EVALUATING SCHOLARSHIP AND RESEARCH IMPACT

Great Debates in Higher Education is a series of short, accessible books addressing key challenges and issues in higher education, on a national and international level. These books are not only research informed but also debate driven. They are intended to be relevant to a broad spectrum of researchers, students, and administrators in higher education and are designed to help readers pick and assess the state of higher education systems, policies, and social and economic impacts.

Published titles:

Teaching Excellence in Higher Education: Challenges, Changes, and the Teaching Excellence Framework
Amanda French and Matt O'Leary

British Universities in the Brexit Moment: Political, Economic, and Cultural Implications
Mike Finn

Sexual Violence on Campus: Power-conscious Approaches to Awareness, Prevention, and Response
Chris Linder

Higher Education, Access, and Funding: The UK in International Perspective
Sheila Riddell, Sarah Minty, Elisabet Weedon, and Susan Whittaker

Forthcoming titles:

The Marketisation of English Higher Education: A Policy Analysis of a Risk-based System
Colin McCaig

Cultural Journeys in Higher Education: Student Voices and Narratives
Jan Bamford and Lucy Pollard

Refugees in Higher Education: Debate, Discourse, and Practice
Jacqueline Stevenson and Sally Baker

Access to Success and Social Mobility through Higher Education: A Curate's Egg?
Stuart Billingham

EVALUATING SCHOLARSHIP AND RESEARCH IMPACT

History, Practices, and Policy Development

BY

Jeffrey W. Alstete
Iona College, USA

Nicholas J. Beutell
Iona College, USA

John P. Meyer
Iona College, USA

United Kingdom – North America – Japan
India – Malaysia – China

Emerald Group Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2018

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-78756-390-2 (Print)
ISBN: 978-1-78756-387-2 (Online)
ISBN: 978-1-78756-389-6 (Epub)

ISOQAR certified Management System, awarded to Emerald for adherence to Environmental standard ISO 14001:2004.

Certificate Number 1985
ISO 14001

INVESTOR IN PEOPLE

CONTENTS

ABOUT THE AUTHORS

Jeffrey W. Alstete is Professor of Management and Business Administration in the School of Business at Iona College, USA, where he teaches courses on business policy and strategy, entrepreneurship, international management, managerial decision-making, competitive intelligence, knowledge management, benchmarking, and other topics. He is Chair of the Intellectual Contributions Committee that evaluates research awards in the School of Business. He is the author of five books including: *Benchmarking in Higher Education: Adapting Best Practices to Improve Quality* (1996); *Accreditation Matters: Achieving Academic Recognition and Renewal* (2004); and *Revenue Generation Strategies: Leveraging Higher Education Resources for Increase Income* (2014).

Nicholas J. Beutell is Professor of Management and Business Administration at Iona College, USA. He has written over 100 publications. He was Dean of the School of Business for 10 years, prior Chair of the Institutional Review Board at Iona College, and has over 40 years of experience as scholar, professor, and administrator. Professor Beutell is a widely recognized and highly cited scholar on work–family, self-employment and entrepreneurship, and wellbeing.

John P. Meyer is Professor and Chair of the Management and Business Administration Department at Iona College, USA. His research includes knowledge management,

innovation, and experiential learning. He has published on intellectual capital and patent citations which, as representations of innovation quality and value, are analogous to the way scholarly citations are now being used to evaluate research quality and value. Professor Meyer earned appointment as Master Teacher in the Iona College School of Business and serves on the institution-wide Rank, Tenure, and Awards (RTA) committee. In these roles, he has experience into how measurement metrics are being used in important organizational hiring decisions and faculty career evaluation.

PREFACE

Academic research and scholarship are experiencing a paradigm shift. Research impact is the most recognizable manifestation of this shift made visible by the omnipresence of scholarly metrics. The "digitization of everything" can be lamented but it is now an everyday reality for people everywhere. If something can be measured, systematized, synthesized, and shared, it will be. Metrics, a proxy for research impact, represent this trend in the domain of academic scholarship. Articles that used to take years to be published in print now appear "online first" before volume, issue, and page numbers are even assigned. This is presumed to increase impact since work can be viewed, downloaded, and cited much more quickly than in the past. Yet, the pressure to make publications available more quickly may be missing a critical component: research quality.

Where should I present or publish my research? Is this a good journal to showcase my work? Will anybody read my article there, let alone cite it? How can I maximize the impact of my scholarly output? These are questions academics and other researchers all over the globe ask themselves and their peers every day. The answers to these and countless similar questions will affect their careers, their reputations, and, in many cases, their paychecks. At the same time, academic and research organizations are under more pressure than ever to create and implement policies that encourage quality,

high-impact research from their members through various evaluation and reward systems. Within this monograph, the authors attempt to provide some guidance to individual researchers and the institutions for which they work, as they struggle with these issues in their ongoing efforts to produce and disseminate valuable, high-quality scholarship.

Universities and research institutions are facing challenging and often conflicting influences from institutional stakeholders and the macro external environment. These forces are somewhat at odds with traditional concepts of scholarship and academic freedom. There is movement toward improved oversight of universities and research institutions along with increased internal accountability at the individual and institutional level. Another related strategic and operational influence affecting change is the nature of academic and scientific publishing in the twenty-first century. These developments suggest that entrenched scholarly traditions are facing very complex and nuanced dilemmas regarding the effective oversight of institutions while simultaneously striving to support creativity and innovation (Scott, 2018). The confluence of increased scrutiny along with easy information access have intensified debates within and outside academe; they have also manifested greater awareness and usage of quantitative analysis of research (e.g., bibliometrics, the statistical analysis of research) in policymaking.

The nature and extent of these changes need to be critically examined. The proliferation of academic research along with advances in information technology have given rise to the visibility and prominence of scholarly metrics, such as author or article citation counts, journal impact factors (JIFs), and related measures of institutional research output that are becoming more widely used (Walters, 2017). Such metrics are readily available on many websites (e.g., http://guides.library.jhu.edu/metrics), but can they be taken at face

value? Are scholarly metrics, such as citation counts and impact factors, useful? Are they reliable and valid indicators of research quality? Do these metrics effect faculty research development and career advancement? Can such metrics help to inform policy development regarding research funding, institutional evaluation, faculty/researcher recruitment, and promotion, as well as overall research strategy? These are some of the important questions that are addressed in this monograph.

The increasing recognition and publicity about scholarly metrics, whether for individuals, authors, articles, departments, or entire institutions, may seem to be just another "number" that needs to be achieved. Proper understanding of the history, recent developments, and possible future trends in scholarly metrics, citation analysis, and publication outlet quality is needed for policymakers to craft cautious yet informed use of these temptingly easy-to-use metrics. While librarians and information scientists have been evaluating journals for at least 75 years, the increasing convenience of automatically tabulated scholarship measurements is now being applied far beyond their original intention. Gross and Gross first conducted a classic study of citation patterns in the 1920s (Gross and Gross, 1927). Other developments followed, such as Estelle Brodman's studies in the 1940s of physiology journals along with subsequent reviews of the process of journal evaluation (Brodman, 1944). The introduction of the Thomson Reuters citation indices then enabled computer-compiled statistical reports to not only do more than tabulate journal articles but also to calculate citation frequency.

Eugene Garfield first mentioned the idea of an impact factor in 1955 (Garfield, 1955). This led to the 1961 publication of the Science Citation Index® (SCI) (Garfield and Sher, 1963) and the JIF to help libraries select additional source

journals. This was conducted by re-sorting the author citation index into the journal citation index. The purpose was to determine highly cited journals that need to be covered in the new SCI®, nothing more. Following the in-house use of journal statistical data to compute the SCI, Thomson Reuters began to publish Journal Citation Reports® (JCR) in 1975 as part of the SCI and the Social Sciences Citation Index® (SSCI). Many other author, article, and journal-level metrics were subsequently developed and continue to evolve and be used today such as Google Scholar citation totals, the Hirsch (*h*) Index, Journal Citation Reports, SCImago Journal Rank, and others.

Thoughtful and cautious use of impact data is important to consider. However, because of their widespread availability, users may be tempted to jump to improper conclusions based on impact factor statistics unless several caveats are contemplated. The various metrics provide quantitative tools for ranking, evaluating, categorizing, and comparing journals and articles. The impact factor is one of these; it measures the frequency with which the "average article" in a journal has been cited in a particular year or period. However, it does not account for various common statistical and other sources of error (such as skewness, bias, self-citations, and other influences). Yet, the impact factor can be useful in clarifying the significance of absolute (or total) citation frequencies. It can be tabulated to remove some of the bias that may favor large journals over small ones, more frequently issued journals over those published less often, or even older journals over newer ones (Garfield, 2012).

There have been many innovative applications of JIFs. Traditional usages involve market research for publishers and as a tool for librarians in their attempts to manage library journal collections. More recently, however, JIFs have taken a different turn. They have quickly become a fast and

convenient metric for evaluating individuals, departments, and institutions. In fact, the founding father of scholarly metrics, Eugene Garfield, was concerned about these and other unintended consequences of his creation. Garfield believed that scholarly metrics might provide a gross approximation of the prestige of academic journals in which individuals have been published. He argued that metrics should be used in conjunction with peer review, overall productivity, and area of academic specialization. The authors of this monograph agree with this assessment.

With respect to faculty tenure and promotion decisions, it is not appropriate to rely solely on the reported metric impact of a journal as a proxy for quality and academic impact of the journal itself. This also applies to any individual articles published in the journal, as well as any authors of those articles. It would be more accurate, professional, and holistic to use the impact factor(s) combined with informed peer review. Furthermore, citation frequencies for specific articles are quite varied among individuals, fields of study, departments, and institutions as a function of their differing missions and institutional characteristics. Some important scholarly work may take many years to develop and publish, additional years to be recognized, and even longer to be cited by others (Kozak, 2013). Such factors should be considered in any tenure and promotion processes that are based in part on research productivity and impact. When rendering important academic and institutional decisions, metrics can and should be considered as contributing to the process; they are not sufficient at present, however, for use in isolation without expert input from human reviewers who understand their limitations, complicated nuances, and intended purposes.

Among additional issues that should give the academic community pause when considering the value of quantitative measures of research value is the fact that a journal's ranking

and an individual scholar's metrics can be affected by the inclusion of such items as review articles or letters. For example, review articles appear to be cited more frequently than typical research articles because they often serve as surrogates for earlier literature. Review journals have some of the highest impact factors when compared to other types of scholarly publications and journals that have a combination of original research articles, and review articles have an advantage in metric tabulation. Other complicating factors in raw article and journal citation numbers are redundant publications in journals, inaccurate tabulations based on similar names, and purposeful manipulation or "gaming" of the metric systems by authors and editors to inflate desirable metric numbers.

To further confound policymaking, it is believed that research method articles tend to attract more citations than other types of articles, yet this is not necessarily the case. In fact, many journals dedicated completely to methods research do not attain unusually high-impact numbers (Elliott, 2014; Seglen, 1997). This mistaken assumption may stem from the fact that some of the most highly cited articles are seminal classics that belie the reputation of the journal. Such journals may not necessarily contain more influential articles than other journals. Decisions about policy formulation and decision-making about hiring, promotion, tenure, reduced teaching loads for research and research funding that are based on scholarly metrics must account for and be aware of these important matters, else improper and unjust evaluations will occur. Readers and naïve users of scholarly metrics need to become more aware of the concerns and pitfalls of using metrics. For example, the chronological limitation of some impact calculations, such as 2- and 5-year rolling periods, is intended to remove the partiality that major breakthrough pieces might produce. Total citation frequencies are influenced in this way, and perhaps that may be important for

some policies and decisions but not necessarily for others. There are also variations between disciplines with different ranges of maximum impact, as well as metric influences affected by the number of item types in journals (original research, reviews, and letters), and certain specified-only journal measures such as the JCR that may involve suspended journals, superseded titles, or journals that have ceased publishing.

The ranking of journals and associated journal metrics are certainly controversial. Many papers written about the strengths and weakness of such metrics identify concerns about what metrics actually measure. The concerns are quite common and repeatedly stated and subsequently have become more urgent in light of current developments in the scholarly academic world. Until fairly recently, it was common practice for academic researchers to concentrate their reading on a limited number of high-quality publications. Subsequently, as the number of research outlets proliferated, researchers were given immediate access to a vast array of journals (print and online) such that less attention was paid to the quality of the outlets. This contributed to an increased reliance on metrics. Scholarly metrics are easy to retrieve and use, supposedly serving as a proxy for quality of content and outlet. Unfortunately, there are many cases of high-quality works that have low citation counts, are published in lower ranked journals, and may be completely overlooked or delayed with respect to advancing knowledge in the discipline (Kozak, 2013).

The current system of scholarly peer-reviewed journals has simultaneously grown, come under increased scrutiny and criticism (for reliability, fairness and validity), yet has also become increasingly relied on. It is an unusual time in higher education, academic scholarship, and funded research. However, a review of the research literature concludes that

journal peer review is valid and still does, in fact, function as a quality filter (Daniel, 2005). Yet, there is a lot of ambiguity in the literature as well. Prepublication peer review could be supplemented with postpublication evaluation to help determine which publications and scientists have contributed most to knowledge advancement in a given field and remove or identify inaccuracies.

The legitimacy of publications and metrics must also be scrutinized carefully and then only be considered as one aspect of quality when implementing policies involving recognition and funding. Holistic policy development is recommended because it is crucial to know what quantitative research analysis can provide to researchers and policymakers and what such measures are unable to deliver. The consensus among many analysts, users, agencies, and faculty researchers is that even the most well-developed scholarly metrics that attempt to account for differences in discipline, age of publication, and other factors are unlikely to be a substitute for human judgment (Reuters, 2016).

In this monograph, a conceptual framework is proposed for using research evaluation methods for developing policies to promote and reward quality research that includes consideration of: research purpose, outputs, forms, funding, and institutional type. Colleges, universities, research institutions, and external funding agencies are currently struggling with questions surrounding how to consider scholarly performance evaluation accurately and fairly. There appears to be a rational understanding that publishing in highly ranked journals, such as those included on various lists of "quality" journals or those that have high-impact factors, does not necessarily equate to true value of the underlying research. Furthermore, when considering fundamental issues such as academic freedom, policymakers should identify the objectives of research publication based on their goals and institution type.

If the goal of the institution and its researchers is to supply new knowledge to the core and foundation of theories in academic disciplines, the policy might seek, encourage, and acknowledge only research contributions accepted in a select group of high-quality publications. Yet, if the goal is practical understanding of a field or maintaining knowledge currency by faculty members, then policies may encourage and acknowledge research contributions accepted in a broader number and type of publications as opposed to only "A-list" journals. In these latter cases, it may also be suitable to recognize publications outside the faculty member's core discipline or traditional field of research. Finally, it is understood that influential stakeholders, including tenure and promotion committees, academic policymakers, funding agencies, libraries, and so on, are not likely to diminish their use of various journal lists and research metrics.

The trend in research evaluation is toward a balanced, hybrid approach that recognizes the value of different publications by examining whether an author, a particular piece of research output (e.g., published article), or an institution as a whole, meets appropriate standards for authors, articles, and journals in their specific research field and type of institution as measured, in part, by certain metric numbers. Further, consideration is also given to whether the work is being cited both within and beyond one's core discipline, by practitioners as well as other academics, bridge media such as professional magazines, and other outlets.

This monograph supports the general philosophy behind the use of hybrid approaches to research evaluation. Organizational policies must be developed with thoughtful examination of the policy objectives, limitations of the measurement systems, differences in disciplines, and institutional types. The use of scholarly metrics has moved beyond philosophical debates about their appropriateness. Scholarly

metrics are already widely used in policy formation and implementation of operational decisions. It is argued that some combination of metrics, perhaps involving citation tabulations, metrics such as *h*-indices and others described later in this monograph, JIFs, and other measures should be combined and used to aid and supplement individual expert analysis (i.e., peer review). In this way, scholarly work can be properly acknowledged, recognized, evaluated, and rewarded in a more holistic manner. There is a need to evaluate such combinations of factors along with a continually evolving approach that considers newly developing metrics. Based on the many concerns about the validity of bibliometric analysis and the use of such tools for evaluating people and research, policies should be developed in light of the agreed upon goals of different organizational processes (e.g., hiring, promotion, tenure, allocation of teaching loads, research grant funding) as well as institutional type.

The aforementioned variable of institutional category is a major consideration in the use of scholarly metrics in policy formulation. For institutions such as 2-year community colleges, the important elements of research geared toward student skill development as identified by Fisher (2009) may be appropriate. For many traditional 4-year institutions, a more holistic and hybrid analysis of intellectual contributions may be more effective. This analysis could include citation counts, scholarly metrics (such as the popular *h*-index, or *h*-index, and its many variants), altmetrics (nontraditional metrics) involving social media, as well as input beyond pure publication and citation activity including impact-generating events that scholarly writers have conducted for constituencies such as undergraduate students, graduate students, faculty, the college or university-at-large, professional societies, the external communities, and others. For large prominently research-oriented universities and research institutions, it may

be more appropriate to place a stronger emphasis on established metrics to quantitatively assess impact of intellectual activities along with expert reviews to establish overall impact.

A common theme synthesized from the literature and current policy practices is that academia and research institution policies should respect and encourage more than publications in certain top journals (Lee, 2014). There should be broader respect for, and acknowledgment of, different types of scholarship such as practice-oriented papers, research-in-progress, book and media reviews, responses to previously published articles, book chapters, conference proceedings, and other intellectual contributions that are often not tabulated in many of the commonly used scholarly metrics. The goals of policies governing intellectual contributions should depend on the institution type, mission, and researcher capabilities so as to open new frontiers of knowledge generation, as well as improvements in teaching, learning, and recognizes the application or integration of knowledge. Therefore, hybrid evaluation approaches that combine quantitative scholarly metrics with more qualitative individual expert analysis are often suggested as the most fair and accurate method to gauge the true impact of scholarly output and measure the ultimate value of the underlying research activity. It is with this goal in mind that this monograph is presented as a guide for the development of more holistic approaches to the measurement and evaluation of scholarship production and research impact.