

Unholy goals and flawed methods

By Professor Charlotte Wien and Library Director Bertil F. Dorch

Abstract

A problematic practice has evolved, which is threatening to undermine research in the social sciences and humanities. Bibliometrics is often claimed to be able to measure researchers' efficiency. We find this quite problematic and, in this article, we illustrate this point by discussing two different bibliometric practices. One is the so-called h-index, the other the so-called BFI-points (Den bibliometriske Forskningsindikator, The Bibliometric Research Indicator). The BFI was never intended to be used for evaluating individual researchers and their productivity. Yet since its introduction in 2008 especially the social sciences and the humanities experience a pressure to deliver "BFI points" and academic job advertisements within the social sciences and the humanities increasingly mention expectations for people's past and/or future production of BFI points.

The h-index is even more problematic because no one academic database covers all the research publications in the world. The whole thing is completely disorganized, and as many as five different h-indexes exist for each researcher. What makes the h-index even more useless is that it will not let you make comparisons across disciplines. Furthermore, like other simple measurements, it is liable to be manipulated and misinterpreted. On that background, it is remarkable that numbers extracted from incomplete databases are used for describing the quality of researchers and their institutions.

Keywords:

Bibliometrics, h-index, the Bibliometric Research Indicator, research evaluation

A problematic practice has evolved, which is threatening to undermine research in the social sciences and humanities in Denmark, and which has managed to infiltrate virtually the entire academic world. The name of the practice is bibliometrics, and it is often claimed to be able to measure researchers' efficiency. That would be excellent, if it were possible, for who in this kingdom would not like to know what we get out of that part of the taxpayers' money that we spend on research? Besides, the idea of measuring fits perfectly into the dominating paradigm of management in the public sector. Everything must be measured, managed, made more efficient, and documented. But things do not really work like that. Ideas that have the potential to change the world cannot be reduced to key figures in a spreadsheet. What is worse is that most of the measures that are used for the purpose are either completely unsuitable for it, or were created for an entirely different purpose, or they are based on insufficient data, which makes the measurements useless.

As a tenured researcher within the humanities or social sciences, you might shrug your shoulders and claim not to care, with a fine display of academic self-confidence and a little bit of arrogance. "After all, quality is the important thing, not quantity". This seems to be how the social sciences and the humanities have survived the initial waves of bibliometrics for the last ten or fifteen years. But the problematic practice has grown and is now turning up in various new connections where it is becoming more and more difficult to ignore it. And if it grows really powerful, it will be very difficult to get rid of, for it is being nourished by strong economic interests both within the universities and from big international businesses.

In the world of Danish research, the practice is found in two important incarnations. One is the so-called h-index, the other the so-called BFI-points (Den bibliometriske Forskningsindikator, The Bibliometric Research Indicator).

To begin with the latter: when New Public Management became a dominating way of thinking in Danish universities in the mid-2000s, the BFI was created with the intention of helping to distribute research money between the institutions. The idea was to reward the highest degree of "productivity". The model came from Norway and was implemented as a system in which researchers in various fields evaluate and establish a hierarchy of scientific journals and publishers according to their perceived quality. When a researcher published something, his university would be awarded a certain number of points, which were then converted into funding for research, provided that the journal or publisher had been included on the "authority list". This system was never intended to be used for evaluating individual researchers and their productivity. Yet since its introduction in 2008, the BFI has been assessed repeatedly, and each time, researchers in especially the social sciences and the humanities experience a pressure to deliver "BFI points". And this is not surprising, when these areas are constantly being cut down, and BFI points mean cold cash. Economic incentives tend to be effective. We know of research institutions which

openly tell researchers that they are expected to deliver a certain amount of BFI points each year.

Just for fun, we have been keeping an eye on academic job advertisements within the social sciences and the humanities, and we have found that it is no longer unusual for advertisements to mention expectations for people's past and/or future production of BFI points. This cannot help influencing the way junior researchers plan their careers. A forthcoming study by Deutz et al. of how junior researchers at SDU within the social sciences and the humanities choose to publish their research demonstrates quite clearly that they are influenced by the BFI. They are more and more focused on finding journals and publishers which will yield the desired points. We also see that calculations of BFI points are part of the numbers that are used for analysing the research coverage of university courses for the purpose of the accreditation of the institutions. As we all know, the problem with having a target is that you tend to aim for it. And it is not a sign of healthy research when a researcher aims at achieving a high BFI score. Increasing researchers' BFI scores was never the intended end goal – for that would determine where and how researchers publish their results – the purpose was rather to measure the perceived quality of the research output, and devise a scheme to direct funding on that basis. The quality of the data produced by any measurement, and the potential for basing an accurate analysis on them, are however damaged when the object of the measuring keeps adapting to the method of measuring. Besides, the authority lists are based on opinions. Nobody has ever yet managed to come up with a watertight definition, guaranteed to work every time, of how you determine the quality of a journal or publisher. And here we have the Achilles heel of the BFI system, and the reason why it should never be used for evaluating individual researchers, but only for distributing funds.

A second incarnation of bibliometrics, the so-called h-index, is even more problematic. The h-index was “invented” by the physician Jorge E. Hirsch in 2005, and since then it has been a success throughout most of the academic world. It is a very simple measurement, based on a combination of the number of texts published by a researcher and the number of times that researcher has been cited by others. You simply list the publications according to the number of citations that have been made to each, and then you go down the list until the number of citations matches the place on the list. Simple and easy. And yet not quite as simple as that.

Because no one academic database covers all the research publications in the world. Some databases cover some publications, others cover others, some include citations, others do not. The whole thing is completely disorganized, and at the University Library of Southern Denmark we need to use as many as five different h-indexes for each researcher, knowing perfectly well that none of them is the correct one.

What makes the h-index even more useless is that it will not let you make comparisons across disciplines. Each field of study has its own traditions for publishing. In the humanities, researchers typically write fewer, longer texts, each with a single author. In the natural sciences, the typical product is a shorter text with many co-authors. On top of that, the international academic databases mainly cover literature in English – literature in the Scandinavian languages within the social sciences or the humanities is rather hard to find. This puts these disciplines at a disadvantage and means that their researchers will typically have much lower h-indexes than those from the disciplines of natural, engineering or health sciences.

But, you may argue, then we will just have to stop comparing people across the disciplines. Unfortunately, things are not as easy as that: the measures are already being used across disciplines. For instance, calculations of researchers' h-indexes are used as a factor when international ranking companies like QS work out the relative rankings of universities. And this is important to the universities: a good position on an international ranking list makes it easier to attract international students, who will bring an income, and star researchers, who may improve the rankings even more.

But then, you might say, why don't people in the humanities adopt the publication traditions of the natural sciences? The answer is that these days, they do. Several studies show that research articles in the humanities list co-authors more and more often, and that at the same time there has been a regular explosion in the number of articles from these disciplines in highly ranked international journals. And though this sounds, on the surface, like a positive development, we will permit ourselves a certain amount of scepticism. There is a very big difference between the exact and the interpreting disciplines and the role they play in society, and this means both that they target very different audiences and that there is a big difference in the media outlets which will suit their purposes. Of course, a discussion of e.g. the development of the Danish welfare state can be conducted in high ranking, international double-blind peer reviewed journals, but this will mean that only very narrow circles of researchers can participate in the discussion, while a good book on the topic will reach a much wider readership.

To get back to the h-index: we are not the first to criticize its use, and by now this discussion is an old one. Like other simple measurements, it is liable to be manipulated and misinterpreted. And obviously, when researchers' h-indexes become an indicator on the list of the ranking firms, it gives research managers an incentive for focusing on it when they plan recruitment and promotions. At the University Library of Southern Denmark, we have attempted to get an impression of Danish universities' recruitment strategies by looking at how often the word "h-index" occurs in Danish academic job advertisements. The h-index is celebrating its fifteenth anniversary this year, and during its first ten years it either did not appear

at all or was mentioned once or twice each year. It appeared 8 times in 2015, 63 times in 2016, and from 2017 onwards it is mentioned in more than 150 job advertisements each year.

The alert and knowing reader may point out that there are several methods for measuring things, and that the more you include, the more sophisticated your measures must become. According to this logic, you should be able to create a more comprehensive and coherent picture of researchers by simply using more measures, e.g. altmetrics, which is meant to measure the impact on society of a researcher's production. Our answer to this is that most bibliometric indicators are relatively easy to game, and that they usually reproduce traditions and thinking from the natural sciences, which may easily result in biased incentives for both researchers and research institutions. For if the object of a measuring – in this case, the researcher – adapts his behaviour to the measuring instrument – in this case, the h-index or BFI points – neither the measuring instrument nor its data will be very precise. In fact, the whole process will become unscientific.

At Danish universities, it has been acknowledged for a long time that the students' potential will only be reflected to some extent by their grades, and therefore tests and interviews are used for recruiting students. On that background, it is remarkable that numbers extracted from incomplete databases are used for describing the quality of researchers and their institutions.

References

Originally published in Weekendavisen (in Danish, toll access) on September 2, 2020.

Daniella Bayle Deutz, Thea Marie Drachen, Dorte Drongstrup, Niels Opstrup og Charlotte Wien (2020) “Quantitative quality: A study on how performance-based measures may change the publication patterns of Danish researchers”, forthcoming

About the authors

Charlotte Wien (<https://orcid.org/0000-0002-3257-2084>)

Professor, Head of Research (Denmark)

Charlotte Wien is professor of Scholarly Communication at SDU/Denmark and Head of Research at the University Library of Southern Denmark and Professor2 at UiT The Arctic University of Norway. Her research focuses on researcher integrity and responsible use of scholarly metrics. She teaches scientific methods, philosophy of science and research integrity. Before she was a full professor of Communication and media science. Her current research team is a cross disciplinary unit covering both STEM and SSH. She is a board member of SPARC Europe and a very active member of LIBER (Ligue des Bibliothèques Européennes de Recherche). Dr. Wien is a frequent contributor to mass media on all kinds of topics related to Scholarly Communication.

Bertil F. Dorch (<https://orcid.org/0000-0003-2594-6778>)

Associate Professor, Library Director (Denmark)

Bertil F. Dorch is Director of the Research and University Library at SDU and teaches responsible conduct of research. Dorch is affiliated to the Department of Physics, Chemistry and Pharmacy at SDU as Associate Professor. Dorch’s professional interests include topics within stellar astrophysics and magnetism, scholarly communication, responsible conduct of research, strategical management, scholarly publishing, especially within the natural sciences. Dorch was three times elected President of the Danish Research Library Association during the period from September 2014 to March 2020.