Research Policy 49 (2020) 103831

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

How to avoid borrowed plumes in academia

Margit Osterloh*, Bruno S. Frey

University of Zurich and University of Basel and CREMA (Center for Research in Management, Economics and the Arts, Suedstrasse 11, 8008 Zuerich, Switzerland

ARTICLE INFO

Keywords:
Journal rankings
Impact factor
Journal quality lists
Skewed citation distribution
Focal random selection

ABSTRACT

Publications in top journals today have a powerful influence on academic careers although there is much criticism of using journal rankings to evaluate individual articles. We ask why this practice of performance evaluation is still so influential. We suggest this is the case because a majority of authors benefit from the present system due to the extreme skewness of citation distributions. "Performance paradox" effects aggravate the problem. Three extant suggestions for reforming performance management are critically discussed. We advance a new proposal based on the insight that fundamental uncertainty is symptomatic for scholarly work. It suggests focal randomization using a rationally founded and well-orchestrated procedure.

1. Introduction

Publication in peer-reviewed scholarly journals has today become the currency of performance for the evaluation of scholars, departments, faculties, and universities. Journals are ranked according to quality criteria, most importantly the journal impact factor. It is defined as the mean number of citations in a particular year of articles published in that journal in the previous two years or five years. Some journals are ranked according to journal quality lists, such as the Association of Business Schools (ABS) Guide in Great Britain (e.g. Mingers and Willmott, 2013) and the "Top Five" in economics (e.g. Hamermesh, 2018).[1] It has been empirically demonstrated that the "Top Five" have a powerful influence on tenure and promotion decisions and has even been denounced as the "tyranny of the top five" by a Nobel Prize laureate (Heckman and Moktan, 2018). Journal quality lists rely not only on journal metrics but also on qualitatively informed indicators of reputation. In both cases, the quality of a journal is widely believed to reflect the quality of any article published therein. Originally designed to evaluate scientific journals, today journal quality lists and impact factors are increasingly used to evaluate individual articles and authors. They strongly influence decisions on tenure, research funding, and the pursuit of career goals. For example, the British ABS Academic Journal Guide claims to give scholars "a recognized currency on which career progress can be based" (ABS The Association of Business Schools ABS, 2015: 5). In many academic institutions, scholars receive a financial bonus for a publication in one of the top journals (e.g. Fuyuno and Cyranoski, 2006; Macdonald and Kam, 2007; Shao and Shen, 2011).

However, this practice has been strongly criticized for several years (Seglen, 1997; Moed and Van Leeuwen, 1996; Laband and Tollison, 2003; Starbuck, 2005; Oswald, 2007; Singh et al., 2007; Adler and Harzing, 2009; Frey and Rost, 2010; Baum, 2011; Macdonald and Kam, 2011; Mingers and Willmott, 2013; Alberts, 2013; Osterloh and Frey, 2014; Wilsdon et al., 2015; Martin, 2016; Larivière et al., 2016; Berg, 2016; Callaway, 2016; Waltman, 2016; Wang et al., 2017), even by Eugene Garfield, the inventor of the impact factor (Garfield, 1973). The San Francisco Declaration on Research Assessment (DORA (San Francisco Declaration on Research Assessment) and DORA, 2012), which has been endorsed by many leading institutions, clearly states: "Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions." The recently released "Statement by three national academies (Académie des Sciences, Leopoldina and Royal Society) on good practice in the evaluation of researchers and research programmes"[2] also asserts that "[i]mpact factors of journals should not be considered in evaluating research outputs". Nevertheless, to date, these critiques have not diminished the impact of either impact factors or journal quality lists. Instead, journal rankings have become more widespread and increasingly important for academic careers and research funding (e.g. Harzing, 2015; Martin, 2016; Vogel et al., 2017). Top-tier journals have become the ultimate fetish token (Willmott, 2011) for many scholars. According to a survey of the perceptions of young economists the pursuit of top journal publications "has become

* Corresponding author.
E-mail addresses: margit.osterloh@business.uzh.ch (M. Osterloh), bruno.frey@bsfrey.ch (B.S. Frey).

[1] See also the *Handelsblatt* Ranking in Germany http://www.handelsblatt.com/politik/konjunktur/vwl-ranking/.

[2] https://www.leopoldina.org/de/publikationen/detailansicht/publication/good-practice-in-the-evaluation-of-researchers-and-research-programmes-2017/

https://doi.org/10.1016/j.respol.2019.103831
Received 8 May 2018; Received in revised form 11 July 2019; Accepted 16 July 2019

the obsession of the next generation" (Heckman and Moktan, 2018: 1).

This paper has two aims. The first is to understand why impact factors and journal lists are still so influential to evaluate individual papers even though they are strongly criticized by many influential scholars and institutions. This criticism is based on the heavily skewed distribution of citations in scholarly journals. Why are impact factors and journal lists not abolished as proxies for the quality of single articles? Second, while the criticisms of this practice are many, few suggestions have been made for changes at the institutional level to overcome the problem. We discuss such proposals and present a novel, radical proposition: purposeful focal randomization. To our knowledge, this is the first proposal for change using the insight that uncertainty is fundamental to research, translating it into performance management.

The second section of this paper complements the literature that questions the use of impact factors and journal quality lists to evaluate individual articles because of the strong skewness of citations in scholarly journals. We ask whether the citation rates of articles accumulated over five years are more useful in evaluating publications than yearly citation rates. We show empirically that this is not the case. There is still a substantial overlap in the distribution of citations between high-, middle- and low-ranked business journals. In the third section, we inquire why impact factors and journal quality lists have not been abolished even though they have attracted such strong criticism. We argue that this is mainly due to the fact that the majority of authors benefits from journal quality lists, which is aggravated by the "performance paradox" and lock-in effects. In the fourth section, we discuss proposals on how the present unsatisfactory situation can be overcome by changes at the institutional level. We present and discuss our own proposal.

2. Skewed distributions of citations

The use of journal lists to evaluate the quality of research – whether derived from metrics or qualitatively-informed indicators - takes for granted that publishing in a "good journal" is a signal of "good research". The most influential journal rankings today rely largely on the two-year journal impact factor (JIF) published by Clarivate Analytics (formerly Thomson Reuters), which owns and publishes the Journal Citation Reports (formerly known as the ISI Web of Knowledge).[3] The JIF was originally developed to help librarians identify the most important journals (see Archambault and Larivière, 2009) according to the numbers of citations of the articles published in those journals.

The use of citation counts as a performance indicator has its own problems (e.g. Starbuck, 2005; Adler and Harzing, 2009; Macdonald and Kam, 2010). To take citations as a proxy for quality is questionable. At best it can inform us whether an article can be considered interesting and influential since citations acknowledge the impact an author has on the work of others (e.g. Antonakis et al., 2014; Alvesson and Sandberg, 2013; Hamermesh, 2018). Nevertheless, citations are widely accepted as a performance indicator for articles and journals (e.g. Goodall, 2009; Vogel et al., 2017), though most scholars agree they should not be used as the only determinant.[4] However, those who use impact factors for an article or a journal – be it as a proxy for quality or for other reasons –must ex ante have accepted that citations matter, because impact factors are based on citations.

It is questionable using the impact factor as a quality indicator for a whole journal, but it is a clear misuse employing the impact factor of a journal as a quality indicator for a *single* article in that journal. This is due to the highly skewed distribution of citations.[5] Nevertheless, such misuse has not decreased (e.g. Heckman and Moktan, 2018), although an increasing number of studies argues that scholars should abolish it.

An impressive example of the misuse of impact factors was published recently in *Nature* (Callaway, 2016). This article refers to a study considering the natural sciences (Larivière et al., 2016), which reveals that 74.8 percent of the articles published in *Nature* (2015) were cited below the 2-year impact factor of 38.1, which reflects the average number of citations for articles in that journal. The most cited paper was referenced 905 times. Three quarters of authors benefit from the minority of authors with many citations. The equally renowned journal *Science* shows almost the same result: 75.5% of the papers published in 2015 garnered less than the impact factor of 34.7. The most successful paper was cited 694 times.

A similar pattern was demonstrated earlier in the field of organization and management by Baum (2011). He examined five journals[6] and collected the citations per year in 2008 of articles published from 1990 to 2007. He concludes that the impact factor has little credibility as a proxy for the quality of an article published in these journals. Using the JIF in such a way results in incorrect attribution of article quality more than half the time. Only a small correlation was found between the number of citations for an individual article and the impact factor of the publishing journal. Baum (2011) firmly recommends that we need to stop this misuse.

Many other influential scholars[7] and academic institutions have banned the use of JIFs as proxy for the quality of a single article, notably the International Mathematical Union (2008), the San Francisco Declaration on Research Assessment (DORA (San Francisco Declaration on Research Assessment) and DORA, 2012), the Leiden Manifesto (Hicks et al., 2015), and the Metric Tide report (Wilsdon et al., 2015).

Yearly citation rates and short-term citation windows might be too narrow to evaluate the impact of articles measured by citations. Annual citation rates typically peak after three to five years (International Mathematical Union [IMU], 2008: 7; Mingers, 2008).[8] Perhaps the accumulation of citations across several years shows a less skewed distribution; this might justify evaluating individual articles by the journal in which they were published. Therefore, we undertake a citation analysis of individual articles and use cumulative citations per article over a five-year period, starting in the second year after publication. In contrast to the five-year Journal Impact Factor, we do not consider citations in the year immediately after publishing, because there is typically a citation lag. Instead, we take all articles published in 2010 in nine management journals and add all citations gained per article during the five years from 2012 to 2016. By doing so, we avoid the weakness of short citation windows (Martin, 2016) that favor "shooting stars" over "sleeping beauties" (Mingers, 2008). However, the period is short enough to avoid significant general changes in citation behavior.[9] We take into account three top-tier journals: *The*

[3] For a review of the literature on different citation impact indicators see Waltman (2016).

[4] See e.g. the extensive model for evaluating research quality by Martenson et al. (2016).

[5] In addition, many other criticisms have been leveled at the robustness of the

(*footnote continued*)
journal impact factor, such as that JIFs are field specific, vary with the type of paper, include self-citations, can be manipulated, and are calculated from data that are neither transparent nor openly available to the public; see Martin (2015; 2016).

[6] *Academy of Management Journal*, *Administrative Science Quarterly*, *Organization Science*, *Journal of Management Studies*, and *Organization Studies*.

[7] See most prominently the panel discussion among five famous economists (Georges Akerlof, Angus Deaton, Drew Fudenberg, Lars Hansen, James Heckman), among them four Nobel Prize laureates, at the American Economic Association Annual Meeting January 7, 2017 in Chicago on "Publishing and Promotion in Economics: The Curse of the Top Five", https://www.youtube.com/watch?v=PqdKMQNXM2A.

[8] Conversely, it has been shown that articles that are not cited within five years are unlikely to be remembered later (Gittelman and Kogut, 2003).

[9] Citation practices have evolved over time. Citations per article approximately doubled between 1980 and 2004 (see Wallace et al., 2009). In management journals, impact factors have evolved accordingly, see e.g. Walsh