

Book Review

Evaluations of Individual Scientists and Research Institutions. Edited by Tibor Braun. Budapest, Hungary: Akadémiai Kiadó Zrt, 2006, Vols. 1 and 2. \$149.00 (ISBN 963-05-8440-9)

Scientists by nature are a contentious group, and this is reinforced when it comes to measuring their own output and productivity. For more than 50 years, bibliometrics has been widely recognized as a core method of evaluating scientific output. Early on, it was accepted that citation counts, journal impact factors, and variations on these can more or less objectively measure the research quality of scientists and their departments and institutions. For nearly 30 years, the journal *Scientometrics* has been a forum for the systematic inquiry into such measurement, and as this collection shows, the debate is far from settled. Published in Hungary, the journal has now become a mainstream title with over 1,700 articles in its 30-year history. Original Editors-in-chief included Eugene Garfield and Derek de Solla Price.

Current members of the editorial and advisory board of the journal have selected 66 papers by 90 authors and compiled them in this first volume of the *Scientometrics Guidebook Series*. The book is intended to inform the debate on assessment of scientific research through bibliometrics. Papers come from well-known authors in the field and include multiple contributions from Garfield, Anthony F.J. Van Raan, and Henrik Moed, and others. Several authors are affiliated with the Hungarian Academy of Sciences, including Editor and contributor Tibor Braun, who has been a scientist himself since receiving his degree in chemistry in the 1950s. In the latter half of his career, he became interested in indicators of scientific achievement, and his *Scientometric Indicators: A 32 Country Comparative Evaluation of Publishing Performance and Citation Impact* (World Scientific, 1985) could be considered the mark of his transition from part-time to full-time scientometrician.

The book is divided into five sections: General framework, Methodological tools, Evaluation at departmental and institutional level, Evaluation at individual level, and Discussion and debate. The methodological and the evaluation sections are the largest, with many of the evaluation papers summarizing studies conducted at scientific organizations at the national or university level. The discussion and debate section, on the other hand, is largely a collection of opinion pieces, letters, and their responses dealing with the validity and use of bibliometrics in science.

Many of the contributions in this 1,000+ page work include statistical formulae and sample data comparing different departments, individuals, or organizations. In addition to the historical context, the book also offers plenty of technical material for the bibliometrician to ponder. Earlier works in this field, *The Web of Knowledge: Festschrift in honor of Eugene Garfield* (Information Today, 2000), and H.F. Moed's *Evaluating Scientific Research* (Springer, 2005) summarized the issues adequately with the *Festschrift* including some statistical formulae and examples. Moed's book is more general in purpose, but because it is more recent, he does devote some space to important and exciting technological trends.

Most of the book consists of studies that propose, explain, and/or justify certain statistical methods. Nearly all bibliometric calculations discussed depend on the Science Citation Index (SCI) for raw data.

* This is a U.S. Government work and, as such, is in the public domain in the United States of America.

© 2008 Wiley Periodicals, Inc.

This is noted throughout the book as a major limitation to scientometrics to date. Coverage limitations and inconsistent forms of entry in the database, many authors argue, place the results of bibliometric analysis in question even before the data are gathered.

In the book's overview, several papers by Braun and András Schubert lay the groundwork for more refined measurement techniques later elaborated on in the evaluation and methodology sections; however, they maintain that "mere publication or citation counts are completely inadequate measures of scientific merit; they can be used for evaluative purposes only after proper standardization or normalization" (p. xx). This is the first of repeated assertions throughout the book that such measurements should not be solely relied upon. Indeed, an editorial from the very first issue of *Scientometrics* states that quantitative methods should always be followed by qualitative analysis. Multiple contributors to the book assert that other institutionalized methods for assessing research output can be supplemented, but certainly not replaced by the use of citation counts and that derived indicators should be taken with other qualitative forms of evaluation, such as peer review. Several researchers attempt to correlate or validate peer review with citation statistics, most showing a positive relationship. Anthony Van Raan, a frequent contributor to the journal, asserts that peer review is strongly correlated with bibliometrics assessment using the method that he and his coauthors describe. Several others in this book concur. This suggests that for evaluation and review committees, it may be useful to select the formula that shows the strongest correlation to the peer-review system in the specific discipline in question.

The introductory section also traces the development of specific statistical indicators including the Relative Citation Rate (RCR), Mean-Observed and Mean-Expected citation rates, and later, the H-index. Each of these was created in an attempt to overcome some inherent disparity in comparing citation counts among disciplines, departments, or individuals. Schubert and Braun's paper on RCR includes formulae for calculating error rates as well as a table listing the RCR for 25 countries as of 1986. The authors note that several of these statistics were made possible only after the SCI introduced new features of search and retrieval that allowed for more elaborate comparison and evaluation of results.

The section on evaluation of departments and institutions includes studies evaluating publications from scientists at the Max Plank Institute, CERN, and national research bodies and universities. Several include funding as a variable to demonstrate some relationship to research quality.

The widespread pessimism about the use of the SCI for accurate citation analysis becomes clear early in the book. Ronald Kostoff, for example asserts that determining a credible control group—an essential step in scientometrics evaluation—requires a manual evaluation of many base papers and the expertise of others in the field. He writes, "... the automated assembly-line approaches to citation analysis, widely used by the decision-aid community today, are highly uncertain at low-to-mid citation levels characteristic of most research" (p. xx). Several other papers trace the development of baselines or data normalization methods, which are required due to the variability in scientific disciplines with regard to collaboration, publishing patterns, norms, and habits. Cross-field normalization studies give rise to consideration of "maximum potential citations" and "relative subfield impact." Several agree that extreme caution is in order with end-user availability of the SCI. The caveats are so abundant throughout the book that the reader begins to doubt the validity of any bibliometric assessment at all.

Along with the technical shortcomings of the database, several author behaviors are noted as also undermining bibliometric methods. Among these are the tendencies of authors toward self-citation, vanity citation, reciprocal (payback) citation, citing upwards (i.e., toward high impact journals), the coauthorship explosion, and the winner-take-all nature of high impact papers. These and other limitations have drawn the attention of research-assessment bodies in Europe and in the English-speaking world in recent years and prompted serious inquiry.

Further, many in the academic community have asked whether the scholarly periodical in its current form will survive. A system where authors are unpaid but journal subscriptions commonly run into the thousands of dollars has led many to this academic giveaway and subsequently to reexamine the current system of research evaluation and assessment. Since almost anyone with a Web site now can be editor and publisher, one response has been to free up this content and make it available at no cost to the reader. Some authors say that bibliometrics will suffer in such a publishing free-for-all while librarians and many of their reader/scientists respond by asking whether traditional methods are still adequate to measure the accomplishment of scientific research. Whatever the solution to this dilemma, it is clear that research assessment through bibliometrics will have to change as technologies change.

Several experiments with open access and what is sometimes called "open peer review" have yielded mixed results, but what is true is that at least among a growing minority, the emphasis placed on journal prestige or citation counts is increasingly viewed as misplaced. For these and many other reasons, the scholarly community has begun to revisit the entire system, including the components of bibliometrics as a tool for research assessment.

As for the book under review, one has to wonder whether those involved in science policy would find it helpful. University and other research-organization officials may be satisfied with journal impact factor, science indicators, and cited reference counts which are readily available from the Institute for Scientific Information, although warnings about their limitations may not be heeded. In the book's final paper, Peter Weingart warns that the wide availability of the SCI has perhaps resulted in the irresponsible use of bibliometric data by policy and administrative types. He states, "Any superficial attempt at interpreting such rankings without the assistance from experts that know how these numbers are created in the first place, and what they represent, is in the context of policy decisions misleading, meaningless and irresponsible" (p. xx).

Scientists are often frustrated by journalists who are forced to boil down research into bite-sized headlines that ignore the nuances of the findings. Newspaper readers want to know, for example, whether the earth is getting warmer, but do not want to be bothered with details of nitrogen levels in Greenland or the salinity of the oceans, and so on. Ironically, one could probably spot a parallel tendency among science policy makers and review panels who may be too busy to concern themselves with a more fine-tuned method of research evaluation, particularly since many methods of measurement are controversial and contentious. They ask how many times a researcher has published and how many times his or her work was cited, but probably do not want to bother with exceptions, scaling, and the publishing idiosyncrasies of certain subdisciplines. Eugene Garfield notes in the book's preface that the occasional article in *Science* or *Nature* regarding bibliometric indicators is probably as far as most busy provosts, deans, and other officials get regarding innovations in scientometrics. Indeed, most scientists themselves probably lack the spare time or energy to concern themselves with the bibliometric alternatives that this book presents. In a recent presentation, Henk Moed (2007) proposed that the only way to get bibliometrics accepted in the scientific community is to openly

discuss its limitations. This book certainly does that. So it is unfortunate that the message of this book, which summarizes the evolution of scientometrics so well, will probably not make it to those who need to hear it most.

It is likely that one could demonstrate a correlation between the technological advancement in citation databases and the increasing precision (or supposed precision) in bibliometrics. Several contributors indicated that refined metrics were made possible only after some advancement in the search and retrieval of the SCI database. This leads naturally to questions of what the future holds for scientometrics given emerging technologies such as data mining, digital object identifiers, unique author identifiers, and other advancements. Most of the papers in this collection are too old to comprehensively address emerging bibliometric alternatives such as CrossRef, the use of download counts, and the open-access movement—all of which will inevitably change the way science is evaluated through publication.

Also note that the humble roots of this journal are noticeable in the quality of the reproductions in the book. Papers are presented in various fonts (including at least one which appears to be generated from a dot-matrix printer) and with tables that are sometimes too small to display details legibly. If this journal had been electronically published or typeset over the years, there would no doubt have been some attempt to ensure uniformity in the text. With the cost of scanning and/or rekeying texts into digital formats falling continually, it is curious that this was not done for the reproduction of these collected papers. This is not a major problem, but it is certainly conspicuous.

The journal *Scientometrics* is held by over 130 libraries that list holdings in OCLC's WorldCat database. For those who have access to the journal, browsing the back issues may serve adequately as a primer on the subject. For those who do not, this book provides in one place a historical background to this increasingly popular field of inquiry. In the coming years, the scholarly and policy community undoubtedly will see a wider variety of statistical measures and indicators of scientific publishing available to them for research assessment purposes. This collection of papers may provide a frame of reference for those who want to more critically evaluate the proposed methods of evaluation. We should look forward to the next edition in the Guidebook series, although with hopes that presentation and readability are given higher priority by the publisher.

References

- Braun, T., Glänzel, W., & Schubert, A. (1985). *Scientometric indicators: A 32 country comparative evaluation of publishing performance and citation impact*. Singapore, Philadelphia: World Scientific.
- Gurfield, E., Crowin, B., & Atkihs, H.B. (2000). *The Web of knowledge: Festschrift in honor of Eugene Garfield*. Medford NJ: Information Today.
- Moed, H. (2007). *The use of bibliometric indicators in research evaluation and policy. Evolution of Scientific Publications: The Point of View of Scientists*. Académie des sciences. Paris, France. May 14–15, 2007. <rtsp://ccwebcast.in2p3.fr/pub/as/EPS/02.Henk.Moed.rm> Listed as Introductory Lecture. See: http://www.academie-sciences.fr/conferences/colloques/pdf/colloque_14_05_07_programme.pdf

Alvin Hutchinson
Smithsonian Institution Libraries
Washington, DC
E-mail: hutchinsona@si.edu

Published online in Wiley InterScience
 (www.interscience.wiley.com).
 DOI: 10.1002/asi.20726