
MEASURING SCIENTISTS' PERFORMANCE: A VIEW FROM ORGANISMAL BIOLOGISTS

Martin Ricker, Héctor M. Hernández and Douglas C. Daly

SUMMARY

Increasingly, academic evaluations quantify performance in science by giving higher rank to scientists (as well as journals and institutions) who publish more articles and have more citations. In Mexico, for example, a centralized federal agency uses such bibliometric statistics for evaluating the performance of all Mexican scientists. In this article we caution against using this form of evaluation as an almost exclusive tool of measuring and

comparing scientists' performance. We argue that from an economic viewpoint, maximizing the number of journal articles and their citations does not necessarily correspond to the preferences and needs of society. The traditional peer review process is much better suited for that purpose, and we propose "rule-based peer review" for evaluating a large number of scientists.

MIDIENDO EL DESEMPEÑO DE LOS CIENTÍFICOS: UN PUNTO DE VISTA DE BIÓLOGOS ORGANÍSMICOS

Martin Ricker, Héctor M. Hernández y Douglas C. Daly

RESUMEN

En la ciencia hay una fuerte tendencia global de cuantificar el desempeño de los científicos (así como a las revistas e instituciones), dando mayor jerarquía a aquellos científicos que publican más artículos y son más citados. En México, por ejemplo, una institución federal centralizada usa tales estadísticas bibliométricas para evaluar el desempeño de todos los científicos del país. En este artículo advertimos sobre los inconvenientes de esta forma de evaluación como una herramienta casi úni-

ca para medir y comparar el desempeño de los científicos. Argumentamos que, desde un punto de vista económico, la maximización del número de artículos científicos y de la frecuencia de sus citas no necesariamente corresponde a las preferencias y necesidades de la sociedad en general. El proceso tradicional de arbitraje por pares es más adecuado para este propósito, y proponemos el "arbitraje por pares basado en reglas" para evaluar a un número alto de científicos.

Introduction

There is an ongoing and necessary discussion of the proper and improper forms of performance evaluation of scientists as well as of scientific journals and institutions. The use of bibliometric data in indices for quantifying performance (Garfield, 1979), such as most recently the Hirsch or h-index to evaluate scientists (Egghe and Rao, 2008), has generated much criticism (Adler *et al.*, 2008; Leydesdorff, 2008).

These evaluation criteria are often inadequate, biased, and unfair to many fields. There have even been bitter complaints about how the pressure to maximize the number of articles and their citations encourages unethical behavior, such as inclusion as co-author without substantive contribution or repeating the same material in different articles, and distorts scientific work that should be aiming for innovative ideas (Lawrence, 2003, 2007; Steele *et al.*, 2006). Consequently, it

would be important to establish appropriate criteria for each field, and to combine quantitative (potentially automated) with qualitative evaluation.

Here we first present the system for quantifying scientific performance that has been used by Mexico's government. Subsequently, we argue with examples from our own fields of study (taxonomy, biogeography, and forestry) against blindly using a journal's impact factor as a measure of quality of the published articles. We

show that one cannot place realistic societal value on a scientist's performance using an ostensibly "objective" algorithm. Finally, we propose an intermediate form of evaluation which we call "rule-based peer review," when the purpose is to evaluate a large number of scientists.

The Sistema Nacional de Investigadores in Mexico

In 1984, Mexico established a centralized "National Sys-

KEYWORDS / Academic Evaluation / Citation Statistics / Peer Review / Sistema Nacional de Investigadores / SNI /

Received: 05/20/2009. Modified: 10/31/2009. Accepted: 11/02/2009.

Martin Ricker. Ph.D. in Forestry and Environmental Studies, Yale University, USA. Researcher, Universidad Nacional Autónoma de México (UNAM), Mexico. Address: Instituto de Biología, Departamento de Botánica. Apartado

postal 70-233 / Circuito Exterior s/n, Ciudad Universitaria, Delegación Coyoacán, México D.F. 04510, Mexico. e-mail: mrricker@ibiologia.unam.mx

Héctor M. Hernández. Ph.D. in Plant Systematics, Missouri Botanical Garden and Saint Louis University, USA. Researcher, UNAM, Mexico. e-mail: hmhm@ibiologia.unam.mx

Douglas C. Daly. Ph.D. in Plant Biology, City University of New York, USA. B.A. Krukoff Curator of Amazonian Botany and Director of the Institute of Systematic Botany, The New York Botanical Garden, USA. e-mail: ddaly@nybg.org

RESUMO

Há uma forte tendência global para avaliações acadêmicas que quantifiquem o desempenho nas ciências através de ranquear os cientistas (assim como revistas e instituições) que publicam mais artigos e têm mais citações. No México, por exemplo, um órgão centralizado do governo utiliza tais estatísticas bibliométricas para avaliar o desempenho de todos os cientistas mexicanos. No presente artigo, chamamos atenção ao uso desta forma de avaliação como ferramenta quase que exclusiva para

medir e comparar o desempenho dos cientistas. Argumentamos de um ponto de vista econômico que maximizar o número de artigos e as suas citações não corresponde necessariamente às preferências e necessidades da sociedade. O processo tradicional de avaliação pelos pares é mais apropriado para esta finalidade, e propomos um sistema baseado em "avaliação pelos pares seguindo regras", que podem avaliar um grande número de cientistas.

tem of Researchers" (Sistema Nacional de Investigadores or SNI; www.conacyt.mx/SNI/Index_SNI.html). Comparable evaluation systems exist in some other countries, for example in Argentina at the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Mexican scientists who wish to advance in their careers apply for acceptance in the SNI, presenting their scientific products. In 2008, 14681 scientists were accepted members of the system. If accepted into the system, members are ranked in four levels, and generally are re-evaluated every three to five years. The period depends on the rank and on whether the scientist is a new member in the SNI; distinguished members can be given 10-year evaluation periods and even life membership. In general, a significant portion of income of Mexican scientists depends on the ranking they are assigned. A Mexican scientist who is not a member of the SNI is generally considered a failure.

The evaluation is carried out by committees for seven different scientific fields, where biology and chemistry are combined in one committee. Each committee has 14 scientist evaluators selected from the highest SNI category; they serve for up to eight years, and each year they evaluate a portion of all SNI-registered scientists, as well as new applicants. On average, each committee annually

has to assess the productivity during the previous 3-5 years of over 500 scientists. Panel sessions of the seven committees discuss each case. It is obvious that the evaluators, who are active scientists themselves, have to assess each case quickly, and that they will not necessarily be specialized in the exact field of the scientist being evaluated. Consequently, they are under high pressure to automate the evaluation process as much as possible. The number of articles published in those journals that are included in the *Journal Citation Reports* of the *ISI Web of Knowledge* is a key element for the assigned SNI category. The ISI impact factors of the journals where articles are published is now often taken as a quantitative measure of the "value" of the journal by the evaluators.

Started by Eugene Garfield in the 1960s, the Institute for Scientific Information (ISI) now belongs to the US company Thomson Reuters. One product of Thomson Reuters is the *ISI Web of Knowledge*, which includes 20 databases, such as *Biological Abstracts*, *Journal Citation Reports*, and *Science Citation Index Expanded*. For 2007, the *Journal Citation Reports* registered 6426 scientific journals with their respective impact factors. Mabe (2003) estimated for 2001 that worldwide there were 14694 peer reviewed scholarly and academic journals, based on *Ulrich's Periodicals Directory* (2001 edi-

tion). He also reported an average annual growth of 3.25% in the number of journals during the 20th Century. The estimated number of journals in 2007 may consequently be estimated as $18787 (14694 \cdot e^{7 \text{ yrs} \cdot 0.0325})$. The selection for *Journal Citation Reports* in 2007 would therefore represent only 34% of all scientific journals worldwide, but the criteria for inclusion of a journal in *Journal Citation Reports* are not made completely transparent by Thomson Reuters, and it is worth noting that the company is under lobbying pressure by publishing houses (Leydesdorff, 2008: 282).

The (mis)use of a journal's impact factor

Developed in 1955, the journal impact factor is a proxy for the mean frequency that articles in a given journal are cited shortly after publication (Garfield, 2006). The ISI impact factor of a given journal for 2007 is the number of citations that articles published in the years 2006 and 2005 received in 2007, divided by the number of articles published during those two years in the journal. The highest ISI impact factor of all journals for 2007 was 69.0, reached by *CA: A Cancer Journal for Clinicians*. This journal, however, publishes review articles rather than original research articles, and review journals have significantly higher impact factors to start with (Leydesdorff, 2008: 281). The

second-highest ISI impact factor for 2007 was 52.6, reached by *The New England Journal of Medicine*. The eminent scientific journals *Nature* and *Science* pale in comparison, with impact factors for 2007 of 28.8 and 26.4, respectively. Numerous authors have pointed out that the most important variable for explaining differences of impact factors is not the quality of the journal, but the differences of citation numbers and citation behavior among different scientific fields (Seglen, 1992, 1997; Leydesdorff, 2008; Althouse *et al.*, 2009; Costas *et al.*, 2009). Taxonomy as a science has fared badly in scientific evaluations that involve impact factors and citation analyses (Krell, 2002; Agnarsson and Kuntner, 2007; Grimaldi and Engel, 2007). Two high-quality international journals of plant taxonomy, *Kew Bulletin* of the Royal Botanic Gardens, Kew, and *Harvard Papers in Botany* of Harvard University, are not even included in Thomson Reuters' *Journal Citation Reports* for 2007.

Scientific evaluations that encourage exclusively the production of articles in ISI-registered journals discourage taxonomists from embarking on long-term, fundamental projects. Currently, for example, there is no research project to produce a comprehensive *Flora of Mexico* that would present descriptions, taxonomy, distribution, and other information for each plant species in this megadiverse country.

The extraordinary diversity of Mexico's flora and fauna is frequently highlighted, but it is well-known only for some charismatic groups (mammals, reptiles, birds). The *Global Strategy for Plant Conservation* (www.cbd.int/gspc/) cannot effectively address any of its targets seriously in Mexico because of the lack of a rigorous biological inventory.

To make the different publication strategies among different scientific fields clearer, we compared one volume in 2008 of each of three international scientific journals that are considered highly regarded in each of their respective scientific fields: *Novon* (for describing new plant species), *Forest Science*, and the aforementioned *New England Journal of Medicine* (this journal had two volumes in 2008 of which we took the second). The impact factor for 2007 of *The New England Journal of Medicine* (52.6) was 339 times higher than *Novon's* impact factor (0.155), and 42 times higher than that of *Forest Science* (1.26). In medicine, there are many more scientists publishing together and citing each other reciprocally and immediately. The median number of co-authors of 114 articles in 2008 in *Novon* was 2 (range 1-5), and the median number of co-authors in *Forest Science* was 3 (range 1-7). These figures contrast with the median number of 16 co-authors in 99 original research articles in *The New England Journal of Medicine* (range 3-84), i.e., 8 times the median number in *Novon* and 5.3 times that of *Forest Science*. The differences are statistically highly significant.

Assume that 16 medical co-authors each write one article per year, in which all other co-authors are always included (for whatever contribution they made to the study). At the end of a year, each co-author has published 16 articles in a "high-impact" journal. In contrast, the two taxonomic co-authors have published "only" two articles each in a "low-impact" journal. They

would be considered much less productive and their work would be labeled as having little impact.

Given the current discussions, the obvious needs to be said: *The New England Journal of Medicine* does not have 339-times the quality and "impact" of *Novon*. Furthermore, any manuscript submitted to one of the journals is not appropriate for submission to either of the other two journals, i.e., they are not alternative choices for authors. The pressure to publish in journals with high impact factor misguides scientists, especially young ones who are starting their careers (Samyn and Massin, 2002; Cheung, 2008). Even the editor-in-chief of the "high-impact" journal *Nature* argues that "for a sure assessment of an individual, there is truly no substitute for reading the papers themselves, regardless of the journal in which they appear" (Campbell, 2008: 7). The choice of a scientific journal for submission of a manuscript should target the corresponding audience and not the journal's impact factor (Macdonald and Kam, 2007).

The descriptive branches of biology generally "suffer" low citation rates, increasing even more the pressure to publish a high number of articles, because each article is considered of low impact and prestige. One factor that influences the citation rate is the small size of many professional guilds. For instance, the systematic study of the Psocoptera, a group of insects containing about 5000 described species distributed in five continents, is currently conducted by seven internationally recognized entomologists, most of them senior researchers working in different universities. This small group of people has the daunting task of describing and interpreting this diverse group of living organisms, for which it has been estimated that at least 5000 additional species are yet to be studied and formally described (Alfonso

García-Aldrete, personal communication, Dec 2008). The combined scientific productivity, and consequently the number of citations generated by this group of scientists, cannot compete statistically with the scientific output registered in popular scientific fields such as biomedical research or biotechnology.

Consider the description of a new biological species in a taxonomic journal. To start with, the publication of a new species will be of utmost interest in the region of its geographical distribution, justifying publication in a less-cited journal that is intended for a more regional audience. Subsequently, it is the species name and not the original publication that will be cited in diverse other publications in the fields of ecology, conservation biology, and systematics. Ideally, and ethically in this context, citation indexes should credit the author of a species name every time a species is cited in scientific literature. Instead, the original publication typically is cited in its complete form (that enters in citation indexes) only by the original author or by another taxonomist revising the taxonomic group to which the new species belongs. If the original species description is accepted and receives only this one formal citation, it has achieved already a major purpose for science. Receiving only this one citation does not mean that it is a low-quality paper or unutilized.

Another factor that causes low citation rates is the long time that articles in taxonomy can remain relevant. Ricker and Hernández (in press) recently compiled an updated list of all Mexican tree species in the gymnosperms, monocotyledons, and tree ferns. For each of the resulting 170 species, the original publication of the species is cited. In 2009, the time back to the species' original description ranged from 8 to 256 years, with a mean (as well as a median) time of 104 years. It is nonsensical for citation analyses involving

such time intervals to make or break careers.

The weak relationship between citation frequency and research quality

We question the premise that "it is better to publish more than less and that the citation count of a paper is a useful measure of its quality" (Lehmann *et al.*, 2006). First of all, with an estimated one million scientific papers being published per year in many thousands of journals worldwide (Mabe, 2003: 193), and ever easier access to publications via the Internet, it is imperative in science to avoid adding to the clutter by publishing the same material numerous times in various forms. We argue that it is better to publish fewer but higher-quality articles, with "quality" in the empirical sciences meaning rigorous obtainment of data, possibly over a longer time, and extensive testing of methods. Even students' theses are nowadays frequently available via the Internet, thus lowering the pressure to publish all the theses' findings immediately. The service to science and society is much higher when authors attempt to produce the best possible science each time they work on a publication. In taxonomy, it will often be more valuable to produce a holistic monograph that describes a number of new species together in the same article or book, comparing and contrasting them in the same publication, rather than to maximize the number of articles by publishing each species separately one after another.

Second, a high number of citations does not necessarily imply high quality. There is no convincing, comprehensive theory for explaining (and distinguishing among) authors' choices of references (Camacho and Núñez, 2009). Many citations are not essential to the substance of an article, and many are replaceable (Cozzens, 1988). For example, often reviewers suggest addi-

tional citations (including their own), and the author includes them simply to expedite the review process. Indeed, there is a lively discussion about what proportion of cited articles has actually been read by the authors citing them, rather than the reference having been simply copied from a list of references. By modeling the propagation of citation misspellings, Simkin and Roychowdhury (2005a) estimate for the physics literature that 70-90% of scientific citations has been copied from the list of references used in other papers, rather than from the cited articles themselves. Todd and Ladle (2008) examined 306 papers from 51 ecology journals, and found that only 76% of the citations clearly supported the assertion they were intended to reinforce. While the accessibility of whole articles via the Internet has vastly improved during the last decade, the increasing number of available publications makes reading whole articles before citing them ever less likely.

An important feature of citation behavior is that once an article is cited (for whatever reason), it has a higher probability of being cited again, compared to articles that have not (yet) been cited. This behavior results in a positive feedback that causes some articles to be cited much more than others. Simkin and Roychowdhury's (2005b) model showed that this phenomenon by itself can explain that in the physics literature only 44 papers (0.18%) out of 24000 were cited 500 times or more. They concluded that this was a result of "mathematical probability, not genius."

Finally, high citation frequencies may also be of little importance to society. An uncited publication is not necessarily an unread publication. Even a publication that has never been cited can have value. The publication could have been an important step for a scientist in the exploration of a research frontier, who subsequently used it to develop

an article that is cited and that includes the experience reported in the uncited publication (Seglen, 1992: 635). It can also be a publication that has been useful for governmental agencies, teaching, technological development, or for consultation by readers who do not write articles themselves. For the preparation of the present article, for example, we consulted a number of articles that were useful but did not make it into the final list of references.

Economic and societal valuation of scientific performance

Ultimately, the discussion of how to assign a value to the work of a scientist must be broadened beyond the number of published articles and indicators of citation frequencies. The topic of valuation is a major focus of economics, and much can be learned from that field in this context. The concept of a market price exists in principle (though hidden) also for scientists and academia. There is public demand for scientific work, and a scientist who performs better has a higher value for society; the problem of course is how to measure performance. The work of scientists in public institutions does not necessarily have a market value, because it is carried out in and for the public domain, and not as a particular response to commercial interests of an individual or a group of people. Such public goods do not enter markets, and they remain non-market goods and services. The fact that they are not found in the market, however, does not mean that their economic value is zero, only that quantifying their value requires indirect methods because the market does not reveal it directly (e.g., Ricker, 1997). In the case of science, this non-explicit value becomes clear when companies hire scientists and the salary is negotiated between the scientist and the company as a function of the scientist's

quality and achievements, as well as of the labor market.

The generation of new knowledge and its value is not measurable as the number of published articles and the received citations in subsequent articles over a given period of time. For example, several publications by the same author can represent variations of the same topic, often including much of the same data, which ideally should have been condensed into a single, integrated, higher-quality article. In the case of several authors who publish an article together, the type and degree of contribution of each co-author can vary dramatically, and credit for the publication cannot be divided without inquiring first with the authors, if at all. Frequently it will be misleading to count "one more article" for each of the participating authors as a measure of their scientific performance. Counting the number of published articles is artificial, because society does not demand a lot of printed paper, especially text that is not intelligible for the general public or even educated laymen. Society instead wants more knowledge that will have a positive impact on human well-being and culture (see Ricker, 1997).

It is nonsensical to search for a single criterion for evaluating all scientists, journals, or institutions, because science maximizes its contributions for society when different scientists focus on different products. For example, specialization of scientists in applied research and in collaboration with industry or governmental institutions generally does not maximize citations (Aksnes and Taxt, 2004: 40). It can, however, produce scientific results that have been "tested on the ground," and can be transferred directly to users. This approach should not be discouraged as being "unproductive," as it would be using current bibliometric criteria.

Furthermore, scientific performance is frequently only a part of overall academic

performance. While scientific performance means advancing knowledge, academic performance also includes teaching, institutional development, consulting for public institutions and companies, knowledge transfer to the general public, creation of new program initiatives, and in the case of the current authors, field exploration, curation of natural history collections, and database development. Furthermore, knowledge is advanced not only by publishing articles in journals that are approved by Thomson Scientific, but also by publishing in other journals, writing books, and developing thesis projects with students. This is particularly true in the cases of organismal biologists, forest scientists, geographers, and geologists, among others, who frequently publish descriptive reports and maps of high regional relevance in journals that are intended for national or even regional audiences, possibly with high impact on society but low or no ISI impact factor. The appropriate combination of academic activities varies among scientific fields and institutions. Consequently, the performance evaluation of scientists must take into account institutional priorities and needs; there is no one-way-fits-it-all algorithm, neither in organismal biology nor in other fields such as physics (Pijpers, 2006).

Rule-based peer review

Science is a key element for generating total economic growth (Ricker, 1997), but it is also expensive. Governmental as well as non-governmental funding agencies obviously want to know what they are getting in return for their investments. Scientists have to think about some form of comparison among their scientific products and results (Giske, 2008). The traditional method for evaluating science and scientists over centuries has been peer review, generally involving external reviewers (see Weller, 2001). This

approach can be supplemented but in no way replaced by careful interpretation of citation statistics (Adler *et al.*, 2008; Giske, 2008).

Peer review as an evaluation method has of course also been criticized. A lively discussion is found in connection with an article by Cicchetti (1991) about the (in her view) unreliability of peer review. The article is followed by comments from 34 experts. We agree with the critical comment from John Bailar (Cicchetti, 1991: 137) that “the purpose of peer review is not reliability (of achieving the same evaluation among reviewers), but to improve decisions concerning publication and funding.” The process of critique and rebuttal between an author and a peer can and should be a highly constructive process that ultimately leads to better scientific products. A good peer reviewer attempts to reach an integrated assessment that serves the best interest of science and society. In this way the peer review system is comparable to that of judges in the legal system.

Peer review, however, is not practical for conducting many hundreds of evaluations during a relatively short time. Stopping short of proposing to dissolve large, centralized evaluation systems like the Mexican SNI altogether, a semi-automated evaluation makes sense. Rather than trying to evaluate each case independently, the evaluation committee members could establish for given scientific fields a point value for each academic product on a predefined scale. All point values for an applicant would be summed up and would need to reach a certain threshold for the applicant to be assigned a certain rank. Applicants could provide all necessary information on-line via the Internet (as is already done for the SNI), submitting corresponding documents as proof. The computer would basically calculate the outcome. The role of the system’s evaluators would be two-fold: First and most importantly, they would establish

the criteria for each field by defining the different products and discussing among them their (point-) values, which could gradually be adjusted during subsequent application cycles. For each scientific field they would also decide which scientific journals are meaningful, rather than blindly accept the list of the *Institute for Scientific Information*. Second, the evaluators would review the outcome of the applied algorithm for each applicant, to see if the assigned rank makes overall sense (i.e., all products fulfill the criteria for the assigned point values), and check aspects of innovation and special achievements. If the point values for different academic products are made public, the applicants will be able to anticipate their expected rank, rather than be surprised by the result. The scientists being evaluated could ask for reconsideration of their case, arguing in which way the semi-automated system has not been just for them.

Consider a hypothetical example of such a system. Assume a scientist has completed the following products over the past three years, for which the corresponding (also hypothetical) point values are mentioned in parentheses:

- One accepted article in an international journal that is recognized in the scientist’s field, where the scientist is the first author (10 points);
- Another article in which the scientist is not the first author (5 points);
- One defended doctoral thesis for which the scientist was principal advisor (10 points);
- Three substantial (open access) technical reports for industrial development, where the scientist is the first author ($3 \times 6 = 18$ points);
- One graduate course taught (8 points);
- Receipt of a significant international award (5 points).

The sum is 56 points. Assume a second scientist who

during the same period has published nine articles in international journals, though none as first author ($9 \times 5 = 45$ points). If the threshold for the highest rank (e.g., SNI level 3) were 50 points, the first scientist would reach it with a diversity of scientific products, the second scientist with a high number of articles as co-author would not. In contrast, in the current form of evaluation by the SNI, the first scientist would not even be accepted into the system (three publications in ISI-registered journals during three years have been the minimum to reach level 1), while the second scientist would probably reach the highest rank in the SNI.

Given the proposed rule-based peer review, the evaluators would answer quantitatively the following conceptual questions, which in the case of the Mexican SNI are currently left open:

- 1) How do different types of academic products and activities, such as scientific articles, book chapters, books, thesis supervision, course teaching, and technical reports compare in value?
- 2) To what extent are different scientific products replaceable? In the case of a stimulus system for research such as the SNI, not having published any article in an international scientific journal during the evaluation period may not be acceptable. On the other hand, presenting for a three-year period one high-quality article in an international scientific journal, several concluded theses supervisions, and constructive collaboration with industry could potentially be of higher value to society than three ISI-registered articles.
- 3) How should scientific innovation and creativity be valued? A point value system obviously has to leave room for considerations about the quality of the presented academic products and activi-

ties. The applying scientists could be asked to summarize their contribution in terms of innovation and creativity, and the contributions presented could be given a proper point value by the evaluators. If it is obvious that scientific articles repeat a large part of previously published ideas, results, and/or data, then points could be subtracted.

4) How are co-authors to be valued in comparison? Without knowledge of what was the contribution of each co-author, the first author could automatically receive a higher point value compared to the other co-authors.

5) How should scientific fields compare in the evaluation, and how long should the evaluation periods be? In the case of the SNI it is nonsensical to have a chemist involved in evaluating a taxonomist. Evaluation committees that are more qualified in a given scientific field, with their proper criteria (and corresponding point values), should be established. Evaluation periods longer than the current 3 to 5 years for most SNI members may make sense and the periods could differ among fields, though it would be especially important in that case that the reconsideration process worked properly. Alternatively, point values could be assigned for substantial progress reports that present the advances in long-term projects that lead to high-quality scientific books or taxonomic monographs.

We call the proposed method “rule-based peer review,” because the evaluators would be peers at least in a broad sense (belonging to the same scientific discipline) and would follow rules to establish the value of distinct products, rather than freely following their opinions. Some academic institutions in Mexico use such an evaluation system internally. The implicit danger is that

scientists are hunting point values rather than addressing important academic issues freely. For such a point-value system to work properly, it is important that it be used as a tool by evaluators for deciding if certain thresholds have been reached, for example for a scientist's promotion. It should not be applied blindly, without evaluators interpreting the overall picture. For cases such as the Mexican SNI, we think that such rule-based peer review would result in a much more integrated and better-balanced evaluation, in which objective indicators and value-based decision-making both have a place.

REFERENCES

- Adler R, Ewing J, Taylor P (2008) *Citation Statistics*. A report from the International Mathematical Union. www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf
- Agnarsson I, Kuntner M (2007) Taxonomy in a changing world: Seeking solutions for a science in crisis. *Syst. Biol.* 56: 531-539.
- Aksnes DW, Taxt RE (2004) Peer reviews and bibliometric indicators: A comparative study at a Norwegian university. *Res. Eval.* 13: 33-41.
- Althouse BM, West JD, Bergstrom CT, Bergstrom T (2009) Differences in impact factor across fields and over time. *J. Am. Soc. Inf. Sci. Technol.* 60: 27-34.
- Camacho-Miñano MM, Núñez-Nickel M (2009) The multilayered nature of reference selection. *J. Am. Soc. Inf. Sci. Technol.* 60: 754-777.
- Campbell P (2008) Escape from the impact factor. *Ethics Sci. Env. Polit.* 8: 5-7.
- Cheung WWL (2008) The economics of post-doc publishing. *Ethics Sci. Env. Polit.* 8: 41-44.
- Cicchetti DV (1991) The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behav. Brain Sci.* 14: 119-186.
- Costas R, Bordons M, van Leeuwen TN, van Raan AFJ (2009) Scaling rules in the sciences system: Influence of field-specific citation characteristics on the impact of individual researchers. *J. Am. Soc. Inf. Sci. Technol.* 60: 740-753.
- Cozzens SE (1988) What do citations count? The rhetoric-first model. *Scientometrics* 15: 437-447.
- Egghe L, Rao IKR (2008) Study of different h-indices for groups of authors. *J. Am. Soc. Inf. Sci. Technol.* 59: 1276-1281.
- Garfield E (1979) *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*. Wiley. New York, USA. 274 pp.
- Garfield E (2006) The history and meaning of the journal impact factor. *J. Am. Med. Ass.* 295: 90-93
- Giske J (2008) Benefitting from bibliometry. *Ethics Sci. Env. Polit.* 8: 79-81.
- Grimaldi DA, Engel MS (2007) Why descriptive science still matters. *BioScience* 57: 646-647.
- Krell FT (2002) Why impact factors don't work for taxonomy. *Nature* 415: 957.
- Lawrence PA (2003) The politics of publication. *Nature* 422: 259-261.
- Lawrence PA (2007) The mismeasurement of science. *Curr. Biol.* 17: R583-R585.
- Lehmann S, Jackson AD, Lautrup BE (2006) Measures for measures. *Nature* 444: 1003-1004.
- Leydesdorff L (2008) *Caveats* for the use of citation indicators in research and journal evaluations. *J. Am. Soc. Inf. Sci. Technol.* 59: 278-287.
- Mabe M (2003) The growth and number of journals. *Serials* 16: 191-197.
- Macdonald S, Kam J (2007) Aardvark *et al.*: Quality journals and gamesmanship in management studies. *J. Inf. Sci.* 33: 702-717.
- Pijpers FP (2006) Performance metrics. *Astron. Geophys.* 47: 6.17-6.18.
- Ricker M (1997) Limits to economic growth as shown by a computable general equilibrium model. *Ecol. Econ.* 21: 141-158.
- Ricker M, Hernández HM (in press) Tree and tree-like species of Mexico: Gymnosperms, monocotyledons, and tree ferns. *Rev. Mex. Biodiv.*
- Samyn Y, Massin C (2002) Taxonomists' requiem? *Science* 295: 276-277.
- Seglen PO (1992) The skewness of science. *J. Am. Soc. Inf. Sci. Technol.* 43: 628-638.
- Seglen PO (1997) Why the impact factor of journals should not be used for evaluating research. *Br. Med. J.* 314: 497.
- Simkin MV, Roychowdhury VP (2005a) Stochastic modeling of citation slips. *Scientometrics* 62: 367-384.
- Simkin MV, Roychowdhury VP (2005b) Copied citations create renowned papers? *Ann. Improb. Res.* 11: 24.
- Steele C, Butler L, Kingsley D (2006) The publishing imperative: The pervasive influence of publication metrics. *Learn. Publ.* 19: 277-290.
- Todd PA, Ladle RJ (2008) Hidden dangers of a 'citation culture'. *Ethics Sci. Env. Polit.* 8: 13-16.
- Weller AC (2001) *Editorial Peer Review: Its Strengths and Weaknesses*. Information Today. Medford, NJ, USA. 342 pp.