

PERSPECTIVE

Evaluate the Evaluations!

RAGHAVENDRA GADAGKAR

Centre for Ecological Sciences and Centre for Contemporary Studies, Indian
Institute of Science, Bangalore, Karnataka 560 012, India. ragh@iisc.ac.in



DI:LOGUE

Science, Scientists, and Society

Evaluate the Evaluations!

RAGHAVENDRA GADAGKAR

Centre for Ecological Sciences and Centre for Contemporary Studies, Indian Institute of Science, Bangalore, Karnataka 560 012, India. ragh@iisc.ac.in

<https://doi.org/10.29195/DSSS.01.01.0005>

Bad practices seem to spread very rapidly and go to near fixation, making it very hard to subsequently replace them with good practices. Hence they linger long after their 'badness' has become widely recognized. Evaluating scientists and scientific papers by the Impact Factors (IF) of the journals in which they are published is one of the most pernicious of such lingering bad practices. In retrospect, it seems shocking that the practice of using IF has become so widely and uncritically accepted. The odds are so heavily stacked against the practice that I would have guessed that it won't take off the ground. As has been mentioned in many forums, IF measures the impact of the journal and not of the paper, citation practices vary from discipline to discipline and 'IF pressure' is sure to lead to bad publishing practices. The bad practices associated with the evaluation procedures go well beyond the use of Impact Factors. Since evaluations are best done by peer groups, the business of eliminating bad practices and ushering in good practices is best attempted as a self-organized process by academics themselves with as wide a participation as possible. Science academies have a critical role in functioning as conscience keepers to usher in good practices and as gatekeepers to keep out bad practices. I believe that science academies around the world are not doing as good a job in this regard, as they potentially can. Recently, three prominent academies, Academie des Sciences of France, Leopoldina of Germany and The Royal Society, London have issued an excellent, joint statement about what they consider good and bad practices.

Although the points they make have been repeated time and again, coming as a joint statement from three of the world's prominent science academies brings with it a certain amount of authority and credibility. Besides, there are some points in this statement that are especially worthy of note. Apart from unambiguously pointing the finger at the excessive use of bibliometric data, the statement suggests reducing the number and frequency of evaluations in the first place; evaluating, training and nurturing the best evaluators, and cautions that the new so-called 'Altmetrics' may not be much better after all. I believe that this statement should be widely read by all scientists and hence I am reproducing it below. Nevertheless, it is a statement by three 'foreign' academies. Hence, I would urge our own three science academies to seriously study the matter in the Indian context and issue our own statement and

bring to bear the pressure of their authority on the conduct of evaluations in India. Clearly, we need an urgent evaluation of the evaluation process itself



INSTITUT DE FRANCE
Académie des sciences



Leopoldina
Nationale Akademie
der Wissenschaften

THE
ROYAL
SOCIETY

October 27, 2017

Statement by three national academies (Académie des Sciences, Leopoldina and Royal Society) on good practice in the evaluation of researchers and research programmes

1. Introduction

The large increase in the size of the international scientific community, coupled with the desire to ensure the appropriate and efficient use of the substantial funding devoted to supporting scientific research, have understandably led to an increased emphasis on accountability and on the evaluation of both researchers, research activities and research projects (including recruitment, as well as the evaluation of grants and prizes). Given that there is a large diversity of procedures currently used in evaluations which have accumulated over time, it is now necessary to provide some guidelines for best practice in the evaluation of scientific research. Peer review, adhering to strict standards, is widely accepted as by far the best method for research evaluation. In this context, the present statement focuses on the evaluation of individual researchers.

Such an assessment by competent experts should be based on both written (journal articles, reviews, books, book chapters, patents, etc.) and other contributions and indicators of esteem (conference presentations, awards, public engagement activity, peer review activity, datasets shared, seminars, etc.). As a careful evaluation of scientific content and quality by experts is time consuming and costly, the number of evaluations should be limited and only undertaken when necessary, in particular for decisions on competitive academic appointments or funding large projects.

With the increase in the number of evaluations and the emergence of easily accessible electronic databases, the use of bibliometric measures has become an additional tool. However, there has been too much reliance on bibliometric indices and indicator-based tools as measures of performance by many evaluation committees and exercises, leading to the

danger of superficial, over-simplified and unreliable methods of evaluation. This bad practice involving the misuse of metrics has become a cause for serious concern.

Of particular concern are the widely used journal impact factors (IF) which are an estimate of the impact of the journal itself rather than the intrinsic scientific quality of a given article published within it – a point that has been made on several occasions and notably in the San Francisco Declaration⁽¹⁾. Outstanding and original work can be found published in journals of low impact factor and the converse is also true. Nevertheless, the use of impact factors as a proxy for the quality of a publication is now common in many disciplines. There is growing concern that such “IF pressure” on authors has increased the incidence of bad practice in research and the ‘gaming’ of metrics over the past two decades, in particular in those disciplines that have over-emphasized impact factors. Also, the so-called ‘altmetrics’ – a new form of impact measure – while adding an important and hitherto overlooked dimension to the measurement of impact, suffers from some of the same weaknesses as the existing citation-based metrics.

There is a serious danger that undue emphasis on bibliometric indicators will not only fail to reflect correctly the quality of research, but may also hinder the appreciation of the work of excellent scientists outside the mainstream; it will also tend to promote those who follow current or fashionable research trends, rather than those whose work is highly novel and which might produce completely new directions of scientific research. Moreover, over-reliance on citations as a measure of quality may encourage the formation of aggregates of researchers (or “citation clubs”) who boost each others citation metrics by mutual citation. It thus becomes important to concentrate on better methods of evaluation, which promote good and innovative scientific research.

2. Principles of good practice in the evaluation of researchers and research activities

Essential elements for the evaluation of researchers can be summarized as follows:

2.1. Selection of evaluation procedures and evaluators

Evaluators

Since the evaluation of research by peers is the essential process by which its quality and originality can be estimated, it is crucial to ensure that the evaluators themselves adhere to the highest standards and are leaders in their field. The selection of evaluators should be based on their scientific excellence and integrity. Their scientific achievements should be widely recognised and their curriculum vitae and research achievements should be easily accessible. Such an open process will ensure the credibility and transparency of the evaluations.

Evaluation processes

Since the number of excellent evaluators is limited, the number of evaluation processes should be reduced in order to avoid over-use of first-class evaluators. There is a concern that different agencies and institutions have carried out an excessive number of routine evaluations over the last decades, putting too much pressure on the best evaluators. First-rate evaluators are increasingly reluctant to commit to time-consuming and unproductive evaluation exercises. It is of great importance to reduce the number of evaluations and to confine them to the core issues of research that only peers are able to judge. Evaluators provide a “free resource” as part of their academic duty and this resource is over-exploited. Evaluating bodies must recognise that good evaluation is a limited and precious resource.

A page limit for submissions to all evaluation processes is needed. Excessively long submissions are counter-productive: evaluators need to be able to concentrate on the essentials, which is problematic with very lengthy submissions.

Rotation of evaluators is essential to avoid excessive or repeated influence from the same opinion leaders. The panel of experts should be adapted to reflect the diversity of disciplines or scientific domains. Although gender and geographical distribution will be factors in the selection of evaluating groups, excellence must remain the primary criterion.

2.2. Ethical guidelines and duties of evaluators

Evaluators should clearly declare possible conflicts of interest before the evaluation process. The confidentiality of expert reviews and of the discussions in the evaluation panel must be strictly respected to protect both the evaluators and the evaluated persons.

While reviewers have often learned the practice of evaluation by experience and self-teaching, this competence cannot be taken as given. Methods and approaches to evaluating and reviewing should become part of all researchers' competence as should the ethical principles involved. Evaluators should be made aware of the dangers of “unconscious bias”. There should, as far as possible, be equivalent standards and procedures for all research disciplines.

The evaluation procedures must also include mechanisms to identify the cases of biased or otherwise inappropriate reviews and exclude them from consideration.

2.3. Evaluation criteria

Evaluations must be based under all circumstances on expert assessment of scientific content, quality and excellence. Publications that are identified by the authors as their most important work, including major articles and books, should receive particular attention in the evaluation. The simple number of publications should not be a dominant criterion.

Impact factors of journals should not be considered in evaluating research outputs. Bibliometric indicators such as the widely used H index or numbers of citations (per article

or per year) should only be interpreted by scientific experts able to put these values within the context of each scientific discipline. The source of these bibliometric indicators must be given and checks should be made to ensure their accuracy by comparison to rival sources of bibliometric information. The use of bibliometric indicators should only be considered as auxiliary information to supplement peer review, not a substitute for it.

The use of bibliometric indicators for early career scientists must in particular be avoided. Such use will tend to push scientists who are building their career into well-established/fashionable research fields, rather than encouraging them to tackle new scientific challenges.

For patents a clear distinction should be made between the stages of application, delivery and licensing.

Success in raising research grant funding should, where relevant, be only one and not the dominant factor in assessing research performance. The main criteria must be the quality, originality and importance of the scientific research.

3. Short summary of the main recommendations

Evaluation requires peer review by acknowledged experts working to the highest ethical standards and focusing on intellectual merits and scientific achievements. Bibliometric data cannot be used as a proxy for expert assessment. Well-founded judgment is essential. Over-emphasis on such metrics may seriously damage scientific creativity and originality. Expert peer review should be treated as a valuable resource.

Reference

1. <http://www.ascb.org/files/SFDeclarationFINAL.pdf>