Journal of Informetrics 11 (2017) 945–947

Contents lists available at ScienceDirect

Journal of Informetrics

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

Correspondence

# Use of the journal impact factor as a criterion for the selection of junior researchers: A rejoinder on a comment by Peters (2017)

The Journal Impact Factor (JIF) is a simple bibliometric indicator measuring the average citation impact of journals in the short run. In a recent study, we tested the ability of the JIF to predict future performance of researchers (Bornmann & Williams, 2017). We did not include the original JIF in the study, but JIF percentiles. In this way, we followed the bibliometric standard of normalizing citation impact for the publishing field and publication year (Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015). We compared the performance in the first five years (based on the JIF) with the performance in the following years (based on the citation impact of papers) for several thousands of researchers. Our results show that researchers with different proportions of papers in prestigious journals (measured in terms of the JIF) early in their careers also differed in their level of publication success later. The results are statistically significant. Since the effect sizes of the results are not better than moderate we concluded that "university departments and granting agencies should not rely solely on early JIFs (in their normalized variants) when rewarding work and allocating resources" (p. 795).

Peters (2017) criticizes our study and deals with possible procedures for the selection of junior researchers. We welcome the comment by Peters (2017) on our paper. However, we do not understand why Peters (2017) describes our study as a "seemingly unquestioning acceptance of JIF as indicator of scientific quality". Since we use empirical data to clarify the meaningfulness of the JIF, our approach cannot be labelled as "unquestioning". The empirical results of our study even give rise to conclusions which might be interpreted as contrary to previous published positions of one of the co-authors (Lutz Bornmann). A simple inspection of the co-authors' publication lists would have shown that this co-author has mostly written on the JIF in a critical way (see e.g., Bornmann, Marx, Gasparyan, & Kitas, 2012; Leydesdorff, Wouters, & Bornmann, 2016).

As a side note, we do not indicate anywhere in our study (or in any other paper) that the JIF is an indicator of scientific quality – as Peters (2017) claims. One can find this claim in many publications about the JIF (or bibliometric indicators in general), but mostly in publications which are not co-authored by professional bibliometricians.

Although Peters (2017) indicates that the statistical analyses of our study "seem to have been conducted very well", he criticizes that we transformed JIF percentiles into quartiles. JIF percentiles are field- and time-normalized metrics, which are based on the JIF provided by Clarivate Analytics in the Journal Citation Reports (JCR). We chose this step to build groups of researchers with different early and late performance. The 'quartilization' is a frequently used statistical method – also in bibliometrics. For example, the Q1 indicator which is applied in the SCImago institutions ranking (see http://scimagoir.com) is the proportion of papers published in journals belonging to the 25% most prestigious journals (in terms of a certain journal metric). Peters (2017) also apparently overlooked that we report in our paper not only results based on 'quartilized' data, but also the results of correlation analyses which are based on not-quartilized performance data.

Peters (2017) criticizes us for arguing that correlation implies causation. We find his claims quite curious, because we explicitly state that the relationship between early success and later success could be causal or it could be spurious. Peters' argument is even more puzzling because he repeats many of the same points we make. According to the "sacred spark" theory, the same qualities that produce early success (e.g., hard work, creativity, intelligence) can also lead to later success. If so, early success does not produce later success; rather, both are caused by ongoing characteristics of the researcher. Alternatively, the relationship between early success and later success could be causal because, as cumulative advantage theory argues, early success leads to greater resources (e.g., research funding, better assistants, and better jobs) which helps produce later success. We do not take a stand on which of these processes is actually operating (although it would not surprise us if both were, see Gould, 2002); but we show that, for whatever reason, early success is modestly predictive of later success.

Peters (2017) then modifies his argument by saying that what we are really describing is a self-fulfilling prophesy: employers create the link between early success and later success by rewarding those who have done well early. He says that doing so offers no advantage to the selector and further says that employers and funders "forego opportunities for selecting on variables that are indicative of relevant competencies." We find Peters' arguments quite problematic here. We never argued that early publication success is or should be the only thing that determines future resources. Further, Peters

http://dx.doi.org/10.1016/j.joi.2017.08.005

(2017) never shows that employers and funders are foregoing opportunities to select on other relevant indicators. He also never shows that partially relying on early success is a harmful strategy or that some other strategies that ignored early success would be both superior and practical. Indeed, if the "sacred spark" theory is correct, employers/funders are wise to consider early success because it provides useful information on the qualities that can lead to later success.

Overall, we think Peters' arguments on correlation and causation are more normative than scientific, and misrepresent our stance. Certainly, researchers' fates should not be set in stone based on the success or failure of their earliest work. As we argue, early success is just a modest indicator of later success (measured in terms of bibliometric indicators), and many other factors (including subjective evaluations of the quality and potential of the work) should be considered.

At the same time, we think it is foolish to say that the link between past and present work is simply the product of a self-fulfilling prophesy that has no real predictive value. Perhaps, early success has too much weight in selecting scholars for later resources (Gregg, 1957). Indeed, according to Merton (1988) early prognostic judgments "lead in some unknown fraction of cases to inadvertent suppression of talent through the process of the self-fulfilling prophecy" (p. 613). However, taken to its logical extreme, Peters' arguments could lead to ignoring early success altogether. We think, our results show that early success does (for whatever reason) somewhat predict later success, and that is the main point of our paper. The modest relationship in our study also shows that funders and employers should take a balanced, broader view when making their decisions. Peters (2017) seems to be implying that we are telling resource providers to place too much emphasis on early success, when really we are cautioning them against that.

Peters (2017) questions the relationship between JIF and scientific quality. Since scientific quality is a latent construct, it cannot be measured directly, but only by proxies (Hug, Ochsner, & Daniel, 2014). The indicators which we used in our study are based on the average number or sum of citations (normalized for time and field of publication). In scientometrics, it has been adopted that citations measure one part of quality – namely scientific impact (Martin & Irvine, 1983). Other parts are accuracy and importance which (probably) cannot be measured by citations. Our study is always interpreted within this framework of interpreting citations (and quality) which is the rule in scientometrics. Peters (2017) recommends using the N-pact factor as an alternative to the JIF (Fraley & Vazire, 2014). However, this indicator also measures only one part of quality, namely methodological accuracy (in psychology). In this respect, it does not differ from the JIF (and other journal metrics).

In an ideal evaluation process, reviewers deal comprehensively with the evaluated entity and break away from indicators which can only be used as proxies. However, the results of Auspurg, Diekmann, Hinz, and Näf (2015) reveal the problem with this ideal world. Auspurg et al. (2015) studied the correlation between experts' assessments of German institutions in sociology and different indicators (e.g., the number of peer reviewed journal papers). The experts in the peer review process knowingly abstained from indicators to evaluate the institutions (e.g., by reading several publications). However, the results of Auspurg et al. (2015) show that "the judgment of research quality is overwhelmingly influenced by output measures such as the number of publications in peer review journals. To a large extent the judgments can be predicted by few quantitative indicators" (p. 177). There are two possible explanations for the results: either the reviewers nevertheless base their scores on indicators, or the indicators correlate with the assessments by peers.

As our response indicates, we have many disagreements with Peters (2017). We think he has mischaracterized many of our arguments, and that the methods he criticizes are, while perhaps not perfect, strongly defensible. Nevertheless, we find his comments valuable. We attempt to address many questions that are not easy to answer, and critiques such as Peters' help us to consider whether other approaches would be superior. While his position seems to be a bit more extreme than ours, Peters (2017) also reinforces our critical point that many factors besides early publication success should be considered when allocating future resources.

## References


Auspurg, K., Diekmann, A., Hinz, T., & Näf, M. (2015). The research rating of the german council of science and humanities: Revisiting reviewers' scores of sociological research units. *Soziale Welt-Zeitschrift Fur Sozialwissenschaftliche Forschung Und Praxis*, *66*(2) [177-+]

Bornmann, L., & Williams, R. (2017). Can the journal impact factor be used as a criterion for the selection of junior researchers? A large-scale empirical study based on ResearcherID data. *Journal of Informetrics*, *11*(3), 788–799. http://dx.doi.org/10.1016/j.joi.2017.06.001

Bornmann, L., Marx, W., Gasparyan, A. Y., & Kitas, G. D. (2012). Diversity, value and limitations of the journal impact factor and alternative metrics. *Rheumatology International*, *32*(7), 1861–1867. http://dx.doi.org/10.1007/s00296-011-2276-1

Fraley, R. C., & Vazire, S. (2014). The N-Pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *Public Library of Science*, *9*(10) http://dx.doi.org/10.1371/journal.pone.0109019

Gould, R. V. (2002). The origins of status hierarchies: A formal theory and empirical test. *American Journal of Sociology*, *107*(5), 1143–1178. http://dx.doi.org/10.1086/341744

Gregg, A. (1957). *For future doctors*. Chicago, IL, USA: University of Chicago Press.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The leiden manifesto for research metrics. *Nature*, *520*(7548), 429–431.

Hug, S. E., Ochsner, M., & Daniel, H.-D. (2014). A framework to explore and develop criteria for assessing research quality in the humanities. *International Journal for Education Law and Policy*, *10*(1), 55–64.

Leydesdorff, L., Wouters, P., & Bornmann, L. (2016). Professional and citizen bibliometrics: Complementarities and ambivalences in the development and use of indicators—A state-of-the-art report. *Scientometrics*, *109*(3), 2129–2150. http://dx.doi.org/10.1007/s11192-016-2150-8

Martin, B. R., & Irvine, J. (1983). Assessing basic research – Some partial indicators of scientific progress in radio astronomy. *Research Policy*, *12*(2), 61–90.

Merton, R. K. (1988). The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *Isis*, *79*(4), 606–623.

Peters, G.-J. Y. (2017). Why not to use the journal impact factor as a criterion for the selection of junior researchers: A comment on Bornmann & Williams (2017). *Journal of Informetrics* (in press).

Lutz Bornmann*
*Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany*

Richard Williams[1]
*Department of Sociology, 4060 Jenkins Nanovic Halls, University of Notre Dame, Notre Dame, IN 46556, USA*

* Corresponding author.
*E-mail addresses:* bornmann@gv.mpg.de (L. Bornmann), Richard.A.Williams.5@ND.Edu (R. Williams)
[1] Web Page: https://www3.nd.edu/~rwilliam/.

14 August 2017

17 August 2017

17 August 2017
Available online 1 September 2017