Margit Osterloh and Bruno S. Frey 28. 01. 19
(appears in Research Policy)

How to avoid borrowed plumes in academia

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.

Keywords

Journal rankings, impact factor, journal quality lists, skewed citation distribution, focal random selection.

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, 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,

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

1997; Moed and Van Leeuwen, 1996, Laband and Tollison, 2003; Starbuck, 2005; Oswald, 2007; Singh, Haddad, and Chow, 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), even by Eugene Garfield, the inventor of the impact factor (Garfield, 1973). The San Francisco Declaration on Research Assessment (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 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

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

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 2008; 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

[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).

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

---

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

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, 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

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, Larivière, and Gingras, 2009). In management journals, impact factors have evolved accordingly, see e.g. Walsh (2011). This problem arises when considering Oswald's (2007) study, which analyzed the cumulative citations of articles in six journals in economics across 25 years. He found that five articles in two top journals had not been cited once during that time, whereas some articles in lower-ranked journals were cited 43 to 199 times. See also Antonakis et al. (2014). They found that 7 percent of all articles published in *The Leadership Quarterly* from 1990 to 2012 were never cited.

take into account three top-tier journals: *The Academy of Management Review* (*AMR*), *The Journal of Management* (*JM*), and *The Academy of Management Journal* (*AMJ*), which take the first three positions out of 121 ranked by impact factor in the Business category in 2017.[10] As a comparison, we analyze three middle-ranked Journals (ranked 49 to 51): *Research-Technology Management (RTM), Small Business Economics (SBE)* and *Journal of Engineering and Technology Management (JET-M),*[11] and three low-tier journals (ranked 99 to 101): *The Asia Pacific Business Review* (*APBR*), *The Journal of Business Economics and Management* (*JBEM*), and *Organization Dynamics* (*OD*).[12] We count the citations of all 348 articles published in these journals in 2010 from 2012 up to 2016.

Figures 1 and 2 show the number of articles published in these journals in 2010, the number of citations over the five-year period 2012-2016, the citations per article, and the average number of citations per article. Table 1 in the appendix shows the statistics.

---

[10] The two-year impact factors of these journals in 2017 are 9.4, 7.7, and 7.4, respectively.
[11] The two-year impact factors of these journals in 2017 are 1,796, 2.857 and 2.686, respectively.
[12] The two-year impact factors of these journals in 2017 are 1.0, 0.97, and 0.93, respectively.

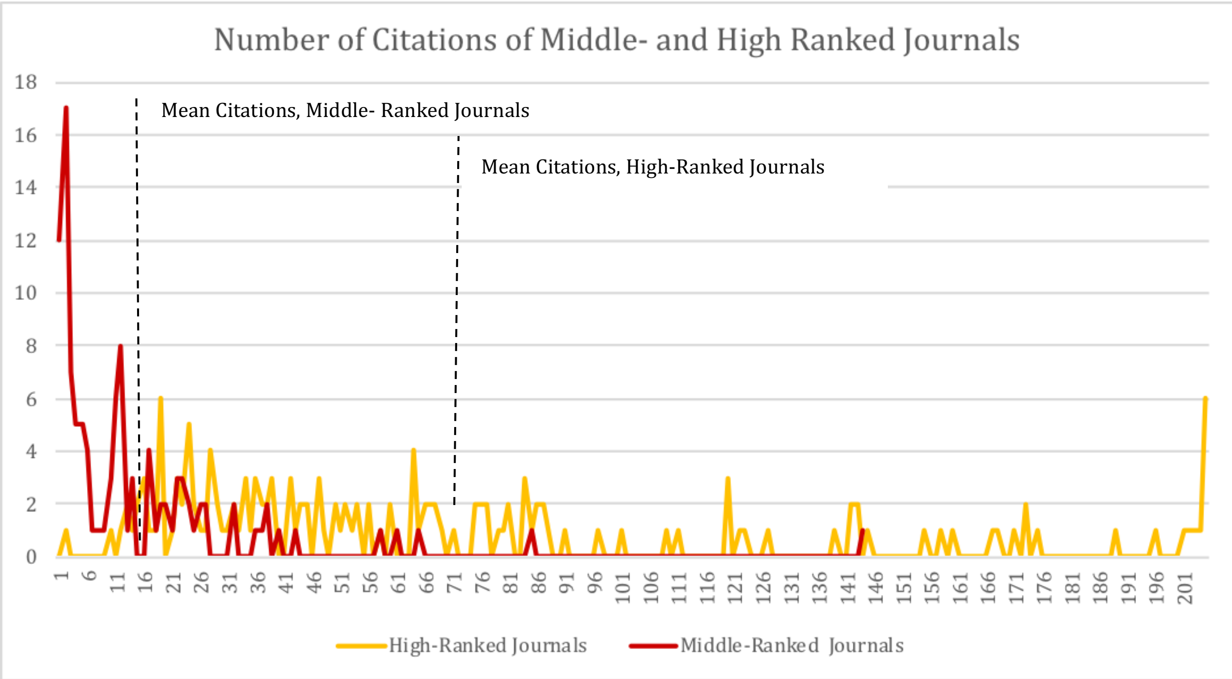

Figure 1: Distribution of Citations in Middle Ranked Journals (red) and in High-Ranked Journals (yellow)

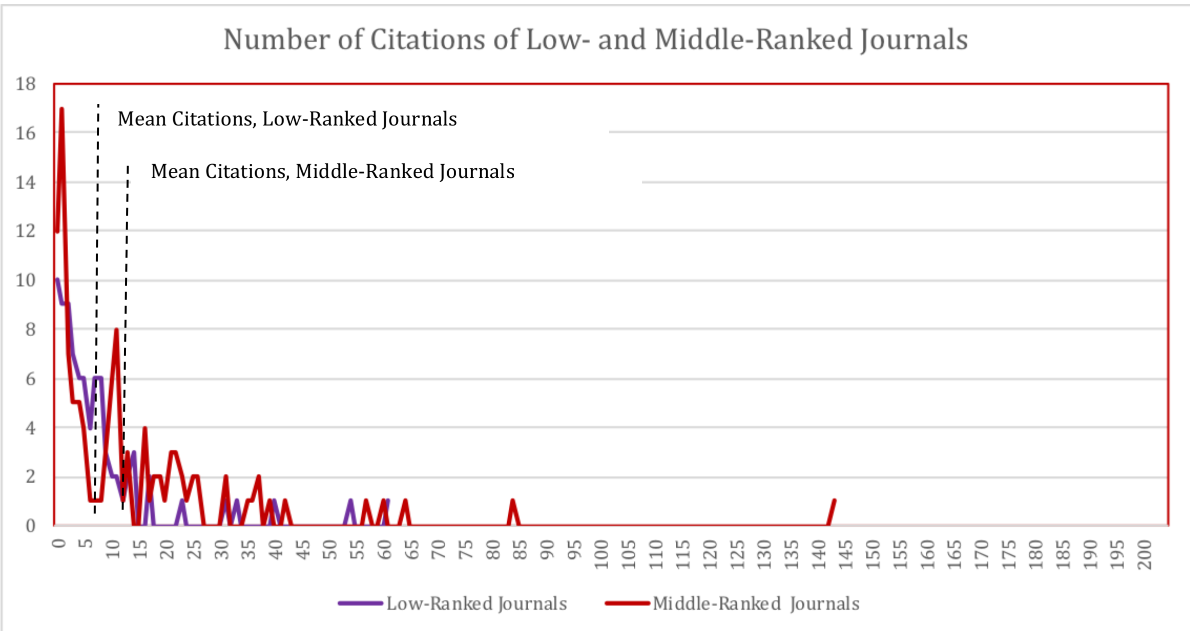

Figure 2: Distribution of Citations in Low Ranked Journals (red) and in Middle-Ranked Journals (blue)

In Figure 1 the yellow line indicates the citation patterns of the high-ranked journals *AMR*, *JM*, and *AMJ*, comprising 149 articles and 10,294 citations. They reveal that there is still a strong skewness and a long tail of the distribution, even when we consider cumulative citations across five years starting with the second year after publication. The most cited article draws 314 citations, more than four times the average citation rate of 69. A large majority of contributions—no less than 64.4%—are cited below average.

The red line indicates the citation pattern of the middle-ranked journals *RTM, SBE*, and *JET-M*. In total, in these journals 110 articles have been cited 1505 times. This distribution is also skewed due to the fact that 12 articles have not been cited at all, but one single article has been cited 144 times. The average number of citations is 13.7; 67.3% of the articles are cited less than the average.

In Figure 2, the red line reproduces the citation patterns of the middle-ranked journals (as in Figure 1). The blue line indicates the distribution of the 84 articles and 641 citations in the low-ranked journals *APBR*, *JBEM,* and *OD*. The citations are also strongly skewed and have a long tail. Of course, the number of citations is much lower than in the high- and middle- ranked journals; the average number of citations being 7.6. Five articles are cited more than 30 times, the maximum is 61. In this group, 65.5% of the articles are cited less than the average.

There is a considerable overlap in the citation distributions between the high-, middle- and low-ranked journals. The least cited article in *AMR* received 15 citations, in *AMJ* 12 citations, and in *JM* 1 citation. To attribute an article that receives 143 citations in a middle-ranked journal (or 61 citations in a low-ranked journal) to be less important than an article cited 1, 12 or 15 times in a high-ranked journal is questionable. One could even argue that being cited from a middle or low-ranked journal has to be valued more highly than being cited from a top journal, since it is harder to be noticed in a low-impact journal (Balaban, 2012).[13]

To sum up, many articles whose frequency of citation is high were published in less well-ranked journals, and vice versa. As we have demonstrated, this is not only true for short-window citations, but also with cumulative citations across five years starting with the second year after publication. Therefore, it is highly problematic to equate publication in “good” academic journals with “good” research and to consider publication in low-ranked journals automatically as signifying less good research.[14]

3. Why are journal rankings still so influential?

Despite the strong criticism, many scholars believe in journal rankings and have even internalized them as part of their identity (Alvesson and Sandberg, 2013). Publishing in a high-impact journal has become far more important than the content of research (e.g. Frey, 2009; Mingers and Willmott, 2013). This might be why the reward center in the brain of authors is activated when they expect a publication in a top journal (Paulus et al., 2015).

[13] This does not mean that we agree with the assumption that high citation rates are a measure of scholarly quality. Instead, we intend to demonstrate that if one adheres to impact factors one has agreed ex ante on citation as a proxy of quality.

[14] We concentrate on journal rankings according to the JIF. Other kinds of journal list such as the British ABS list and the h-index for journals might lead to different journal rankings. In particular the h-index for journals provides a more accurate measure of journal quality than JIF (Harzing and van der Wal, 2009; Martin 2015). However, the problem remains that evaluating single articles based on the quality of the publishing journal leads in the majority of cases to incorrect assessments, due to the skewed distribution of citations (e.g. Hamermesh 2018).

Could it be the case that impact factors and journal lists are still so influential because they possess positive qualities that outweigh their disadvantages? Advocates of the "paper quality theory" (Mingers and Xu, 2010) argue in this vein that top journals have more qualified reviewers and have editors who are better able to select promising articles than those of less highly ranked journals. This is certainly correct for journals on average. It is exactly what the JIF establishes, provided citations are taken as a proxy for the scholarly influence of a paper. Moreover, high journal rankings of management journals not only display some discriminatory power in interdisciplinarity, theoretical diversity, and (recombinant) innovativeness (Vogel et al., 2017), but also indicate a minimum threshold of quality. High impact factors also correlate with high rejection rates and thus stronger competition (e.g. Haensly et al., 2008). Further, the strongest driver of citations in management journals is the ranking of the journal itself (Mingers and Xu, 2010), which might be interpreted as a signal of the quality of high-ranked journals.

However, there are two arguments against the "paper quality theory" which assumes that high-ranked journals publish only the best papers (Mingers and Xu, 201O). First, although top journals on average publish more highly cited articles, there is a great deal of randomness in their editorial selections (Rothwell and Martyn, 2000; Bedeian, 2003; Starbuck, 2005; Siler, Lee, and Bero, 2015). As discussed, the great majority of articles published in top-tier journals are cited far below the impact factor of the publishing journals. Most articles are cited little. This suggests that even the best referees and editors are able to assess the future impact of an article to only a limited degree. Reviewers' ratings of impact correlate only 0,14 with later citations for published articles (Gottfredson, 1978; Starbuck, 2015). The reason is not any lack of expertise or fairness, though biases may play a role (e.g. Bornmann, 2011). More importantly, it is a consequence of fundamental uncertainty in research (Bush, 1945; Dasgupta and Davis 1994; Nelson, 1959, 2004; Stephan, 1996); that is, possible innovations are unknown, outcomes and alternatives are ambiguous, [15] serendipity is ubiquitous, [16] and individual ambiguity-aversion differs much (Krahnen et al.2014). Such uncertainty is demonstrated by inconclusive reviews (Nightingale and Scott 2007), low prognostic quality of reviews and low interrater reliability between the judgments of peers (Peters and Ceci, 1982; Starbuck, 2005, 2015; Bornmann, 2011; Nicolai, Schmal, and Schuster, 2015). It is also indicated by empirical findings on the "luck of the reviewer draw" (Cole, Cole, and Simon, 1981; Bornmann and Daniel, 2009), which in many cases is decisive for the acceptance or rejection of a grant proposal or paper. This phenomenon is illustrated by rejections of articles by authors who later won the Nobel Prize (Gans and Shepherd, 1994; Campanario, 1996; The Guardian, 2013[17]). This is not very often the case. However, Campanario (1995; 2009) discusses no less than nineteen Nobel class papers in the natural sciences that were rejected or had major difficulties during the review process.

Second, the journal effect theory (Mingers and Xu, 2010) argues that journal rankings activate strong Matthew effects, by which "success breeds success" (Merton, 1968; Starbuck 2005; Espeland and Sauder, 2007). The high rank of a journal attracts more readers and thus more citations, which leads to a circular causality. This means that, in contrast to what Garfield (1973) intended, the impact factor of a journal has a considerable impact on the average citation rate. This consequence was shown in a natural experiment by Larivière and Gingras (2010). Duplicate articles published in

[15] in the sense of Knightian uncertainty (Knight, 1921), see e.g. Dosi et al. (2006).
[16] that is, search might lead to results far from the expected ones.
[17] In this article, Daniel Shechtman, the Nobel prize winner for chemistry in 2011, talks about the massive initial rejection of his research even by a former Nobel prize winner.

high-ranked journals produced twice as many citations on average as their identical counterparts in lower-ranked journals.

Summing up the arguments, many influential scholars and institutions are justified in their assertion that - as the International Mathematical Union stated - classifying articles according to the ranking of the journals in which they were published is an “insidious misuse” (IMU, 2008: 9). Nevertheless, the role that impact factors and journal quality lists play in the evaluation of single articles has not diminished (e.g. Heckman and Moktan, 2018; Vogel et al. 2017). Baum’s (2011: 464) statement is still valid: “Typically, a measure found to be ill-conceived, unreliable, and invalid will fall into disrepute and disuse among the members of a scientific community. Remarkably, this has not been the case with the IF among organization theorists; indeed it is, if anything, gaining attention and being applied more frequently....”. Why is this the case?

First, a majority of the authors whose papers are accepted for publication benefit from this measure. It is exactly the skewed distribution of citations that is beneficial for many authors. As argued, the quality of two thirds to three quarters of all articles is overestimated if they are evaluated according to the impact factor of the journal in which they were published. Thus, a majority of authors in a good journal can claim to have published well even if their work has been cited little. They are able to adorn themselves with borrowed plumes, while only a minority[18] would benefit from being accepted in a higher-ranked journal. It is not surprising that the majority of winners are not inclined to abolish the present system.

[18] except the authors in the highest-ranked journal

Second, performance indicators tend to establish a "performance paradox" (Gupta and Meyer, 1994; Frost and Brockmann, 2014).[19] Indicators not only cause reactivity (Espeland and Sauder, 2007) but may also cause perverse learning or lock-in effects (Osterloh, 2010). This is the case when people focus on performance indicators but not on the performance they are supposed to indicate. They tend to improve indicators ("playing to the test") without improving the performance characteristics the indicators are designed to measure. This practice may even worsen performance, for instance by goal displacement (Ordonez et al., 2009), gap-spotting research (Alvesson and Sandberg, 2013), and ranking games (Osterloh and Frey, 2014). Once a certain performance indicator has become established, people who have gained success with this indicator will make a strong effort to maintain its relevance, even if it has been proven to be misleading.

Such lock-in effects are reinforced by ever-growing bureaucracies. In many universities, report and reward systems are established that are aligned to journal rankings and impact factors. Research administrators increasingly allocate budgets and funds according to these criteria (e.g. Laudel, 2006; Bleiklie et al. 2015). Because funding inequality has increased strongly (Zhi and Meng, 2016; Katz and Matter, 2017), authors, deans, and research communities have "to play the game" (Macdonald and Kam, 2007; Frost and Brockmann, 2014). As a consequence, a ranking bureaucracy and even a ranking management industry have emerged (Mingers and Willmott, 2013).

Lock-in effects are also reinforced by adaptive expectations. Organizations' members are willing to adopt certain measurement criteria when they assume that

---

[19] The performance paradox literature argues similarly as the literature on organizational path dependencies, see e.g. Sydow et al. (2009). However, path dependencies usually start with a useful innovation. This is not the case with the JIF as a performance indicator for single articles.

others do so. If scholars expect influential scholars or committees to use impact factors as a proxy for quality, they adopt these criteria for their own work. They also direct their attention accordingly. A self-fulfilling prophecy may set in (Ferraro, Pfeffer and Sutton 2005; Espeland and Sauder 2007).

Lock-in effects might also be strengthened by the fact that the information about the acceptance of a paper is available earlier than that about citation counts. In contrast, citation counts as a proxy for quality need several years to make any sense. The impact factor of a journal provides scholars seemingly with a speedy quality indicator, in particular because impact factors are freely available.[20]

Lastly, it might be argued that no suitable alternatives exist to impact factors and journal lists, which are easy to handle.[21] Because time and resources are limited for assessing the huge amount of research we face, heuristics to select what to read are desirable. However, heuristics may be misleading. As we have demonstrated, this is the case when using quality indicators of journals (such as JIF or quality lists) to evaluate particular articles. We therefore focus on institutional changes inducing the use of more helpful heuristics.

4. Proposals for change

Although the use of journal rankings has been widely criticized, few proposals exist for changing the current practice of performance management in academia. Most concern the individual level. In particular, it has been suggested that the papers should

[20] Impact factors are readily available, but unfortunately, they are not easy to check. The data used by the providers of the JIF are not open to the public, see Martin (2016).
[21] There are suggestions to use other indicators than impact factors, (e.g. Rost, Teichert, and Pilkington, 2017) or to apply a mix of different indicators (Aguinis et al., 2014). These suggestions are welcome; however, they are not easy to handle.

be read instead of relying on journal rankings (e.g. Moed, 2007; DORA, 2012; Wilsdon et al., 2015; Alberts, 2013; Berg, 2016; Heckman and Moktan, 2018). This is certainly good advice, but hard to put into practice. We first discuss three extant proposals to reform performance evaluation. We then introduce our own suggestion based on the insight that research is characterized by fundamental uncertainty. All four proposals refer to the institutional level.

A first proposal intends to change the academic journal system as a whole. It suggests to evaluate scholarly work through "open post-publication peer review" (Kriegeskorte, 2012; Osterloh and Kieser, 2015). The internet allows manuscripts to be published as they are and to be evaluated ex post. This procedure starts with the publication of a paper in an online public repository. The author asks a senior scholar to try to find two to four reviewers willing to comment publicly on the paper. This creates transparency within the reviewing process and a plurality of perspectives. Some contributions will elicit inspiring debates; others will be ignored. The papers that have inspired the most interesting discussions might be presented to a broader audience as the state of art in special issues. However, unintended consequences may occur. First, the reputation of the senior scholar and of the reviewers will have a great impact on the attention that the paper receives. In contrast, today it is the reputation of a journal that has been acquired for a long time within a research community that counts for the attention for an article. Second, since comments and reviews are conducted publicly, junior scholars may be reluctant to critique the work of senior scholars. In addition, old boys´ networks might play an undesirable role, and cronyism could arise. Ultimately, the system of open post-publication peer review could lead to a ranking of publication outlets that produces similar problems as the evaluation of single articles according to the quality of a journal.

In contrast to the first proposal the following three accept the crucial role of journals to focus on topical and relevant issues. The second proposal suggests that every journal publishing its JIF should also publish the distribution of citations (Larivière et al., 2016). In the meantime, this proposal has been taken on board by Clarivate Analytics.[22] This proposal could apply to journal quality lists in general. For those who believe in citations as a signal of scholarly impact it can be used to reveal the extensive overlap between the citation distributions of different journals. It will broaden awareness of the spread of citations. It can also be used to measure how often an author's publications are cited above (or below) the impact factors of the journals he or she has appeared in. An alternative would be to provide parameters of distribution such as median or inter-quartile ranges, but a visual representation is more powerful. This suggestion meets the demands that editors and reviewers usually make on authors to make their data traceable.[23]

This suggestion has the advantage of being close to current practice and therefore of being accepted widely. It should, however, be taken into account that the time frame used by JIF is too narrow to evaluate a paper´s influence.[24] Moreover, the distribution of citations still relies on the questionable assumption that citations are a good measure of scholarly impact and that the present reviewing and acceptance procedures accurately reveal the "collective wisdom" (Laband, 2013) of the scientific community.

A third proposal is the publication of a manuscript on an "as is" basis (Tsang and

---

[22] See https://clarivate.com/blog/science-research-connect/the-2018-jcr-release-is-here/

[23] In earlier times the data that Thomson Reuter uses to produce the JIF were not openly available, and efforts to replicate individual impact values had failed (Rossner, Van Epps, and Hill, 2007, 2008).

[24] This is the reason why our own analysis presented above uses cumulative citations over a five year time span.

Frey, 2007). A paper is reviewed double-blind as usual. The reviewers are given only two options when advising to the editor: to accept or reject the paper. The option to revise and resubmit is ruled out. The editor then decides whether the manuscript is published as it is or not. If the paper is accepted, then it is up to the authors to incorporate the comments of the reviewers into the paper. The editor also publishes a comment that addresses differences of view among reviewers and him- or herself. This suggestion would speed up the review process and the dissemination of new knowledge. It would unburden reviewers from evaluating revised and resubmitted papers. It also would avoid that authors feel as if they were coerced by the reviewers instead of being advised (Bedeian, 2003; Frey, 2003). Most importantly, this suggestion would make clear to both the authors and the readers that being accepted by a high-impact journal is not a universal quality indicator. The editors would be burdened with a higher responsibility than today to achieve and to demonstrate the state of "organized skepticism" (Merton, 1942) and “creative disagreement” (Harnad, 1979) that is at the heart of scholarly work. But it might encourage editors to publish more imaginative studies.

Our own – the fourth - proposal to overcome the performance paradox and the lock-in effect is based on the insight that uncertainty about future success is symptomatic of scholarly work (Bush, 1945; Nelson, 2004; Stephan, 1996). This insight can be liberating (Starbuck 2015). Therefore, we translate it into the peer review system. Uncertainty can be used to the advantage of scholarship with the following procedure:

When reviewers agree on the excellent quality of a paper, it should be accepted, preferably on an “as is” basis (Tsang and Frey, 2007). Papers perceived unanimously as valueless are rejected immediately. Papers that are evaluated differently by the referees are randomized. Empirical research has found reviewers´ evaluations to be more congruent with poor contributions (Cicchetti, 1991; Bornmann, 2011; Moed, 2007; Siler et al., 2015) and fairly effective in identifying extremely strong contributions (Li and Agha, 2015). However, reviewers' ability to predict the future impact of contributions has been shown to be particularly limited in the middle range in which reviewers´ judgements conform to a low degree (Fang et al., 2016).[25] Such papers could undergo a random draw.

Why should contributions to which the referees do not agree be randomized? This procedure reduces the "conservative bias", that is the bias against unconventional ideas. Referees subjectively have more information on research projects that are close to existing knowledge. Moreover, information on those contributions is more consistent. With unorthodox contributions referees have less – and usually inconsistent - information. But such ideas yield may well high returns in the future. Under these circumstances a randomized choice among the unorthodox contributions is advantageous. Brezis (2007) shows in a numerical model that the optimal ranking mechanism is to accept contributions to which all referees have agreed and to reject those that all referees have put on the bottom and the variance is high.[26] It is the different level and different consistency of information between conventional and unorthodox contributions that is key to focal randomization among papers that referees disagree upon. Gilles (2008) and Engwall (2014) argue in a similar vein. They refer to the theory of statistical tests involving two types of error: type I errors ("reject errors") implying that a correct hypothesis is rejected, and type 2 errors implying that a false

[25] Li and Agha (2015) as well as Fang et al. (2016) refer to grant applications.
[26] Brezis (2007) refers to R&D projects.

hypothesis is accepted ("accept errors"). The former matters more than the latter. "Reject errors" stop promising new ideas, sometimes for a long time, while "accept errors" lead to a waste of money, but may be detected soon once published. This is the reason why it is more difficult to identify "reject errors" than "accept errors".[27] To avoid the negative consequences of "reject errors", risks must be diversified. Fang and Casadevall (2016:158) support this argument by stating that " [j]ust as passively managed diversified stock portfolios that rely on random fluctuations of the stock market generally outperform active management based on expert predictions, a modified lottery-based funding strategy would maximize the return on society's investment". The suggestion of partly focal randomization of grants has already been put in practice by two big funding agencies.[28] Other research councils share such considerations.[29]

Our proposal applies these insights to the selection of journal articles. Disagreement among journal referee reports matters more than those among those on grant applications. In the latter case referees usually engage in extensive consultation and mutual adjustments before the final decision is made (Reinhart, 2010). Reducing the "conservative bias" by focal randomization of controversial papers not only diversifies risk of rejecting fruitful ideas, but in addition has an incentivizing effect. It

---

[27] Engwall (2014) argues that „reject errors" will become larger the higher the percentage of desk rejections is. He presumes that due to „reject errors" the most innovative research will be found in low impact factors. See for empirical evidence Siler, Lee and Bero (2016).

[28] Volkswagen Foundation , https://www.volkswagenstiftung.de/en/funding/our-funding-portfolio-at-a-glance/experiment
Health Research Council of New Zealand Explorer Grants.
http://www.hrc.govt.nz/funding-opportunities/researcher-initiated-proposal/explorer-grants;

[29] E.g. German Council of Science and Humanities
https://www.wissenschaftsrat.de/index.php?id=1405&L=;

encourages researchers to submit unorthodox ideas that otherwise have a hard time being published (e.g. Alvesson and Sandberg, 2013).

Rational scholars might feel uneasy with randomization mechanisms, However, with focal randomization scholars remain in power. They decide which papers are published or rejected immediately and which enter the randomization process. The purposeful use of random mechanisms in academia is not new. It played a role in the 18th century at the University of Basel. Vacant professorial chairs were filled by lot from a list of three candidates (Burckhardt, 1916; Stolz, 1986; Frey and Osterloh, 2015).[30] At that time the main purpose was to weaken old boys´ networks. Today the main purpose is to ensure diversity that is crucial for the progress of scholarly work (Starbuck, 2015). It also serves to encourage the submission of unorthodox yet promising ideas. The "tyranny of the top five" and their role in tenure and promotion decisions is de-emphasized, and the signaling function among a diversity of journals is redistributed. These goals are explicitly stated by Nobel Prize laureate Heckman (Heckman and Moktan 2018: 54). Moreover, Matthew effects and lock-in effects are mitigated.

Our proposal moreover unburdens editors considerably from the problem of dealing with low interrater reliability and contradictory reviews. In contrast to the unintended randomness attributed to the peer review process (e.g. Peters and Ceci, 1982; Starbuck, 2005; Bornmann and Daniel, 2009; Rothwell and Martyn, 2000; Graves, Barnett, and Clarke; 2011; Smith, 2015; Nicolai, Schmal, and Schuster, 2015), which is sometimes close to an unintended lottery (Rothwell and Martyn, 2000; Bedeian, 2003;

---

[30] In political governance too, mixed procedures of random elements and voting were common, for instance in classical Athens and in medieval Venice and Florence (Manin, 1997; Buchstein, 2009; Van Reybrouck, 2016).

Siler, Lee, and Bero, 2015), this suggestion applies randomness in a strictly controlled and rational way.

Such a system would also possess some disadvantages. First, random procedures do not differentiate between good and bad quality. This is the reason why they are preceded by a pre-selection based on quality. It is important to note that the better the pre-selection works, the less the quality of the remaining papers can be distinguished. In this case, the variance in quality is reduced. It becomes much harder to decide which is “the best” or the “second best” paper (March and March 1977; Denrell, Fang, and Liu, 2014). Through focal randomization, the seeming disadvantage becomes an advantage, since otherwise personal preferences and unintended randomness might be decisive (Brezis, 2007). Second, random decisions are considered by many people to be “irrational”. However, seemingly rational decisions are often marred by many biases (Kahneman, 2011). An example is awarding prizes in some competitions, which turns out to be unintentionally random (Ginsburgh and Weyers, 2014). In such cases, the rationality of decision processes is a façade; an intentionally random decision based on mathematical probabilities would be much more rational. Third, more articles of low quality could be submitted if scholars knew that random selection played a role. But it could equally be the case that more unorthodox high-quality articles would be submitted because authors would feel more encouraged than with the present system.

5. Concluding remarks

The present practice of performance management in academia based on journal quality lists and impact factors needs reform. Publication in a “good” journal does not indicate that the article is "good". Empirical research shows that about two-thirds to three-quarters of all published articles are overvalued by these criteria. In contrast, frequently cited articles which have had the misfortune to be published in low-ranked journals are undervalued. We show that this is true for both short citation windows and five-year spans.

We discuss why the present practice has gained so much influence. We suggest this is the case because a majority of authors benefits unduly from the present system. Moreover, performance paradox effects, lock-in effects, and ranking bureaucracies block reforms. Therefore, appealing to scholars individually is not sufficient to change the present practice of performance management. Instead, proposals are needed for changes at the institutional level that give incentives to mitigate the obsession of top journal publications. We discuss three suggestions made in the literature. The first is to inform scholars regularly about the skewed distribution of citations of articles and to show the overlap in the distributions for different-tier journals. The second, more far-reaching, proposal is "open post-publication peer review", which abolishes ex-ante double-blind peer reviews. The third proposal is the publication of manuscripts on the basis of double-blind ex-ante reviews but "as-is".

Our own proposal is the most radical. It is based on the insight that fundamental uncertainty is symptomatic for scholarly work. This is indicated by the low prognostic quality of reviews and the low interrater reliability revealed by many empirical analyses. Our suggestion takes this evidence into account. It suggests the introduction of a partly random mechanism. Focal randomisation takes place after a thorough preselection of articles by peer reviews. Such a rationally founded and well-orchestrated procedure promises to downplay the importance (or even "tyranny") of top journals and to encourage more unorthodox research than today.

All four proposals could be initiated in an experimental way, preferably as field experiments. Their outcomes could be evaluated after some years. In any case, they serve to enrich the discussion about the inevitable uncertainty of quality indicators in science.

References

ABS The Association of Business Schools [ABS], 2015. *Academic Journal Guide 2015.* Retrieved September 4, 2017, from http://www.kfs.edu.eg/com/pdf/20820152253917.pdf.

Adler, N. J., Harzing, A.-W., 2009. When knowledge wins: Transcending the sense and nonsense of academic rankings. *Academy of Management Learning*, 8(1), 72-95.

Aguinis, H., Shapiro, D. L., Antonacopoulou, E. P., Cummings, T. G., 2014. Scholarly Impact: A Pluralist Conceptualization. *Academy of Management Learning & Education*, 13(4), 623-639.

Alberts, B., 2013. Editorial: Impact Factor Distortions. *Science*, 340(6134), 787. Retrieved August 31, 2017, from http://science.sciencemag.org/content/340/6134/787

Alvesson, M., Sandberg, J., 2013. Has Management Studies lost its way? Ideas for more imaginative and innovative research. *Journal of Management Studies,* 50(1), 128-152.

Antonakis, J., Bastardoz, N., Liu, Y., Schriesheim, C. A., 2014. What makes papers highly cited? *Leadership Quarterly,* 25, 152–179.

Archambault, É., Larivière, V., 2009. History of the journal impact factor: contingencies and consequences. *Scientometrics,* 79(3), 639-653.

Balaban, A. T., 2012. Positive and negative aspects of citation indices and journal impact factors. *Scientometrics*, 92(2), 241-247.

Baum, J. A. C., 2011. Free-Riding on Power Laws: questioning the validity of the Impact Factor as a measure of research quality in organization studies. *Organization*, 18(4), 449-466.

Bedeian, A. G., 2003. The manuscript review process: The proper roles of authors, referees, and editors. *Journal of Management Inquiry,* 12(4), 331-338.

Berg, J., 2016. JIFfy Pop. *Science*, 353(6299), 523.

Bleiklie, I., Enders, J., Lepori B., 2015. Organizations as Penetrated Hierarchies: Environmental Pressures and Control in Professional Organizations. *Organization Studies,* 36 (7), 873-896.

Bornmann, L., 2011. Scientific Peer Review. *Annual Review of Information Science and Technology*, 45(1), 199-245.

Bornmann, L., Daniel, H. D., 2009. The luck of the referee draw: The effect of exchanging reviews. *Learned Publishing*, 22(2), 117-125.

Brezis, E. S., 2007. Focal randomization: An optimal mechanism for the evaluation of R&D projects. *Science and Public Policy,* 34, 691-698.

Burckhardt, A., 1916. Ueber die Wahlart der Basler Professoren, besonders im 18. Jahrhundert. *Basler Zeitschrift für Geschichte und Altertumskunde*, 15, 28-46.

Buchstein, H., 2009. *Demokratie und Lotterie*. Frankfurt am Main: Campus Verlag.

Bush, V., 1945. *Science: The endless frontier. A Report to the president by Vannevar Bush, Director of the Office of Scientific Research and Development. July, 1945.* Washington, DC: United States Government Printing Office. Retrieved September 4, 2017, from https://www.nsf.gov/od/lpa/nsf50/vbush1945.htm.

Callaway, E., 2016. Beat it, Impact Factor! Publishing elite turns against controversial metric. *Nature*, 535(7611), 210-211.

Campanario, J.M., 1995. Commentary on influential books and journal articles initially rejected because of negative referees evaluations. *Science Communication,*16, 304-325.

Campanario, J. M., 1996. Using citation classics to study the incidence of serendipity in scientific discovery. *Scientometrics,* 37(1), 3-24.

Campanario, J.M., 2009. Rejecting and resisting Nobel class discoveries: accounts by Nobel Laureates. *Scientometrics,* 81(2),549-565.

Cicchetti, D. V., 1991. The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, 14(1), 119-135.

Cole, S., Cole, J. R., Simon, G. A., 1981. Chance and consensus in peer review. *Science,* 214(4523), 881-886.

Dasgupta, P., David, P. A., 1994. Toward a New Economics of Science. *Research Policy,* 23(5), 487-521.

Denrell, J., Fang, C., Liu, C., 2014. Perspective–Chance explanations in the management sciences. *Organization Science,* 26(3), 923-940.

DORA (San Francisco Declaration on Research Assessment), 2012. Retrieved August 30, 2017, from http://www.ascb.org/dora/

Dosi, G., Malerba, F., Ramello, B., Silva, F. ,2006. Information, appropriability, and the generation of innovative knowledge four decades after Arrow and Nelson: an introduction. *Industrial and Corporate Change,* 15 (6), 891-901.

Engwall, L., 2014. On the quality of quality assessments. In: W. Blockmans, D. Engwall., D. Wearie (Eds) *Bibliometrics: Use and abuse in the review of research performance.* Portland Press, 93 - 104. London,

Espeland, W. N., Sauder, M., 2007. Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology,* 113(1), 1-40.

Fang, F. C., Bowen, A., Casadevall, A., 2016. NIH peer review percentile scores are poorly predictive of grant productivity. *eLife 5:* e13323. DOI: 10.7554/eLife.13323.

Fang, F. C. Casadevall, A., 2016. Grantfunding: Playing the odds. *Science,* 352(6282), 158.

Ferraro, F., Pfeffer, J., Sutton, R. I., 2005. Economics language and assumptions: how theories can become self-fulfilling. *Academy of Management Review*, 30(1), 8–24.

Frey, B. S., 2003. Publishing as prostitution? – Choosing between one's own ideas and academic success. *Public Choice,* 116(1-2), 205-223.

Frey, B. S., 2009. Economists in the PITS. *International Review of Economics*, 56(4), 335-346.

Frey, B. S., Osterloh, M., 2015. Zufall in der Wissenschaft. *Ökonomenstimme,* December 16, 2015. Retrieved from http://www.oekonomenstimme.org/artikel/2015/12/zufallsauswahl-in-der-wissenschaft/

Frey, B. S., Rost, K., 2010. Do rankings reflect research quality? *Journal of Applied Economics,* 13(1), 1-38.

Frost, J., Brockmann, J., 2014. When quality is equated with quantitative productivity – Scholars caught in a performance paradox. *Zeitschrift für Erziehungswissenschaft,* 17(6), 25–45.

Fuyuno, I., Cyranoski, D., 2006. Cash for papers: putting a premium on publication. *Nature*, 441(7095), 792.

Gans, J. S., Shepherd, G. B., 1994. How are the mighty fallen: Rejected classic articles by leading economists. *The Journal of Economic Perspectives,* 8(1), 165-179.

Garfield, E., 1973. Citation impact depends upon the paper, not the journal! Don't count on citation by association. *Current Contents*, 22, 5-6.

Ginsburgh, V., Weyers, S., 2014. Nominees, winners, and losers. *Journal of Cultural Economics*, 38(4), 291-313.

Gittelman, M., Kogut, B., 2003. Does Good Science Lead to Valuable Knowledge? Biotechnical Firm and the Evolutionary Logic of Citation Patterns. *Management Science,* 49(4), 366-382.

Goodall, A. H.. 2009. Highly cited leaders and the performance of research universities. *Research Policy*, 38(7); 1079-1092.

Gottfredson, S. D., 1978. Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist* 33(10), 920–934.

Graves, N., Barnett, A. G., Clarke, P., 2011. Cutting random funding decisions. *Nature,* 469(7330), 299.

Gupta, V., Meyer, M. W., 1994. The performance paradox. *Research in Organizational Behavior*, 16, 309–369.

Haensly, P. J., Hodges, P. E., Davenport, S. A., 2008. Acceptance Rates and Journal Quality: An Analysis of Journals in Economics and Finance. *Journal of Business and Finance Librarianship,* 14(1), 2-31.

Hamermesh, D. S., 2018. Citations in Economics: Measurement, Uses, and Impacts. *Journal of Economic Literature,* 56(1), 115-156.

Harnad, S., 1979. Creative disagreement. *The Sciences,* 19, 18-20.

Harzing, A.-W., 2015. Journal Quality List, 62th ed. Retrieved from https://harzing.com/resources/journal-quality-list

Harzing, A.-W., van der Wal, R., 2009. A Google Scholar h-index for Journals: An alternative metric to measure journal impact in Economics & Business**?** *Journal of the American Society for Information Science and Technology,* 60 (1), 41-46

Health Research Council of New Zealand, 2017. *Explorer Grants.* Retrieved September, 4, 2017, from http://www.hrc.govt.nz/funding-opportunities/researcher-initiated-proposals/explorer-grants

Heckman, J. J., Moktan, S., 2018. Publishing and Promotion in Economics: The tyranny of the Top Five. © NBER Working Paper Series Working 25093.

Hicks D., Wouters P., Waltman L., de Rijcke S., Rafols I., 2015. Bibliometrics: The Leiden Manifesto for research metrics. *Nature*. 520, 429-31.

International Mathematical Union [IMU], 2008. *Citation Statistics. A report.* Corrected version, 16/12/08. Retrieved October 4, 2018, from https://www.mathunion.org/imu-net/archive/2008/imu-net-029b

Kahnemann, D., 2011. *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux.

Katz, Y., Matter, U., 2017. *On the Biomedical Elite. Inequality and Stasis in Scientific Knowledge Production.* Unpubl. Manuscript, Harvard Medical School and Faculty of Economics and Business, University of Basel.

F. Knight: *Risk, Uncertainty, and Profit*. Mifflin, Boston, New York 1921.

Krahnen, J.P., Ockenfels, P., Wilde, 2014. *Measuring Ambiguity Aversion: A Systematic Experimental Approach*. SSRN_ID2457026

Kriegeskorte, N., 2012. Open evaluation: a vision for entirely transparent post-publication peer review and rating for science. *Frontiers in Computational Neuroscience,* 6, 1-18.

Laband, D. N., 2013. On the Use and Abuse of Economics Journal Rankings. *The Economic Journal,* 123, F223-54.

Laband, D. N., Tollison, R. D., 2003. Dry holes in economic research. *Kyklos,* 56, 161-174.

Larivière, V., Gingras, Y., 2010. The Impact Factor's Matthew Effect: A Natural Experiment in Bibliometrics. *Journal of the American Society for Information Science and Technology,* 61(2), 424-427.

Larivière, V., Kiermer, V., MacCallum, C. J., McNutt, M., Patterson, M., Pulverer, B., Swaminathan, S., Taylor, S., Curry, S., 2016. *A simple proposal for the publication* *of journal citation distributions.* Retrieved September 11, 2018, from http://www.biorxiv.org/content/early/2016/09/11/062109.

Laudel, G., 2006. The 'quality myth': Promoting and hindering conditions for acquiring research funds. *Higher Education*, 52(3), 375-403.

Li, D., Agha, L., 2015. Big names or big ideas: do peer-review panels select the best science proposals? *Science* 348, 434–438.

Macdonald, S., Kam, J., 2007. Ring a Ring o' Roses: Quality Journals and Gamesmanship in Management Studies. *Journal of Management Studies,* 44(4), 640–55.

Macdonald, S. , Kam, J., 2010. Counting footnotes: citability in Management Studies. *Scandinavian Journal of Management,* 26(2), 189-203.

Macdonald, S., Kam, J., 2011. The skewed few: people and papers of quality in management studies. *Organization*, 18(4), 467-475.

Manin, B., 1997. *The principles of representative government*. Cambridge: Cambridge University Press.

March, J.C., March, J.G., 1977. Almost random careers: The Wisconsin school superintendency, 1940-1972. *Administrative Science Quarterly,*22 (3), 377-409.

Martenson, P., Fors, U., Sven-Bertil, W., Zander, U., Nilsson, G. H. ,2016. Evaluating research: A multidisciplinary approach to assessing research practice and quality. *Research Policy,* 45, 593-603.

Martin, B. R., 2015. An new editorial team, and a new indicator of where RP stands. *Research Policy,* 44 (1), 1-3.

Martin, B. R., 2016. Editorial: Editors´ JIF-boosting stratagems – Which are appropiate and which not? *Research Policy,* 45(1), 1-7.

Merton, R. K., 1942. Science and technology in a democratic order. *Journal of Legal and Political Sociology,* 1, 115-126.

Merton, R. K., 1968. The Matthew Effect in Science. *Science,* 159(3810), 56-63.

Mingers, J., 2008. Exploring the dynamics of journal citations: modelling with Scurves. *Journal of the Operational Research Society,* 59(8), 1013-1025.

Mingers, J., Xu, F., 2010. The drivers of citations in management science journals. *European Journal of Operational Research,* 205(2), 422-430.

Mingers, J., Willmott, H., 2013. Taylorizing business school research: on the 'one best way' performative effects of journal ranking lists. *Human Relations,* 66 (8), 1051-1073.

Moed, H. F., Van Leeuwen T. N., 1996. Impact Factors Can Mislead, *Nature,* 381, 186.

Moed, H. F., 2007. The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science and Public Policy,* 34(8), 575-583.

Nelson, R. R., 1959. The Simple Economics of Basic Scientific Research. *Journal of Political Economy,* 67(3), 297-306.

Nelson, R. R., 2004. The Market Economy, and the Scientific Commons. *Research Policy,* 33(3), 455-471.

Nicolai, A. T., Schmal, S., Schuster, Ch., 2015. Interrater Reliability of the Peer Review Process in Management Journals. In I. M. Welpe, J. Wollersheim, S. Ringelhan, M. Osterloh (Eds.), *Incentives and Performance. Governance of Research Organizations***,** 107-120. Cham: Springer.

Nightingale, P., Scott, A., 2007. Peer review and the relevance gap: ten suggestions for policy-makers. *Science and Public Policy*, 34 (8), 543-553.

Ordonez, L. D., Schweitzer, M. E., Galinsky, A. D., Bazerman, M. H., 2009. Goals gone wild: the systematic side effects of overprescribing goal setting. *Academy of Management Perspectives,* 23(1), 6–16.

Osterloh, M., 2010. Governance by Numbers. Does it Really Work in research? *Analyse & Kritik: Zeitschrift für Sozialtheorie,* 32(2), 267–283.

Osterloh, M., Frey, B. S., 2014. Ranking Games. *Evaluation Review,* 39(1), 102-129.

Osterloh, M., Kieser, A., 2015. Double-Blind Peer Review: How to Slaughter a Sacred Cow. In I. Welpe, J. Wollersheim, S. Ringelhan, M. Osterloh (Eds.), *Incentives and Performance. Governance of Research Organizations,* 307-324. Cham: Springer.

Oswald, A. J., 2007. An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers. *Economica,* 74(293), 21-31.

Paulus, F. M., Rademacher, L., Schäfer, T. A. J., Müller-Pinzler, L., Krach, S., 2015. Journal Impact Factor Shapes Scientists' Reward Signal in the Prospect of Publication. *PLoS ONE*, 10(11). Retrieved September 11, 2018, from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0142537

Peters, D., Ceci, S. J., 1982. Peer review practices of psychological journals: The fate of published articles, submitted again. *The Behavioral and Brain Sciences,* 5(2), 187-195.

Reinhart, M., 2010. Peer Review Practices: Epistemic and Social Aspects in the Assesment of Scientific Work. *Research Evaluation*, 19(5), 317-331.

Rossner, M., Van Epps, H., Hill, E., 2007. Show me the data. *The Journal of Cell Biology*, 179(6), 1091-1092.

Rossner, M., Van Epps, H., Hill, E., 2008. Irreproducible results: A response to Thomson Scientific*. The Journal of Cell Biology,* 180(2), 254-255.

Rost, K., Teichert, T., Pilkington, A., 2017. Social Network Analytics for Advanced Bibliometrics: Referring to Actor Roles of Management Journals instead of Journal Rankings. *Scientometrics***,** 112(3), 1631-1657.

Rothwell, P. M., Martyn, C. N., 2000. Reproducibility of peer review in clinical neuroscience. Is agreement between reviewers any greater than would be expected by chance alone? *Brain,* 123(9), 1964-1969.

Seglen, P. O., 1997. Why the Impact Factor of journals should not be used for evaluating research. *British Medical Journal,* 314(7079), 498-502.

Shao, J., Shen, H., 2011. The outflow of academic papers from China: why is it happening and can it be stemmed? *Learned Publishing,* 24(2), 95-97.

Siler, K., Lee, K., Bero, L., 2015. Measuring the effectiveness of scientific gatekeeping. *PNAS,* 112(2), 360-365.

Singh, G., Haddad, K. M., Chow, S., 2007. Are articles in “top” management journals necessarily of higher quality? *Journal of Management Inquiry,* 16(4), 319-331.

Smith, R., 2015. Ineffective at any dose? Why peer review simply doesn't work. *Times Higher Education Supplement*, May 28, 2015 (2205) available at https://www.questia.com/magazine/1P3-3701977521/ineffective-at-any-dose-why-peer-review-simply-doesn-t.

Starbuck, W. H., 2005. How much better are the most prestigious journals? The statistics of academic publication. *Organization Science,* 16(2), 180-200.

Starbuck, W. H., 2015. Issues and trends in publishing behavioral science: A quarrelsome crew struggling with a disintegrating boat on a stormy sea. In I. M. Welpe, J. Wollersheim, S. Ringelhan, M. Osterloh (Eds.), *Incentives and Performance. Governance of Research Organizations,* 3-18. Cham: Springer.

Stephan, P. E., 1996. The economics of science. *Journal of Economic Literature,* 34(3), 1199-1235.

Stolz, P., 1986. Parteienwettbewerb, politisches Kartell und Tausch zwischen sozioökonomischen Gruppen. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik,* 122, 657-675.

Sydow, J., Schreyögg, G. Koch, J., 2009. Organizational path dependence: opening the black box. *Academy of Management Review*, 34(4), 689–709.

The Guardian, 2013, Jan. 6. Dan Shechtman: "Linus Pauling said I was talking nonsense". Retrieved Januar 8, 2018, from https://www.theguardian.com/science/2013/jan/06/dan-shechtman-nobel-prize-chemistry-interview.

Tsang, E. W. K., Frey, B. S., 2007. The as-is journal review process: Let authors own their ideas. *Academy of Management Learning and Education,* 6(1), 128-136.

Van Reybrouck, D., 2016. *Against Elections. The Case for Democracy*. London: Penguin Random House.

Vogel, R., Hattke, F., Petersen, J., 2017. Journal rankings in management and business studies: What rules do we play by? *Research Policy,* 46, 1707-1722.

Wallace, M. L, Larivière, V., Gingras, Y., 2009. Modeling a century of citation distributions. *Journal of Informetrics,* 3(4), 296-303. doi: 10.1016/j.joi.2009.03.

Walsh, J. P., 2011. 2010 Presidential address: Embracing the sacred in our secular scholarly world. *Academy of Management Review,* 36(2), 215-234.

Waltman, L., 2016. A review of the literature on citation impact indicators. *Journal of Informetrics,* 10(2016), 365-391.

Willmott, H., 2011. Journal List Fetishism and the Perversion of Scholarship: Reactivity and the ABS List. *Organization,* 18(4), 429-442.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., Johnson, B., 2015. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management.* doi: 10.13140/RG.2.1.4929.1363. Retrieved September 11, 2018, from http://blogs.lse.ac.uk/impactofsocialsciences/files/2015/07/2015_metrictide.

Zhi, Q., Meng, T., 2016. Funding allocation, inequality, and scientific research output: an empirical study based on the life science sector of Natural Science Foundation of China. *Scientometrics,* 106, 603-628.

Appendix

Statistics

Number of Citations

low-ranked journals	N	Valid	83
		Missing	0
	Mean		7.5904
	Std. Error of Mean		1.17660
	Median		5.0000
	Std. Deviation		10.71930
	Variance		114.903
	Minimum		.00
	Maximum		61.00
	Percentiles	25	2.0000
		50	5.0000
		75	8.0000
middle-ranked journals	N	Valid	110
		Missing	0
	Mean		13.6818
	Std. Error of Mean		1.85594
	Median		9.0000
	Std. Deviation		19.46530

	Percentiles	25	1.0000
		50	9.0000
		75	19.2500
high-ranked journals	N	Valid	154
		Missing	0
	Mean		71.5000
	Std. Error of Mean		4.53468
	Median		52.5000
	Std. Deviation		56.27387
	Variance		3166.748
	Minimum		1.00
	Maximum		205.00
	Percentiles	25	27.7500
		50	52.5000
		75	87.7500

Table 1: Statistics of Citations in Low-, Middle- and High –Ranked Journals over five years 2012-2016.