are traditionally compilations of existing knowledge and usually do not announce new findings or discoveries, or have the same level of rigour as a peer-reviewed journal paper. Note also that in private industry, internal reports may sometimes have commercial significance and may surpass a journal paper in commercial and intellectual value but be subject to protection issues relating to intellectual property.

Reviewers for a refereed journal may reject the manuscript outright, or require re-submission with significant changes, or accept it with minor changes. The process can be slow and demoralising, often taking years for a high impact journal. For example, *Nature* has a rejection rate that is not far short of 100%, whilst the *IEEE Transactions in Engineering* series can have a turn-around time of several years.

One result of this is that harried researchers often complain about the harsh external assessment of manuscripts by academics and professors is not fully appreciated by administrators whose jobs involve a more subjective and less stressful internal appraisal process.

Most publications fall into three main categories:

1. New theory and methodology,
2. Experimental results and field measurements, and
3. Strategy, commentary and critical review.

In some cases, these three categories are integrated into the same study to either provide overviews or frameworks, or to serve as a vision paper for future research. Publications reporting new theory, new methodology, new statistical techniques, new scientific technology or quality improvements are often rated highly whereas relative weighting may be lower for routine reporting of field measurements, surveys, trade journal publications and newsletters.

A difficult assessment problem occurs whenever a paper is based solely upon the development of a new software package, which in itself often represents a creative work. The importance of computer coding and software models has been recognised in recent research, although stand-alone papers on software issues and interfaces are still rare and computer source code is usually not provided. Hence there is possibly a need for increased publishing, and more frequent citation of original software and scientific 'Apps' for tablets and mobile phones.

Accordingly, the publisher Elsevier recently introduced an open access journal devoted to software research. The journal *SoftwareX* claims to bestow academic recognition upon software writers and it encourages application of software developments beyond the original domain of application. The journal is multidisciplinary and it accepts submissions from a variety of subject areas.

For example, an article by Hanwell et al (2015) describes an open source, cross-platform toolkit for scientific visualisation and data analysis, and other software papers include descriptions of domain-independent software tools relevant to chemistry, medical imaging and cosmological data.

An important consideration when it comes to assessing publications in refereed journals is the quality of the journal, as judged by its impact and ranking (see SCIMAGO, 2014). However, relying on a journal's impact and ranking is seldom straight forward because good papers can appear in low-impact journals and poor papers can appear in high-impact journals.

The value of any journal's Impact Factor is based on the citation rate of the journal. A citation index called The Science Citation Index (SCI), which originated with the Institute for Scientific Information (ISI), has been greatly expanded over the years and is now available on the Web of Science platform (Thomson Reuters) covering over 6000 journals (Garfield, 2010), and another example of a citation index is Google Scholar.

But the conventional Impact Factor is not the only metric used, and others are shown in Table 1. In all cases there are concerns about their value - they can be blunt instruments when applied unthinkingly across all disciplines (Jarwal et al, 2009; Smith et al, 2013). The highest scoring papers are often just literature surveys.

By way of definition, the **Impact Factor (IF)** is the number of cites per document over the previous two years (or longer) and a journal has an **H- index** of H if H of N papers have at least H citations each. For example, H equals 5 if 5 papers have been cited 5 times each (this indicator is highly regarded in the physical sciences). The **SJR indicator** applies different values to citations according to the international importance of the journal in question, and the algorithm used here is from the SCOPUS database. Finally, the **Quartile Ranking** is the relative standing of the paper in the journal in the discipline of interest. For example, Q2 means the paper is ranked in the second quartile. Note correlation between indicators may not be strong.

Journal:	**Impact Factor (IF)**	**H-Index**	**SJR**	**Quartile Ranking**
Journal of Neural Engineering	4.27	50	1.51	Q1
International Journal of Digital Earth	2.27	10	0.92	Q1
Applied Optics	1.89	128	1.14	Q1
Crop and Pasture	1.35	58	0.60	Q2
Animal Science Journal	1.14	20	0.60	Q1
Animal Production Science	1.04	48	0.63	Q1

Table 1 – A selection of metrics commonly used.

It should be remembered that the *citation number*, which underlies these methods, itself has limitations. For example, it is associated with authors only, and not all research practitioners are authors.

Moreover, it measures citation rate for the whole journal and benefits all papers published in that journal. It does not necessarily relate to an individual paper. In other words, it is an *implied credit* only for all papers in the journal and may not hold up under independent scrutiny.

Also, reasons for citing a particular paper as a reference may not relate to quality, or precedence. Reasons can also include convenience of access due to title key words used in Google-based search, or even self-citation to support a personal impact record.

A possible problem following manuscript submission is that the quality of a manuscript may be overlooked by a poor reviewer who might reject a paper due to ignorance or bias. The onus is on the editor for fair arbitration. However, because some journal editors use key words to search for the most recent issue of articles on the topic and then invite one of the authors to act as a referee, the consequences can be unjust. For example, an inexperienced post doctoral student may be selected as referee and then reject a prescient paper by a team of professors due to the referee's personal ignorance of the subject matter. Requesting the authors to provide a list of referees to the editor is not a solution because it introduces possible problems of selection bias and cronyism.

The underlying problem for editors is the difficulty in finding reviewers. Overloaded academics are now reticent, and so some other approach needs to be considered - perhaps actually paying for professional referees from academia.

Note that there is an effective metric which is not often appreciated in academia - the total number of downloads from the internet. It is usually much greater than the citation number. Yet cursory inspection of many online journals shows that there is not necessarily a correlation between download count and citation rate because the pool of readers is much larger than the pool of writers. Authors and researchers, however, are likely to be much more informed and critical than other readers, and so perhaps the Impact Factor and H-index remain resilient measures despite their shortcomings.

Nevertheless, Impact Factor scores do seem to vary greatly between disciplines, as indicated in Table 2. They tend to be much lower in the physical sciences than in the life sciences. This is due in part to the fact that highly mathematical and theoretical papers have a much smaller readership than papers reporting experimental results, as commonly found in the natural sciences.

Moreover, articles in the biological sciences are more likely to be understood by the interested public, including journalists, and therefore more likely to be cited than papers in engineering and technology. Curiously, papers in economics, commerce and management have a low IF, but this may be due to alternative sources of knowledge distribution such as trade magazines and internal reports.

Discipline	**Impact Factor**	
	Median	**Normalised**
Biological Sciences	5.63	1.00
Clinical Sciences and Clinical Physiology	4.15	0.74
Physical, Chemical and Earth Sciences	3.11	0.55
Public Health and Health Services	2.03	0.36
Engineering and Technology	1.48	0.26
Economics, Commerce and Management	0.65	0.12

Table 2 - Impact Factors by discipline. Data have been normalised here from the median results derived from Jarwal et al (2009).

Obviously, any Impact Factor-based comparison between two journal papers can only be made *within the same discipline* and not across disciplines. For example, the Biological Sciences average Impact Factor is *three times* that of Engineering and Technology.

Recently, one of the most cited authors in Australia remarked that his publications averaged 20 authors per paper, with his own name being typically 10th in the list, and so it is hardly surprising that issues relating to authorship weightings have been discussed in the recent literature (e.g. Rennie et al, 1997; Martinson et al, 2005).

The first author is generally credited as the main contributor, but in many disciplines, other 'authors' are included because they are members of the team, or represent management, or funding bodies. Such 'gift authorship' remains a continuing problem. It is theoretically possible to build a publication record, and a high IF or H-index, without writing a single word.

One way to minimise this effect is to collate citations for authors who are listed as the *first author* on papers. Another is to provide a statement at the end of articles detailing precisely what contributions were made by each 'author' (noting that valid contributions can also be made by so-called 'non-writing' authors). This latter editorial request is already becoming more common in medical journals.

Alternatively, in the USA and Germany, as well as with ARC grant applications in Australia, a trend has been to submit a shortlist, of say five career best publications, and then let the examiner evaluate research potential based on various metrics (and dispense with publication volume as a criterion). We suspect that this approach might become more widespread in future.

Possible metrics for a limited shortlist may include:

1. Type of publication (e.g. peer reviewed journal papers)
2. Order of authorship (e.g. citation rate as a first author, not coauthor?)
3. Number of authors listed (normalisation?)
4. Bibliometric indicators (H-index vs Impact Factor?)

In summary, the process of fairly attributing credit for the work of academics and scientists remains a complicated and increasingly fraught issue. We have discussed some of the problems involved, along with some possible solutions. Despite its shortcomings, peer-reviewed publication is still a very important metric for research evaluation. Nevertheless, it needs to be supplemented, in many instances, with data about patents awarded, innovative software written, financial return and socio-economic impact.

References

Bailey, C.A. (2006) - *A Guide to Field Research*, 2nd Ed., Sage, Thousand Oaks, CA (USA).

Bernhard, H.R. & Ryan, G. (2009) - *Analysing Qualitative Data*, Sage, Thousand Oaks, CA (USA).

Box, G.E.P. & Draper, N.R. (1987) - *Empirical Model-Building and Response Surfaces*. John Wiley, Hoboken, NJ (USA).

Garfield, E. & Malin, M.V. (1968) - December). Can Nobel Prize winners be predicted? in *135th Meetings of the American Association for the Advancement of Science, Dallas, TX (USA)*.

Garfield, E. (2010) - The evolution of the science citation index. *International Microbiology*, *10*(1), 65-69.

Hanwell, M.D., Martin, K.M., Chaudhary, A., & Avila, L.S. (2015) - The Visualization Toolkit (VTK): Rewriting the rendering code for modern graphics cards, *Software X, 1-2, 9-12*.

Jarwal, S.D., Brion, A. M. & King, M. L. (2009) - Measuring research quality using the journal impact factor, citations and 'Ranked Journals': blunt instruments or inspired metrics? *Journal of Higher Education Policy and Management*, *31*(4), 289-300.

Kilic, S. (2014) - *An Evaluation Manual. Agriculture Research Division*, Department of Environment and Primary Industries (DEPI), Parkville Centre, Victoria, Australia, page 62.

Laurance, B., Useche, C., Bradshaw, C. & Laurance, S. (2013a) - Predicting publication success for biologists. *BioScience* 63:817-823.

Laurance, B., Useche, C., Bradshaw, C. & Laurance, S. (2013b). Predicting who will publish or perish as career academics. *The Conversation*. http://theconversation.com/predicting-who-will-publish-or-perish-as-career-academics-18473. accessed 21 Jan. 2015.

Martinson, B.C., Anderson, M.S. & De Vries, R. (2005) -. Scientists behaving badly. *Nature*, *435* (7043), 737-738.

Rennie, D.. Yank, V. & Emanuel, L. (1997) -. When authorship fails. A proposal to make contributors accountable. *JAMA,* 278(7): 579–85.

Smith, K.M., Crookes, E. & Crookes, P. A. (2013) - Measuring research 'impact'for academic promotion: issues from the literature. *Journal of Higher Education Policy and Management*, *35*(4), 410-420.

SCIMAGO (SCOPUS Database) (2014) - http://www.scimagojr.com/. accessed Sept 2014.

Scoble, R., Dickson, K., Hanney, S. & Rodgers, G.J. (2010) - Institutional strategies for capturing the socio-economic impact of academic research. *Journal of Higher Education Policy and Management*, *32*(5), 499-510.
