Nostalgia for the world without numbers

Von Peter Weingart

Abstract: The argument between propagators and opponents of quantitative performance measures in research and higher education is at a stalemate. On the one hand are those who promote enthusiastically quantitative performance measures, not giving much thought to issues of misrepresentation and unintended consequences. On the other are the 'essentialists' denying the admissibility of transforming qualitative assessments into quantitative ones and reverting to a fundamental argument about the nature of science in general and the university as its core institution in particular. In between are the 'pragmatists' who occupy a strategic position in being the forward oriented vanguard combining reflective analysis and shaping the technology of indicators and their applications. I argue that the confrontation of promoters and essentialists is missing the point. Blind belief in the technology of numbers is as misplaced as its outright rejection that does not recognize the strength of social change driving it. In fact, individual scientists, universities and research institutions, large scientific publishing companies as well as science policy and bibliometricians are entangled in a tight arrangement in which quantitative measures have become the central currency and everyone profits from dealing with it in some way. Control by numbers is a social technology fired by digitization and has replaced trust in institutions. The old world of academia is thus past. There is no alternative to the pragmatists' efforts to shape that technology.

1. Discourse stalemate

For some time now, the debate over pros and cons of performance measures in higher education and research institutions has turned around in circles. Specifically, no new evidence of unintended consequences has been put on the table (which is not to say that such evidence could not be produced by appropriate research efforts). The arguments being exchanged between propagators and opponents are repetitious. In fact, the underlying structure of the dispute has not changed ever since the first introduction of quantitative indicators of the research process at the beginning of the 1980 s (Weingart 2005). On one side are those who promote enthusiastically quantitative performance measures, not giving much thought to issues of misrepresentation, unwarranted simplifications and unintended consequences. They argue for progress and claim to represent it. On the other side are the 'essentialists'. Their principle denial of the admissibility of transforming qualitative assessments into quantitative ones is coupled with a fundamental argument about the nature of science in general and the university as its core institution in particular. Their critique goes far beyond technical details – e.g. why the journal impact factor is a flawed measure of the quality of individual articles – by attacking the "cultural flattening of market economic imperatives", i.e. the replacement of what used to be the specific value of the pursuit of knowledge by the single criterion of economic utility (Burrows 2012: 1). Variants of this position deplore the power shift from individual scientists to university administrations, or they warn of a loss of diversity of research as the basis of its innovativeness and an increasing pressure to mainstreaming (Martin / Whitley 2010). They are fundamentally opposed to indicators. Those who are indifferent – probably the large majority – and suffer or profit quietly don't let us share their thoughts.

In situations of incommensurable arguments it is most productive to change the perspective altogether. I want to give reasons why in spite of all criticism the resistance against quantitative performance measures is based on futile arguments, which explains why it is and probably will be in vain. To be sure, many if not most of the more technical critiques of

Soziale Welt 66 (2015), S. 243 – 250

the indicators are justified.[1] In fact, between promoters and opponents is the group of 'pragmatists', i.e. researchers who are agnostic with respect to the possibility of expressing the qualitative in quantitative terms. Their empirical studies are focused on the claimed effects of incentivizing indicators on publication behavior (e.g. redundant publishing) the results of which leave room for both, success stories (scientists' productivity has increased) and skepticism (if growth can be established, not all of it is substantive). They readily concede that indicators are selective, and that the gold standard of their application is their use in a model termed 'informed peer review' (e.g. Butler 2008, 2010; Osuna / Cruz Castro / Sanz Menéndez 2011). It is probably no accident that the pragmatists don't raise questions about the possibility or impossibility of measuring quality. After all, indicator construction presupposes just that. So, basically they are in favor of indicators. But, as I will try to show, they occupy a strategic position in being the forward oriented vanguard combining reflective analysis and shaping the technology of indicators and their applications. My argument will be that the confrontation of promoters and essentialists is missing the point. Blind belief in the technology of numbers is as misplaced as its outright rejection that does not recognize the strength of social change driving it. Instead, I will argue in favor of the pragmatists' role of engineering. As a first step it will help to raise some questions: What are the motives of propagators and critics? What future do they have in mind?

2. Motives and interests in the régime of quantitative performance measures

Turning first to the propagators, one can distinguish three subgroups. In the first subgroup we identify what could be called the data collectors and processors. Thomson Scientific in Philadelphia, the successor of the parent institute of all indicator production, can be counted in this category. It is owned by Thomson Reuter, a Canadian media concern that operates globally and with its *Web of Science* has virtually monopolized the production of bibliometric data used to construct indicators. It is only challenged by one other media concern, Elsevier of the Elsevier Reed group, that produces *Scopus*, "the largest abstract and citation database of peer-reviewed literature".[2] It is these two companies that share a global oligopoly of data production, processing and dissemination which is the basis of all performance measures in higher education and research institutions wherever they are applied. The above mentioned 'pragmatists' are dependent on their data, they clean and customize them in order to serve the needs of particular clients. They are the retailers in the data business while the oligopolists are the wholesalers. The retailers have to buy their data bases for hefty fees.

Bibliometric data have become a highly successful business model, especially for Elsevier, due to a growing demand from governments and university administrations worldwide. One just has to look at university rankings, which have become a global obsession and have given rise to enterprises such as the Center for World-Class Universities (CWCU) of Shanghai Jiao Tong University that produces the Shanghai University ranking. Rankings based on bibliometric data have developed into an industry whose business model, closely tied to the data producers, is not going to lose its customers any time soon.

The second group propagating bibliometric indicators is constituted by the scientific community itself, or rather parts of it. One may establish the following equation: the more extensive the coverage of a discipline in *Web of Science* journals and the higher the remunerations based on bibliometric performance measures the greater the enthusiasm and the stronger the support of these measures. After the German Science Council had carried out its pioneering

1 In the meantime these critiques have also been voiced by international scientific associations and have led to official protests or critical declarations (cf. DORA; IMU 2008).

2 See online at: http://www.elsevier.com/online-tools/scopus, last access: 1.7.2015.

nationwide evaluation of chemistry and sociology, the chemists declared satisfaction with the indicators applied whereas the sociologists asked for various supplementary measures. The council's effort to add history to the process failed altogether due to the historians' resistance. This is exemplary. In the hard sciences and medicine two conditions reinforce each other: coverage of journal articles in *WoS* is very good, so that citation counts cannot be disputed very well, and the amount of money to be distributed on the basis of performance indicators is considerable. Department chairmen of medical departments like to defer budget decisions to indicators because it helps to avoid personal internal conflicts.

What holds for the attitudes of disciplines is true for individual scientists as well. Since its introduction, grown up, intelligent scientists, acting much like teens in the facebook mode, advertise their h-index on their websites. Similarly, the impact factor of journals (JIF) is taken as a symbol of distinction attached to individual articles even though the connection is tenuous at best. But the perception is one of high quality. In other words, the conspicuous exposition of quantitative measures is not only widely accepted especially among scientists in the STEM disciplines but has become common currency in documenting and exhibiting reputation. Individual vanity is a strong force on the science fair. Even more disturbing is the observation that above all young and successful scientists support the new currency while older and less visible scientists tend to be more skeptical. This seemingly paradoxical development is fired both by technical innovation such as the social media and altmetrics and by links to monetary incentives. The former facilitates a fast and far-reaching communication of the symbols of reputation, the latter ensures quick monetary satisfaction. Compare this to the slow and haphazard processes of attributing reputation by peer review to explain the attraction that the measures have especially for young scientists. The numbers provided by citation counts and impact factors promise just and fair competition beyond the reach of entrenched gate keepers and their ossified hierarchies. In the new academia a double reward is to be gained: reputation *and* money, not just reputation. Performance measures have achieved the miracle of tying tightly the attribution of reputation internal to science to the distribution of monetary resources and remunerations. Academia is no longer determined by immaterial judgments of quality alone but by money as well. Status by attribution of excellence and status by budgetary clout have become largely indistinguishable.

The third and last group contributing to the establishment of quantitative performance measures is that of policymakers and administrators. Policymakers have, after some time of skeptical reluctance, embraced the new instruments of performance measures as part of the 'new public management' (NPM). With science budgets pushing the 3% GDP limit in some countries and hovering around 2.5% in most others, there is direct as well as indirect pressure to demonstrate 'value for money'. All across Europe policies have shifted from 'science' to 'innovation' as their chief objective, signaling a closer link between research and economic benefit. In the context of this régime it is no longer feasible to rely on the promises of the academic community that their governance of self-regulation will ultimately pay the expected dividends, and that they know best how to achieve that. The entire science system, universities and research laboratories alike, has been subjected to both direct and indirect pressures of 'economization'. More interesting than the former (e.g. patenting, MTAs, IPRs) are the latter: through the introduction of rankings, the promotion of competition between entire universities, let alone between individual scientists, and various performance measures in conjunction with monetary remunerations for chairs, institutes or personal income the system has been transformed to a market logic. Government policies have pursued and succeeded to implement a "shift from a model of science based on the philosophy of the public domain to one leaning toward notions of proprietary ownership and control" (Rhoten / Powell 2007: 355).

The latest development – not so recent anymore – is the introduction of ad hoc selective funding initiatives designed to entice universities to enter national competitions. China, Japan, Korea, Denmark, UK and notably Germany with its 'excellence initiative' have followed this strategy. Although robust data of outcomes are still hard to come by, the impression of science policymakers and analysts in Germany is that overall it was a success. Regardless of measurable success or failure, however, more relevant is the sheer magnitude of the effort. Although the German initiative involved peer review, it is no secret that the international reviewers in many cases could not do justice to the extensive documentation provided by the contending universities simply by 'reading' them. It will remain a secret how many assessments were quietly done on the basis or at least with the help of quantitative indicators. In other words, the more selective, comprehensive and penetrating the initiatives become the more demanding are the evaluation processes they require both in terms of reviewer time and specialized expertise. One of the lessons of the 'Excellence Initiative' is that it cannot be repeated another time in the same way, for precisely that reason.

The science and higher education policy constitutes the framework and the conditions within and under which the administrators, university presidents, deans, department chairmen and institute directors have to operate. Driven by the insatiable need for funding they respond to opportunities of gaining resources offered by governments. But these are now tied to means of verification. University administrations now act as mediators between governments on the one hand and the members of their respective institutions on the other. Legitimating their decisions by referring to quantitative performance measures is in many cases mandated by law or ministerial decrees. It makes their job much easier by allowing them to make distributive and structural decisions which before would have been vetoed by penalized departments. The blame for defeat in the competition for funding can now be deferred to the quantitative and seemingly objectified evidence of the performance measures. It is not likely that methodological critique or excuses with special circumstances that justify exception will carry very far. Any such argument is tainted by self-interest. Perhaps the most important function of quantitative measures is that they de-legitimate the qualitative assertions of academic excellence in the context of pompous 'old boys networks'. They are instrumental in shifting power and influence.

All this is irrespective of real and/or assumed methodological weaknesses of particular measures. It does not matter if they misrepresent certain aspects of university processes and products. Due to the fact that they are accepted and used as the common currency by all parties in the system they shape the reality of that system.

3. Control vs. trust

The introduction of performance measures is not just the fleeting idea of some number crazy bureaucrats who want to make life difficult for 'lazy professors' but rather the expression of a much more fundamental development that reaches the last resort of self-regulation hitherto condoned by society: academia. For some time, before measuring quality by indicators was introduced into the world of academic research, it had been applied to practically all areas of public administration – the shift from input legitimation to output accounting. This delay may be due to the fact that public trust in science as an institution has remained high in comparison to other institutions (Peters et al. 2007). Another reason is probably the fact that science, in contrast to many other policy fields, is inherently 'inaccessible' to the lay public and its representatives in parliaments due to its highly specialized languages, methods and theories. The lack of knowledge and understanding of how science 'ticks' and how to decide competently about the means with which to steer it in desired directions makes it dependent on judgments of scientists themselves. In no other policy arena is the dependence of the reg-

ulators from the judgments and information of the regulated, of the principal from the agent, greater than in science policy. "'For that reason »science policy is largely played out as science budget policy«' (Sarewitz 2007: 32), while the capacity to undertake centralized, strategic science policy planning is mostly lacking" (Weingart 2013: 87). If those are the ground-rules by which policymakers had to play so far, their sustained effort to gain control over this area is served ideally by quantitative performance measures. They promise to provide a more fine-tuned steering instrument, if not to set priorities at least to define focused conditionalities for the behavior of scientists and universities. "Performance measures are basically the necessary correlate to a science policy that can only shape the context of the research system in order to influence it" (Weingart 2013: 87).

Although the circumstances in science policy may differ from those in other policy arenas, it should be clear that the gradual replacement of trust by control is a secular development that Power has termed the advent of an 'audit society' (Power 1997). This development has not exempted the institution of science, and it is not likely to be stopped, let alone reversed, in the foreseeable future because it rides on the wave of very strong values. The introduction of performance measures comes under the label of 'transparency' and 'accountability' – fundamental values in democratic societies. NPM has two aspects which are both related to legitimacy: democratic control by the public through securing transparency of internal practices, and efficiency through management designed to save public funds. Thus, NPM is supposed to re-establish 'legitimacy' to publicly funded institutions (Weingart 2013: 85). The issue of control vs. trust reaches far beyond the realm of science. It affects the fabric of democratic societies, the philosophy underlying their institutional structures and the governance by which they are operated.

4. Nostalgia is for sissies – academia in a new epoch

I have started the argument with characterizing the discourse on quantitative performance measures in academia as the debate between 'promoters' and 'essentialists' over the issue if quality in science can be measured. As it turns out, that debate is inconclusive, the opponents talk past each other. But more importantly, that is not even the crucial issue. The triumph of quantitative indicators in spite of justified critique of their shortcomings and unintended prejudicial consequences is explained by the emergence of an interrelated network of mutually reinforcing factors. This network is constituted by publishers and data management concerns, by governments and university administrations using performance measures to improve the effectiveness of their policies and budgetary decision making, by scientists who are driven by severe competitive pressure and vanity and strive for fairness and transparency in the attribution of reputation, and finally by subsidiary academic institutes acting as research and application agents supporting and improving the implementation of the measures. This tightly knit network is welded together by reinforcing confluence of interests that are stronger than the misgivings some may have about the negative consequences of these measures. One must add to all this two supportive factors. One is the added legitimacy through reference to strong values like transparency, fairness and accountability to the public. The other is literally the propellant that adds velocity and penetrating power to the domination of indicators:

- a rapidly developing technology propelled by digitization and new algorithms that make the construction, widespread use and constant improvement of performance measures possible in the first place,
- the continued innovation and expansion of the data industry, starting with the publishers' editorial management systems which capture the relevant data of the publication

process in journals, continuing with the translation and distribution of these data into 'ready-made' tool-kits such as the Web-of-Science,

- and, further on, the capture of the academic publication process by the social media such as Google Scholar and Research Gate that lure scientists to check their h-Index, their citation score or their RG score daily and with a precision and reach that exceeds anything ever even imagined just a short while ago.

The logic of the fusion of new digital technology and scientists' behavior is poignantly illustrated by Enserink (2009): "Are You Ready to Become a Number?", he entitled an article describing recent efforts to establish unambiguous identification in scholarly publishing. The solution is to attach a number to everyone:

> "A universal numbering system could aid scientists trying to stay on top of the literature, help universities more readily track staff productivity, and enable funding agencies to better monitor the bang they're getting for their buck." (Enserink 2009: 1662)

Schemes like Researcher ID and ORCID have already been put in place and ideally fuse the academy with the economy: "It's the creation of digital information company Thomson Reuters, which hopes to enhance the value of its paid services" (Enserink 2009: 1662). These services rely on the cooperation of the scientists. They have to provide the necessary input.

> "For such a voluntary disambiguation system to work, scientists need an incentive, however. Researcher ID's carrot is that scientists can analyze citations to their papers, or place a 'widget' on their web site or blog that automatically retrieves a list of their most current papers whenever someone clicks on that page." (Enserink 2009: 1663)

This is the very logic of the digitized world: the possibility of monitoring the impact of your own actions is probably nowhere else a stronger temptation than in a social subsystem where impact is geared to status and at the same time used to be elusive and indirect.

Quantitative performance measures are a new *social* technology and a byproduct of digitization as another new technology, which has already revolutionized the entire world. Digitization is such a revolutionary power because it allows people and organizations to become reflective and monitor their own actions, thus fine-tuning their control. Like every other successful technology it entails gains and losses, but it would not be successful if the losses would outnumber the gains. To try to stop it by reverting to a supposedly ideal past is unrealistic. The only option is to improve performance measures even if that means accepting their logic.[3]

This is the business of the reflective pragmatists, the *craftspeople*, the bibliometricians, i.e. all those scholars who have made the construction and application of quantitative indicators their main research endeavor and/or who are running institutes whose business are evaluations based on performance measures such as CWTS in Leiden, IfQ in Berlin or CREST in Stellenbosch. Some are more academic than others but all of them are involved in research on the construction and use of indicators for clients who pay. One cannot expect from them that they will question the applicability and utility of their products in principle, but they are interested to *improve* them if that will help their sales. The more academically minded engage in research oriented discussions about technical improvements. In other words, they know best why and how the indicators may misrepresent what they are supposed to measure, how they should be applied and how they should not be applied, and above all, how their unintended consequences may be avoided. They, if anyone, are the

3 This is also the logic that underlies the UK's government review committee that looked into the uses and limitations of quantitative indicators. Cf. Wilsdon 2015; Wilsdon et al. 2015

voice that will be listened to by those who would otherwise continue to apply the performance measures mindlessly. They occupy the strategic position of *shaping* the new technology rather than opposing it head on. Because the history of technology shows that rarely ever does it have merit to fight a new technology by referring to a supposedly ideal past. Here it should not go without stressing that among the pragmatists, the social scientists, notably the sociologists of science, take center stage by contributing reflective analysis of science. After all, indicator construction for science depends foremost on reflective, theoretical knowledge of 'how science operates'.

As Burrows comments, "academic accounts of the life world of the post war concur that measure in any systematic form, with accompanying material consequences is new. Measure as we would recognize it, simply did not exist in the post war university" (Burrows 2012: 3). Referring to the fictional representation of the university in campus novels in the 1980 s he finds testimony of the "transformations that were subsequently unleashed upon the academy: a world in which relations between measure and value have become increasingly enacted via code, software and algorithmic forms of power; a world in which the role of number and numbers has come to take political precedence over the aesthetic, the affective and the hermeneutic; and a world in which structures of feeling have been, consequently, fundamentally altered" (Burrows 2012: 4). It could have hardly been said in clearer words, except to add that this applies not just to the academy but to all other institutions in modern societies. It is inescapable.

In light of this the plea for a 'return to academic virtues' such as 'reading' the works of colleagues when sitting in a search committee rather than relying on citation scores and impact factors is like opting for stagecoach travel in order to get a deeper appreciation of the landscape. As justified as the lamentation over the orientation to the symptom rather than the examination of the original may be, the world in which that was still reality is gone. To long for its renaissance is pure nostalgia, and nostalgia just ain't enough.

Postscript

If the reader wonders why I have written this essay in English for a German language sociology journal in which all other articles are in German, the explanation should be obvious at this point: to beef up my Research Gate score and my h-Index!

References

Burrows, R. (2012): Living with the h-index? Metric assemblages in the contemporary academy, in: The Sociological Review 60 / 2, S. 355-372.

Butler, L. (2008): Using a balanced approach to bibliometrics: quantitative performance measures in the Australian Research Quality Framework, in: Ethics in Science and Environmental Politics 8, S. 1-10.

Butler, L. (2010): Impacts of performance-based research funding systems: A review of the concerns and the evidence, in: OECD (Hrsg.), Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings, OECD Publishing, S. 127-168.

DORA, The San Francisco Declaration on Research Assessment, online abrufbar unter: http://am.ascb.org/dora/, letztes Abrufdatum: 2.5.2015.

Enserink, M. (2009): Are You Ready to Become a Number?, in: Science 323, S. 1626-1664.

International Mathematical Union (IMU)(2008): Citation Statistics, online abrufbar unter: http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf, letztes Abrufdatum 22.3.2015.

Martin, B. / R. Whitley (2010): The UK Research Assessment Exercise: a Case of Regulatory Capture?, in: R. Whitley / J. Gläser/ L. Engwall (Hrsg.), Reconfiguring Knowledge Production: Changing authority Oxford relationships in the sciences and their consequences for intellectual innovation, Oxford, S. 51-80.

Osuna, C. / L. Cruz Castro / L. Sanz Menéndez (2011): Overturning some assumptions about the effects of evaluation systems on publication performance, in: Scientometrics 86 / 3, S. 575-592.

Peters, H.P. et al. (2007): Culture and technological innovation: Impact of institutional trust and appreciation of nature on attitudes towards food biotechnology in the USA and Germany, in: International Journal of Public Opinion Research 19 / 2, S. 191-220.

Power, M. (1997): The Audit Society: Rituals of Verification, Oxford.

Rhoten, D. / W.W. Powell (2007): Property: Expanded Protection versus New Models of Open Science, in: Annu. Rev. Law Soc. Sci. 3, S. 345-373.

Sarewitz, D. (2007): Does Science Policy Matter?, in: Issues in Science and Technology 2, S. 31-38.

Weingart, P. (2013): The loss of trust and how to regain it: performance measures and entrepreneurial universities, in: L. Engwall / P. Scott (Hrsg.), Trust in Universities, London, S. 83-95.

Weingart, P. (2005): Impact of bibliometrics upon the science system: Inadvertent consequences?, in: Scientometrics 62 / 1, S. 117.131.

Wilsdon, J., et al. (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, © HEFCE 2015.

Wilsdon, J. (2015): We need a measured approach to metrics, Nature, 9. Juli, VOL 523, 129

Prof. Dr. Peter Weingart
Universität Bielefeld
Institut für Soziologie
Postfach 10 01 31
33501 Bielefeld
weingart@uni-bielefeld.de