Journal of Informetrics 10 (2016) 658–660

Contents lists available at ScienceDirect

Journal of Informetrics

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

Productivity, performance, efficiency, impact—What do we measure anyway? Some comments on the paper "A farewell to the MNCS and like size-independent indicators" by Abramo and D'Angelo

1. Introduction

Abramo and D'Angelo (2016) publish a critical study on size-independent citation indicators in which they address two important issues, firstly the fallacy of interpreting (size-independent) citation indicators as measures of research performance, and secondly the inappropriateness of such indices for ranking exercises. Taking a strict and focused perspective on the indicators only, both statements are as such correct. But Abramo and D'Angelo are not the first to articulate these concerns. However, the authors' argumentation, which is based on a simple economic interpretation of research productivity and efficiency, results in their suggestion to eliminate so-called size-independent citation measures and to replace those by indicators that take also input measures, in particular, measures of research investment into account. Finally, they propagate a paradigm shift towards "true measurement of research efficiency" and formulate several recommendations addressed to various actors and stakeholders for the use of performance measures in this context. This endeavour is certainly guided by the best intentions, however in the course of the argumentation and the development of their new strategy, the authors fall into the same trap that they basically intended to dismantle. In what follows, we will point to some problematic issues we have found in their study. First, we discuss the arguments put forward by the authors, next the examples they provided are examined and last, we comment their recommendations.

2. Discussion

As already mentioned in the introduction, we basically agree that citation measures, in general, and normalised or size-independent citation indicators, in particular, cannot by themselves reflect the complexity of research performance. Nonetheless, publication and citation measures do – if properly applied – measure important aspects and dimensions of research productivity and impact in a statistically reliable manner, notably if these are based on appropriate mathematical models and sufficiently large and preferably longitudinal data sets. It is up to the bibliometricians' responsibility to choose indicators cautiously and to inform readers and users of bibliometric results about their correct interpretation and where necessary about their limitations. Thus there is no reason to throw the baby out with the bathwater and to condemn "classical" citation indicators as such. Unfortunately Abramo and D'Angelo mistakenly assume the equation of classical moment, share or percentile based citation indicators with performance measures and conclude their invalidity as they "violate an axiom of productivity theory". However, productivity and performance are not the same and an increased production of unmerchantable wastrel might be a sign of high productivity but certainly not of high performance, efficiency or even competitiveness. Just producing more or the same with less resources is probably not the most suited goal in a model of science systems. As economists know, this challenge is at the heart of the efficiency—effectiveness debate. System performance is about the effectiveness of a system. Productivity only hints at its efficiency. Definitely when it comes to the science system, effectiveness should be the main concern. In addition, efficiency or productivity also poses its specific challenges in terms of both operationalization and quality of the underlying data. This observation is important because the authors first correctly suggest the inclusion of input measures to capture efficiency but, again, this is then implemented in the same way as they criticise in the context of normalised citation measures. The FSS indicator group proposed by the authors in an earlier study (Abramo & D'Angelo, 2014) is designed as a normalised citation rate, similarly to MNCS and related indicators, but now with one financial component. Thus FSS suffers from the same flaws as the criticised MNCS, namely that indicator values cannot be decomposed by components, i.e., in this case by research investment (expressed by salary), publication output, observed

http://dx.doi.org/10.1016/j.joi.2016.04.008

and expected citation rate (restricted to cited papers).[1] In addition and beside this methodological remark, the quality and comparability of the type of financial data used by the authors across levels and units of analysis is questionable as well. Salary and investment financial structures differ hugely between countries, and salary levels differ hugely between functions, organizations and countries. To paraphrase Belgian surrealism: a salary is not a salary, while a research investment is not a research investment. Comparability (and hence validity) of the underlying data themselves not only is a challenge, it is a problem.

As the authors illustrated, their design can and is applied to ranking of individuals, departments and institutions. In a recent study Glänzel and Debackere (2009) have discussed the "multi-dimensionality" of ranking and the role of bibliometrics in university assessment. Without loss of generality one could say that in the case of complex systems, such as the subjects of analysis of bibliometric studies, whether merely one single aspect – representing only one dimension of the underlying multidimensional space – can be chosen for the ranking, or a composite measure resulting in a projection on one dimension must be built. They concluded that "the difficulty of quantification as well as the all too frequently experienced arbitrariness in defining composite indicators often results in inadequate representation and irreproducibility." This implies that both essential information and the demand for adequate depiction of complexity is lost, whenever variables representing different dimensions are mathematically combined in a non-decomposable way. The stated complexity of the indicator proposed by Abramo and D'Angelo cannot disguise that this measure is subject to the mentioned effect as well.

In a following step Abramo and D'Angelo compare ranking by FSS and MNCS as argumentative support for their paradigm shift. Table 2 refers to individual scientists. Wouters, Glänzel, Gläser, and Rafols (2013) discussed the assessment of research output of individual researchers and formulated 10 don'ts and dos in individual-level bibliometrics as pars pro toto. One of the "dos" referred to the fact that bibliometrics is statistics also at this level of aggregation, which means that the requirement of a sufficiently large dataset for the calculation of indicators applies to all paradigm systems. We just mention in passing that in this context the second example in Table 2 is not the best recommendation.

The paper is concluded by a number of recommendations, which are, however, problematic. In particular, the first one refers to the already mentioned paradigm shift and is as such formally justified but the text in the corresponding bullet point has practically no content link with this demand. Frankly speaking, the second recommendation addressed to scientific editors sounds almost like a demand for censorship and confinement of the freedom of research. The criticism of current ranking practices formulated in the third and fourth recommendation is, as already mentioned, partially justified but applies, as we said to all multi-dimensional or -variate cases, not specifically just to normalised citation indicators. This makes the idea of the development of a distortion-free or, otherwise, unbiased ranking remains a dream. The advice of using a possibly fine-grained subject classification scheme along with unique assignment of actors is in practice quite remote from reality. Granularity of cognitive assignment and measurement in bibliometrics – and more generally in information and library science – is by far too coarse. Besides, the classification of researchers to one and only one field completely neglects the possible individual multi- or interdisciplinary profile of scientists (also see Wouters et al., 2013). The incommensurability of sectoral and national salaries and the impossibility of objectively crediting individual co-authorship might serve just as further examples of this problem in the context of building FSS-type indicators. And last, both individual classification and financial indicators are very sensitive to inconsistencies and even to manipulation. Researchers might adopt their publication strategy to acquire the lowest possible baseline to achieve the maximum "performance".

3. Conclusions

To conclude, maybe time is ripe indeed for a paradigm shift and bibliometrics needs to be opened to broader scopes and contexts; application to social sciences, arts and humanities, the web, the search for new and alternative metrics and the measurement of societal impact of research are certainly signs of an upcoming new age in our field too. And this new age requires also new models and tools. But replacing one imperfect tool by another one, even if guided by the best intentions, does not help meet this challenge.

References

Abramo, G., & D'Angelo, C. A. (2014). How do you define and measure research productivity? *Scientometrics*, *101*(2), 1129–1144.
Abramo, G., & D'Angelo, C. A. (2016). A farewell to the MNCS and like size-independent indicators. *Journal of Informetrics*, *10*(2), 646–651.
Glänzel, W., & Debackere, K. (2009). *On the multi-dimensionality of ranking and the role of bibliometrics in university assessment*. In C. Dehon, D. Jacobs, C. Vermandele (Eds.), Ranking Universities, Editions de l' Université de Bruxelles, 65–75.
Wouters, P., Glänzel, W., Gläser, J., & Rafols, I. (2013). The dilemmas of performance indicators of individual researchers—an urgent debate in bibliometrics. *ISSI Newsletter*, *9*(3), 48–53.

Wolfgang Glänzel*
Bart Thijs
Koenraad Debackere

[1] Although one can detect conceptual issues in the definition of FSS, our comments will not extend to the design of this indicator because this is not the subject of the present study.

Center for R&D Monitoring (ECOOM) and Dept. MSI, KU Leuven, Belgium

* Corresponding author.
E-mail address: wolfgang.glanzel@kuleuven.be (W. Glänzel)

7 January 2016
Available online 4 May 2016