Research Evaluation, 28(1), 2019, 73–83
doi: 10.1093/reseval/rvy032
Advance Access Publication Date: 16 November 2018
Article

OXFORD

Does counting change what is counted? Potential for paradigm change through performance metrics

Ann Rudinow Sætnan[1], Gunhild Tøndel[2,*] and Bente Rasmussen[1]

[1]Department of Sociology and Political Science, Norwegian University of Science and Technology, Trondheim, Norway, and [2]Department of Interdisciplinary Studies of Culture, Norwegian University of Science and Technology, and NTNU Social Research, Trondheim, Norway

*Corresponding author. Email: gunhild.tondel@samfunn.ntnu.no

Abstract

This qualitative case study of emergent practices in a multi-paradigmatic field at one department in Norway explores how performance metrics in science intra-act with staff composition and resource allocation, potentially affecting paradigm trends. Inspired by agential realism, we discuss four key metrics as the material core of an assemblage through which Academe is iteratively enacted. Separately, together, and in concert with other metrics and practices, these play major roles in steering disciplinary development. Our empirical material consists of official documents, bibliometric outputs, and auto-ethnographic observations. Metrics and the practices and institutions they measure are iteratively co-emergent. Responses to the assemblage include hyper-cooperation, cooptation, and resistance. Effects are not uniform and may add to discrimination according to field and position status. Emphasis on international publication may also lead to increasing distance between social sciences and the local society they are entrusted to study. Recently, we and other colleagues have become concerned that performance metrics may also be contributing to paradigm bias in hiring procedures.

Key words: performance management, bibliometric accountancy practices, productivity indicators, sociology, paradigm, qualitative vs quantitative tradition.

1. Introduction: Productivity indicators in academe

Since at least the 1970s (De Rijcke et al. 2016: 162), performance metrics have been spreading, diversifying, and becoming entrenched in Academe (Wouters 2014: 429; Hicks et al. 2015). As standardized units of data travelling within public institutions, they occupy multiple roles—measuring productivity, justifying funding, and distributing stigma and status. The techno-political infrastructures implementing these metrics have expanded the managerial frame for science with a detailed repertoire of means for describing, ranking, and rewarding scientific performance, quality, and impact. Policy propelled top-down—e.g. through initiatives of the Organisation for Economic Co-operation and Development (OECD) and the European Union to internationalize and mainstream the higher education system for a more competitive European knowledge market—intertwines with bottom-up practices, sometimes producing unintended outcomes.

Dating back to studies of Scientific Management in industry, it is long established that metrics are also actions with intended and unintended organizational and personal effects, including gaming and goal displacement (Merton 1936; Butler 2007; De Rijcke et al 2016), task reduction (Laudel and Gläser 2006), proletarization and loss of autonomy (Braverman 1974; Burrows 2012), and affective outcomes such as stress and feelings of inadequacy (Burrows 2012). Common trends across this literature have been to emphasize the responses of employees already in place (Butler 2003; De Rijcke et al. 2016), validity issues with indicators (Moed and van Leeuwen 1995; Aksnes and Rip 2009; De Rijcke and Rushforth 2015), and how the sum of strategies and outcomes may pervert even the best designed metric systems (Merton 1936; Laudel and Gläser 2006; Bevan and Hood 2006; Hood 2006; Butler 2007; De Rijcke et al 2016; Sætnan, Lomell, and Hammer 2011; Tøndel 2011, 2017). Metrics in academic evaluation and assessment systems have a range of constitutive effects that have been studied system by system

(Dahler-Larsen 2014). Less is known about the interplay among metrics, although there is a growing recognition that these confront academics not only one metric system at a time but also as an assemblage of interlocking metrics (Burrows 2012). Effects arising through intra-actions within the assemblage as a whole have only scarcely been documented and analyzed (De Rijcke et al. 2016). This article contributes to filling that gap.

Our approach is a case study of emergent practices of evaluation, resource allocation, and hiring processes at one Sociology department in Norway. This study offers a forewarning of role(s) the assemblage may play in (re-)configuring the content of local disciplinary practice. We ask: *How do (especially biblio-)metrics intra-act with staff composition and resource allocation, and might these affect paradigm trends?*

The interlacing of metrics is neither just a question of more or less individualized employee understandings or strategies nor reducible to single mechanisms such as goal displacements or task reductions, although those too are part of the picture. However each metric's outputs are produced, whatever their visions and blind spots, whatever the involvement(s) of gaming, internalization, compliance, task reduction, goal displacement, and so on—their outputs become inputs, premises, motivations, filters, and focal lenses for other metrics in the assemblage. Thus, while not ignoring documented effects such as gaming, goal displacement, etc., we use Barad's conceptual framework of measuring apparatuses as entangled intra-actively with the phenomena they at once constitute and describe. In what follows, we present Barad's agential realism and discuss how we apply it to academic performance metrics. We then describe the emergence of this assemblage of metrics in a knowledge policy context internationally and nationally; the trickle-down movement of bibliometrics through the university; the intertwined metrics of national evaluations, research proposals, and hiring processes; and how these intra-act with paradigms' publishing styles.

2. Applying agential realism to academic performance metrics

As a governance tool, bibliometrics can train us to see knowledge in specific ways. Mertonian sociology of science acknowledged this insight early, focusing on topics such as the stratification of science and the development of specialties (Wyatt et al. 2017). Taking measurement of 'spacetimematter' as the nodal point of analysis and focusing on constant change through the co-emergence of practices and phenomena, Barad (2007) offers a perspective better adapted for exploring metrics-related changes within disciplines.

Barad (2007), a physicist, explains why one cannot answer questions such as 'is light formed of particles or waves' experimentally. Apparatuses designed to measure or test for one hypothesized property cannot at the same time measure or test others. Measured as particles, light behaves and is registered as particles; measured as waves, it behaves and is registered as waves. Furthermore, each apparatus has impacts on its object. Measuring momentum slows it; measuring position moves it. Some would claim this supports epistemological, perhaps even ontological relativism, but Barad argues for *agential realism*.

According to agential realism, no entities are defined prior to and outside the apparatus. Rather, the apparatus *intra*-acts with those entities to constitute phenomena of which the apparatus is itself an inseparable part. Intra-action stands in contrast to 'interaction' (which presumes the prior existence of independent entities) and involves a specific material configuration of 'the apparatus of observation' (Barad 2003: 815). This does not render the resulting phenomena fictional. Phenomena are real if reproducible within the precisely described context of the experiment in question and can be generalized as true/real for any equivalent apparatus.

Furthermore, not only are apparatus and object mutually entangled and co-emergent. Researchers too are entangled with their tools, objects, and practices (see Barad 2007: 161–8). We are inescapably part of our research apparatuses and study phenomena. No procedures we might follow can extract us from the part(s) we play in their (and our) co-emergence.

Barad (2007) holds this true not only for nano-phenomena studied in quantum physics but also for macro-phenomena observable through the human senses, phenomena where we ourselves *are* the research apparatus. It holds true for social sciences and humanities as for the natural sciences, in life as in laboratory experiments, for all cognition and all reality. The cosmos is a relational space enacted through intra-actions: 'Space, time, and matter are mutually constituted through the dynamics of iterative intra-activity' (Barad 2007: 198).

Measurement does not stand outside this intra-activity but is a particularly powerful form of intra-action not reducible to the intentions of the measurer. Barad (2007: 19) argues that each measuring apparatus renders certain properties determinate while excluding others. This is not governed by the desires or will of the experimenter, but by the specificity of the apparatus. We take this perspective, viewing (especially biblio-)metrics as the material core of an intra-active repertoire of practices through which the relevant parts of Academe iteratively emerge.

3. Methodological considerations

Barad says little about methodology, except that one should describe in detail the workings of the apparatus. In a similar vein, Latour (2005: 144) admonishes to 'just describe the state of affairs at hand', including only those actants and actions that leave a trace in the outcomes. Inspired by this advice and the approach known as organizational auto-ethnography (AE), we have grounded the analysis in our own experiences with metrics through our positions at a Norwegian university department, complementing these data with organizational documents and bibliometric records from the same department. Data sources are further described below.

We ourselves are obviously both participants in and objects of these metric systems. We have experienced them as producers and reporters of data, as ranked academics and members of ranked departments, as recipients of incentive rewards (or not) and the kudos and/or scorn of our colleagues, as submitters and evaluators of research proposals, and as job applicants and search committee members. We use our personal knowledge of these systems to help explain how they function, beyond what the formal documents and produced statistics themselves can tell.

Using own experiences as qualitative data is seen by some as essentially lazy, biased by familiarity, unethical to publish, experiential rather than analytic, and/or overly focused on the self (Delamont 2007; Doloriert and Sambrook 2012). We acknowledge these warnings but heed also strategic considerations, practically and epistemologically. Staying local meant we already knew much about the production of the data we found. We are *in* the data, a familiarity offering advantages—guiding us to relevant documents and helping us interpret them. Yet we are not focused on ourselves, but rather on the organizational context, aiming to contribute an understanding of the

phenomenon to the greater culture (a prerequisite for AE (Doloriert and Sambrook 2012)). We treated our personal experiences analytically reflexively, discussed findings with colleagues, and were committed to the chosen theoretical framework (Anderson 2006). Used thus, organizational AE opens up for exploring mundane and silent dimensions of the organizational context that may otherwise be difficult to find (Doloriert and Sambrook 2012).

An agential realist study of a knowledge-producing apparatus means there is a double apparatus to describe: the object of our study, and the means we have used to study it. In this instance we ourselves are knowledge-producers within both apparatuses. The apparatus we study provides us with the means to study it even as we participate in it studying us—all densely intertwined. In our methods section we therefore take double steps—describing the principles underlying our choices, and placing ourselves into the context entailed by following those principles. We then move on to describing and analyzing the apparatus's intra-actions and outputs.

4. Methods 1: Choosing discipline and department

We are looking for signs of paradigm shifts, i.e. changes in the types of questions asked and the means used to answer them. According to Kuhn (1970), such changes occur when the accumulation of anomalies overwhelms the capacity of existing theories to explain empirical results. According to others (Collins and Pinch 1993), 'scientific revolutions' are not data-driven so much as socially driven. Since data are always interpretatively flexible, scientists maintain a research trajectory in spite of anomalies, explaining the anomalies within their preferred paradigm. However, funding and recruitment may wither for one paradigm while blossoming for another, creating paradigm revolution via a generation shift.

We are looking for potential paradigm shifts constituted through the workings of a performance metrics-based apparatus that is neither internal nor external to science but rather an inseparable part of science as a phenomenon. While not defining this constitutive mechanism as 'social' or 'external', we would nevertheless expect to see paradigm shifts soonest in multi- or pre-paradigmatic fields, especially when undergoing a rapid generational change. Sociology as a field and our department–the Department of Sociology and Political Science (ISS) at the Norwegian University of Science and Technology (NTNU)–as an example fill both criteria.

Sociology, while not uniquely so, is clearly multi- or pre-paradigmatic. Sociology has been referred to as a 'low-consensus discipline' (Leahey and Reikowsky 2008: 425), and less optimistically, as an 'archipelago of poorly connected islands of specialization' (Calhoun 1992: 25). Yet, at some risk of oversimplification, Sociology's many paradigms can be lumped into two groups, and some such lumping is necessary to render a small total number of people studyable.

One traditional divide lumps together paradigms associated with and pejoratively referred to as 'Positivism'. Positivism *stricto sensu* refers to Comte's strategy of using data from elements of social life that seemed (matter-of-)factual and readily quantifiable—elements he called 'positifs'. Once quantified, these can be correlated statistically to identify social patterns. Without necessarily committing to Comte's theories, many sociologists practice simplification, operationalization, quantification, and various measures of correlation as the mainstay of their research.

Another porous and disputed cut through the field is the quantitative: qualitative divide. That categorization is also inaccurate. Researchers sometimes mix methods or switch methods from project to project. Some collect textual data and then analyze it quantitatively, or collect numerical data and analyze it interpretatively.

To avoid the established, inaccurate assumptions implicated in these traditional categories, we will from here on use a newly minted (though still potentially distortive) one focusing on the analysis stage of research, hereafter calling the first paradigm group 'correlationist' and the second 'heuristic'.

Correlationist work need not rely specifically or exclusively on correlation analysis; regression, variance, or factorial analyses are also about patterns of co-occurrence of data categories and values within data sets. Furthermore, correlation in itself may not be the goal. Rather, mathematical relationships may serve as indicators of causality and/or tools for prediction. Nevertheless, within this group, mathematical relationships constitute the core of analysis and primary source for conclusions, with interpretation of texts playing at most supporting roles, e.g. as tests for survey questionnaire items.

The heuristic group also includes a mix of theoretical approaches: phenomenology, interactionism, constructivism… As varied in their theoretical commitments as the correlationists, heuristicians too share an analytical preference—in this case, the interpretation of various forms of texts and a focus on meanings as the key to understanding social action. Again, the group's boundaries are porous. Mathematical analyses may well be included but play at most a supporting role, with interpretation the core of knowledge production.

Differences between these two groups may be quite subtle. Had we been categorizing strangers, the work would have demanded close reading of publications and cross-checking category choices with one another. As we were categorizing fellow professionals we have known as teachers and colleagues for decades, we found the divide fairly easy to follow and, by the same token, difficult to cross-check as we are likely to share our biases. Regarding biases, it behooves us to mention that we categorize ourselves as heuristicians, although we have all also studied, occasionally performed, and even taught quantitative analyses.

Both groups are widely represented where Sociology is performed, with tendencies for one or the other to be hegemonic. Our own department was solidly correlationist until the late 1980s, although heuristically oriented graduate students were actively present since the 1970s. In the late 1980s, during an economic downturn, Norway's government addressed unemployment by expanding university admissions. Universities hired staff to increase teaching capacity, and many heuristicians, having conducted research elsewhere, were positioned to land the new jobs. Since then, the department has comprised both correlationist and heuristic approaches, with some emphasis on the latter. This is not to say that an ideal or 'true' balance has been achieved. Rather, this demonstrates again that factors affecting university funding and recruitment can effect paradigm shifts, even aside from data- or anomaly-driven processes.

Observing recent hires at the department, we wondered whether performance metrics were triggering a pendulum swing toward correlationism. We rejected notions that writing might be easier in one paradigm or the other, thus inflating CVs. Such accusations have been hurled across social sciences' methodological divide from both sides, but each side faces its own quality demands, making such a comparison moot if not impossible. Rather, we explore whether the metrics apparatus might be inscribed with properties that support a

quantitative research framework, defining and rewarding those properties as signs of quality, thereby favoring correlationist hires.

Factors we consider are the age composition of the Sociology staff and the various intertwined roles bibliometrics play in resource distribution and hiring processes. Two-thirds of the tenured Sociology staff (12 of 18 tenured faculty members, from both sides of the correlationist/heuristic 'divide') have either recently retired or will be retiring by 2021; up to 12 positions may be filled over less than a decade, i.e. a generation shift in a multi-paradigmatic field and department. Placing ourselves in this context, one of us is partially retired, and she and another expect to fully retire within 2 years, while the third is an early career researcher seeking tenure.

5. Methods 2: Metrics systems and data sources

Defining bibliometrics broadly (for the purposes of this article) as any metric focusing on publications, we include publication points, publication counts, and some aspects of discipline evaluations and grant proposal metrics within that umbrella category. Bibliometrics are part of a larger metric assemblage academics now encounter in their everyday work. We will situate bibliometrics within a larger assemblage context, though a complete picture of all the metrics-based systems that intra-act strikes us as impossible given the limitations of a journal article. We will discuss four key (biblio-)metrics within the ever-changing assemblage: publication points in the national bibliometric system, bibliometrics in national disciplinary evaluations, metrics in/of research grant proposals, and publication numbers in connection with hires. Though a far greater portion of university incomes is linked to other metrics (numbers of registered students, study program commitments, and completed exam programs), publications and research awards resonate with core academic values, thus towering high in academics' awareness. Separately and together, the many metrics play major roles in determining the direction of research and thus of disciplinary development.

The details of the metric systems will be discussed below. Important to note here is that each of these systems' origins, procedures, algorithms, and outputs are documented in available sources. Those will be our primary data sources. We have checked official sources regarding the formal details of each system. From the records of the publication points system, we have downloaded the titles and authors of the 405 journal articles, books, and book chapters registered to Sociologists at our department from 2000 to 2016. Regarding the hiring system, we have accessed public as well as internal documents (job announcement, candidate list, and evaluation report) covering one case out of five recent hires. We are now ready to describe the metric assemblage.

6. Seven cuts through the metric assemblage

6.1 Cut 1: Historical context

Some may remember the implementation of the Research Assessment Exercise (RAE) system in the UK in 1986—marking a turning point for acceleration in the use of performance-based systems as academic policy tools (Aagaard, Bloch and Schneider 2015). RAE entailed international panels evaluating their respective disciplines at several-year intervals. Evaluations covered numerous criteria and data types, and use of bibliometric indicators varied across disciplines. The RAE process was expensive, and panels were found to favor disciplines' own quality criteria, basic over applied research, and academic over social or economic impacts (Barker 2007). It was recommended to make the RAE less 'high science' oriented, less controlled by disciplinary concerns, more amenable to applied science in collaboration with industry and government, and more directed towards producing socioeconomic impacts. Bibliometrics were proposed as a tool for achieving these aims, as well as for saving money (Ibid.). In other words, bibliometrics were expected to produce changes in science content, not just accidentally but also intentionally.

Later, the OECD's recipe for internationalization and competition (in all fields of endeavor) was to separate the regulatory role of the state from its role as owner by deregulating state monopolies to improve performance, efficiency, and effectiveness of public spending (OECD 2003). Managerial autonomy and accountability for results would, it was thought, improve performance (OECD 2002). In 2001, a right-wing Norwegian cabinet introduced the OECD policy to Academe, setting the universities 'free' to govern themselves as part of a trend turning state-owned agencies into business(-like) enterprises. In 2005, ours was the first Norwegian university to introduce the new management principles (Rasmussen 2015).

Accompanying the new institutional 'autonomy', however, were new forms of control: direct control of education quality by a certifying agency, and indirect control through competition in quasi-markets for institutions' shares of the national higher education budget. Systems were developed to register institutions' production of credits and candidates ('Common Student System', CSS) and publications ('Cristin') and algorithms to calculate budget shares. Whereas CSS registers credits automatically, Cristin is only partially automated, requiring manual checking and completion by authors.

As Lund (2015) describes, further OECD recommendations in 2005 commodified and standardized educational qualifications to make them transferable internationally, facilitating academic mobility. Recommendations emphasized monitoring research output and dissemination and measuring higher education quality through self-assessment and comparison across institutions and nations (Lund 2015: 21). Incoming and outgoing international students and international research partnerships count as indicators of internationalization and university competitiveness, while internationally, publications in top international (i.e. English language) journals have become the main measure of institutional and individual performance (Tienari 2012).

Though publications were registered earlier, Norway implemented governance by bibliometrics from 2005, with budget impacts from 2006 lending even more assessment prominence to bibliometrics. The Norwegian model rewards publication by quantity and channel, i.e. where and how much you publish (Aagaard, Bloch and Schneider 2015). Using the system's own indicators, early evaluators of the Danish model (largely a copy of the Norwegian one), unsurprisingly, found increases in numbers and prestige of publications but did not examine publications' content or style (Sivertsen and Schneider 2012). Wright (2014) also found changes to researchers' health, practices of resistance, institutional culture, and what aspects of scientific production were pursued and published (on such changes, see also Aagaard, Bloch and Schneider 2015; Berg and Seeber 2016).

In the initial years of bibliometric assessment, one might expect changes in reporting strategies (report more thoroughly) and/or submission strategies (submit more and to higher-ranked journals), but

given the time lag from submission to publication, those latter changes would only barely be registerable in publication results after 2 years. Changes in content would likely take longer to emerge through slower processes such as research financing, university restructuring, and recruitment.

Discussing the aims and consequences of bibliometric routines, Sørensen (2010a,b) argues that metrics inject an element of commodification suited to economistic public administration. Metrics simplify and deskill evaluation work—an argument that managers and policymakers often find appealing (Woelert 2013). Sørensen (2010b) also discusses over- and misinterpretation of bibliometrics in connection with a specific promotion application at a Norwegian university. Sørensen treats the described promotion as a one-off case: a warning rather than a diagnosis, alarming but not (yet) a pattern.

6.2 Cut 2: Bibliometric management in Norway

Though bibliometrics play a marginal role in university budgets in Norway, intra-actions with other routines and metrics give them considerable impact. To explain their impact, we start from the top with the national budget for the university sector.

Norwegian universities and colleges are financed over the national budget. About half the national financing is calculated according to long-term obligations: study programs and tenured positions. For 2016, these account for 46.9% of the budget at our university (NN). Nearly as much (42.8% of the 2016 budget at NN) is based on students' course credit points and completed PhD degrees. From 2017, completed undergraduate and master degrees will be included in the algorithm. Monetary rates per credit point and degree are fixed for periods of several years, providing a planning horizon to manage degree programs. Deducting these from the overall university share of the national budget, any remainder is divided among the universities according to their respective shares of the biblio- and other metric scores 2 years previously. That budget portion is small (approximately 10% in 2016) and unpredictable (depending on overall size of the national budget, where higher education and research stand among national priorities, and universities' respective ranks in the metrics). Furthermore, it is difficult to predict what impact organizational measures might have on publication rates, and, as numbers of publications rise, the marginal value of each new publication falls. Thus, it is near impossible for universities to plan strategic measures aimed toward increasing net gain from these metrics. Nevertheless, universities do invest both time and money into attempting to increase their research- and bibliometric results.

Research- and bibliometrics make it possible to monitor developments in the different departments of the university and across universities. By comparing results, it is then possible to 'promote constant improvement' as New Public Management promises (Miller 2005). Although the system is supposed to work only on an institutional level (Sivertsen 2009), the data are recorded at individual level and used to control and reward departments and individual academics.

Metrics-based rewards make counted activities count the most (Whitley 2011; Hammarfelt and Rushforth 2017). Chief among these activities is publishing. Until 2014 and effective through the 2016 budget, bibliometrics linked to university budgets have consisted of three categories: articles, monographs, and anthologies. Articles in peer-reviewed journals scored either 1 or 3 points depending on journal rank.[1] Monographs published through peer review scored either 5 or 8 points, depending on publisher's rank. Editorship of an anthology did not count, but each anthology chapter scored 0.7 or 1 point, again depending on publisher's rank. In the case of multiple authors, points were divided equally among all authors and Norwegian-employed authors' portions credited to the budgets of their respective institutions.

Starting from 2015 with budget effects from 2017, a new algorithm applies to works with multiple authors from multiple institutions. Instead of allocating points by the simple ratio of local authors to total number of authors, the square root of that ratio will be the multiplier. For an article published in a Level 1 journal with 10 authors from University X and one from University Y, the old algorithm looks like this: University X, 1*10/11 points = 0.91 points; University Y, 1*1/11 points = 0.09 points. The new algorithm will look like this: University X, $1*\sqrt{10/11}$ points = 0.95 points; University Y, $1*\sqrt{1/11}$ points = 0.30 points. Thus, publications with multiple authors from multiple institutions will score higher than single-authored and/or single-institution publications. How much higher will depend on the number of authors and institutions and authors' distribution across those institutions. Additionally, points for publications with at least one author at an institution outside Norway will be multiplied by 1.3. This change arose from academics in medicine, engineering, and natural sciences complaining that, because they published with many other authors, they had to publish far more than those in social sciences or humanities to achieve as many points. In response, a committee tweaked and retweaked, ran and reran the algorithm until they found a formula whereby professors came out with the same point averages across disciplines.[2] The new algorithm is publicly presented as 'fair' and 'equitable', although as we shall see, it has differentiating effects among early career academics.

6.3 Cut 3: Discipline evaluations

Academics know the publication points of other disciplines and departments through universities' annual reports and national evaluations. In 2017, social sciences went through a new model for national evaluation–EVALuering av norsk SAMfunnsvitenskapelig forskning (SAMEVAL)–to be repeated pentennially hereafter. In addition to bibliometric results, SAMEVAL also requested the 10 most important publications of each discipline department or section, documentation of impact, and presentation of the 10 most important examples of dissemination.

Furthermore, SAMEVAL evaluated the production of these metrics results, requesting results representing large, permanent research groups. As with the change in the biometric algorithm, attention to and valuation of large research groups reflects the introduction into social sciences of norms from the STEM (Science, TEchnology, and Medicine) disciplines. Because hardly any of the Sociology departments in Norway had established such groups, many strategically 'gamed' this aspect by 'constructing' groups out of successful research areas or by lumping together productive researchers—reporting the 'groups' to SAMEVAL regardless of how little collective activity actually occurred.

6.4 Cut 4: Research proposal metrics

Unsurprisingly, one's publication record also counts when applying for research funding; yet the research funding context introduces some additional metrics to the assemblage as well. One of these consists of anonymous reviewers' numerical scores on a number of assessment points. Across the major funding sites, it has become

standard that questions such as originality, importance, appropriateness of methods, budget realism, strength of the research team, principal investigator's leadership experience, and potential impacts of the project are each scored on a scale from 1 (lowest) to 7 (highest).

Though simply an expression of reviewers' subjective opinions, once in numerical form, scores take on a glow of objectivity. Funding committee members are presented with a score-ranked list of proposals. Although they also have access to the full proposals and the reviewers' descriptive reasoning for their scores, it takes considerable effort to convince ones fellow committee members to change the score-ranked order. Proposals with score averages of 4 or more are, in principle, considered fundable; scores of 5 or more are considered good, but, in practice, even proposals that score all 6 s and 7 s may not achieve funding from the most competitive programs. Nevertheless, however slim the chances of funding might be, proposals in themselves are considered important enough to be tallied. Proposal tallies, scores, and funding results are all outputs from the funding process and inputs elsewhere in the overall metrics assemblage.

6.5 Cut 5: 'Trickle-down' intra-actions

Publication metrics and research funding convey cumulative advantage to one another. Long publication lists serve as indicators of competence, boosting funding proposal success.[3] Funded research allots time to the production and publication of results. Even non-funded research can lead to further funding options when universities use numbers of proposals submitted and referee scores on proposals as criteria for allocating internal research funds. The cumulative intra-actions of these metrics create the well-documented 'Matthew effect'[4] both for the researchers counted and for the cultural weight attached to each metric.

Of these mutually supportive metrics, the publication metric counts most directly in university budgets. The others come into play by increasing the likelihood of publications and through local allocations as university budgets 'trickle down' through institution-specific routines to faculty budgets, trickling in turn to departmental budgets, where there may finally be practices allocating discretionary funds to individual faculty members. Budget routines vary somewhat from institution to institution, faculty to faculty, and department to department, each potentially involving new metrics-based algorithms. At each trickle-down step, algorithms also set aside strategic funding for investments in projects, positions, proposals, publications, and material resources. We will here present each of the trickle-down steps in turn as currently practiced at our university, faculty, and department.

The 'trickle-down' process begins with the national budget's allocations to the universities. Besides the three budget elements discussed before (programs, credits and degrees, and publication points), some funding is allotted for 'centers of excellence' and other earmarked programs. Competition for such program funds includes documentation of 'excellence', where international networks and publication records play key roles.

Once the universities' shares of the national budget are known, each university sets aside sums for central administrative functions and strategic investments, distributing the remainder among centers and faculties. At NTNU, faculty budgets are again based on programs, credits and degrees, and relative proportion of the university's publication points, along with a redistributive algorithm to stabilize faculty budgets (after all, tenured faculty cannot be fired on short notice when budgets fall nor is it always wise to expand immediately when budgets rise). Each faculty repeats this distributary process among its departments. Finally, our department uses some of its budget to offer incentive moneys directly to teaching staff. These are fixed sums: NKR 6000 for supervising an MA candidate through to completion in the previous year, 20,000 for the same with a PhD candidate, and 8,000 per publication point—not large sums, but not insignificant either. Incentive moneys are for academic purposes only—equipment, academic travel, research assistants, and access to data sets. Thus, incentive moneys well used can pave the way to further publications.

Some of the strategic funds serve as research investments. Here the proposal metrics come into play. Our own faculty invests in writing proposals, thus boosting that metric. They subsidize/expand projects that received funding and partially finance some projects that were not (yet) funded despite good review scores.

Once allocated, most of the budget is spent on wages. When it comes to filling the positions, a simpler bibliometric is used. It is in the context of hiring routines that we see the clearest difference metrics make in the department's paradigm balance. To show how those consequences arise, we will now take you through the hiring process and a recent example from our department.

6.6 Cut 6: Hiring process metrics

Hiring involves several steps. The department formulates a job announcement, published subject to faculty approval. After deadline, a committee is appointed to review the applications. As a rule, one committee member is on faculty at the hiring department, while two others are from other institutions. Anyone who has supervised and/or co-authored with any of the applicants is disqualified from the committee. With long applicant lists in a small country, this can entail that the committee lacks top qualifications in the specific (sub-) field of the job. Bibliometrics can help build an appearance of expertise and unbiased evaluation.

The committee writes a report including comments on each applicant and each of the publications applicants submit for consideration (until 2011/2012 up to 10 publications per applicant; currently 5 per applicant for an Associate Professorship, 10 for a Professorship). This is a considerable amount of work. Bibliometrics are nearly always used to simplify the job.

The committee must base their assessment on the criteria mentioned in the job announcement, but the announcement gives leeway for operationalization and judgment calls. The committee should arrive at a ranked short-list of about three qualified applicants. A second committee, including a student representative, then interviews the short-listed applicants, and assesses their job lectures. Based on the two committee reports, the department chair writes a recommendation to the faculty, which checks for biases and acceptability of the reasoning behind the assessment. Bibliometrics, with their aura of objectivity, may help the recommendation pass that hurdle. The faculty then finalizes the hire.

In connection with a strategic management reform, our university introduced standardized job announcements reflecting policies on internationalization and excellence. The standard text for a teaching position states that a relevant PhD or equivalent thereof (more on PhD equivalents below) is required and emphasizes the relevance of (especially international) publications 'in the last 5 years'. Departments may adapt this text, especially to describe the position and any particular qualifications looked for. In a

recruitment process at our department concerning two positions, one in quantitative and one in qualitative methods, the standard text was used for the first time with only a minimal supplementary text specifying that the jobs involved teaching quantitative and/or qualitative methodology.

The PhD 'or equivalent' requirement entails a bibliometric element. According to regulations, a PhD consists of either a monograph approved by a degree assessment committee or a composite thesis of three to five related but non-overlapping articles of publishable quality together with an extended introduction discussing theoretical and methodological issues within the article set. The PhD candidate must be first or sole author of at least half the articles within the set. Articles published in refereed, international journals are, by definition, publishable. Thus, in the hiring context, a committee can count their way to job candidate's PhD equivalent(s) by grouping and counting sets of published articles, where each set preferably includes at least one that goes into some depth on theoretical and/or methodological issues. Again, bibliometrics can provide task reduction and an appearance of expertise and objectivity.

The hiring process for two recent methodology jobs can serve as a case in point. We chose this case, as it was the one that set off alarm bells at the department, resulting in measures to counterbalance the influence of metrics on the hiring process. Nevertheless, several later hires have followed a nearly identical pattern regarding reliance on metrics, thus making this case both seminal and typical at the same time.

In this case, the committee operationalized job criteria as simply as possible. All candidates with a relevant PhD were judged as qualified in the data collection methods (regardless of analytical approach) of their research, mainly their PhD project. Candidates who had used any form of interviews, documents, or observations in any of their research were considered qualified in qualitative methods. Of these, some had written monograph-style dissertations, others composites. Once the PhD requirement was controlled for, the past 5 years' publications were used to short-list qualified applicants. As a result, candidates with article-based dissertations or PhD equivalents based on articles had their PhD work counted twice. Monograph dissertations qualified at the first step but did not add to candidates' international publications, leaving no chance to be short-listed and thereby no chance to get the job. Furthermore, although briefly mentioned in the report, the committee attached little importance to whether the candidates had written or reflected about the methods used in their publications, or if their methodology was well chosen and fruitful in their research. Nor did experience in teaching methods courses count in ranking applicants. The count of international publications was clearly the method of selecting short-listed nominees.

A simple, quantitative approach enables evaluation committees to reach decisions about the possible candidates quickly, with little effort and little risk of being accused of bias. However, there are risks to the department and to other staff members who may have to pick up the slack after unfortunate appointments. The emphasis on international publications favors foreign applicants, some of whom have subsequently proven unable to teach lecture courses, grade exams, or take on their share of administrative duties conducted in Norwegian.

Emphasizing publications in the past 5 years also means that speed of publishing becomes more important, in addition to quality or even amount over time. PhDs who write composite dissertations and those immediately granted a research post doc have the best chances to be nominated. Contract researchers and teaching staff with less time to publish will tend to fall short. Effects on gender composition of staff are discussed in Rasmussen (2015).

So, the new bibliometrics, along with other metrics, can affect the composition of the staff. Furthermore, that composition can affect the working conditions, not only of newly hired staff but also of the staff as a whole. Working conditions can in turn affect productivity, which then feeds back into later rounds of hiring and other resource distribution processes. But can they affect the balance of paradigms? To answer this question, we examine the department's bibliometric results since 2000 (Figure 1).

6.7 Cut 7: Publishing styles and bibliometrics

We see in Figure 1 that publication numbers have risen—unevenly, with occasional falls, but overall they have risen, especially since 2006. A number of factors are involved. First, especially the pre-2006 numbers may be incomplete, since registering was a nuisance job with no clear incentives for department or staff. Second, staff size has varied. For instance, the correlations group was down to two people at its nadir. Those two had to pick up vacancy slack in teaching quantitative methods, among the department's largest courses. Eventually, the department hired three new tenured staff members to teach quantitative methods. One of these was awarded external funding, supplementary strategic funding, and a reduced share of teaching duties to build a research team. Meanwhile, the heuristic group has had variable luck with external funding applications—several large projects, several NTNU strategic PhD-ships, sabbaticals, but also periods of limited funds. In 2015–16, the department was again understaffed due to postponed hires. In collating the table, we saw publication falls for individuals who have had extended periods of sick leave, but also publication bursts for several emeriti around the time of their retirement. In sum, in light of what we know about staffing levels and working conditions at the department, we conclude that publication numbers change with work situations and cannot be readily attributed to skill levels, work ethics, or to bibliometric incentives directly.

Note too that article-based dissertations contribute to boosting the total numbers. Such dissertations are becoming more common. In 2000–07, monographs were the norm (21 out of 26). In 2008–11 practices had changed, with 13 composite dissertations out of 21 in total. The past 6 years have seen a slight majority of composite dissertations, but even those who submit monographs now tend to publish an article or two in addition, having seen what it means for their job chances.

We did not calculate publications per staff member. The numbers are simply not readily divisible down to the individual staff member level, in spite of authors being listed in the records we worked from. Correcting for situations such as sick leave would require going beyond our personal recollections and accessing data in confidential personnel files, obviously not available to us. Furthermore, total staff numbers are not always clear. Graduate student publications are included in the totals, but graduate students' presence on the staff is a fluid concept. Some in our graduate programs work and register their publications elsewhere. Some have officially left the program, yet suddenly reappear with publications finished and a dissertation ready for submission. Thus it is nigh impossible to calculate averages per staff member. What did strike us as we tallied publications was the difference between paradigm groups in numbers of authors per title (Figure 2).

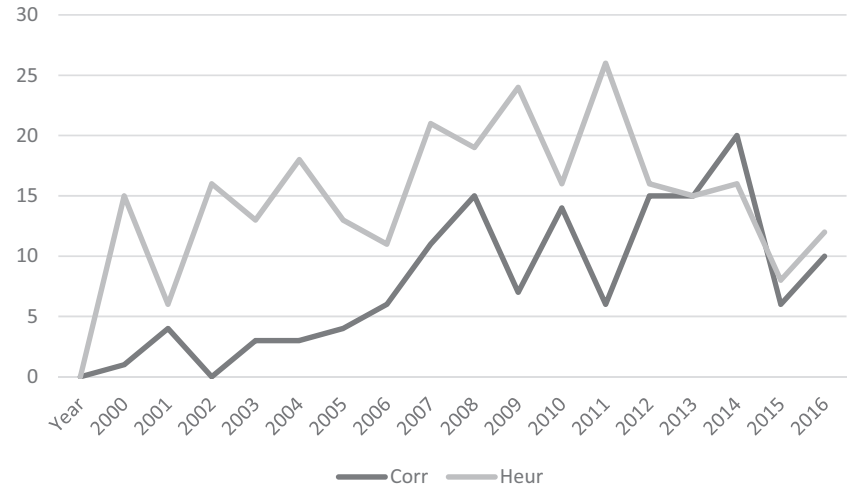

Figure 1. Numbers of journal articles, book chapters, and books in the two paradigm clusters at ISS [Department] over the years 2001–2016.

Source: Cristin. N = 405.

The heuristic group shows a weak trend toward more shared authorship, an international trend across disciplines (Tarkang et al. 2017). Some co-authored titles stem from project work involving two or more of the tenured staff, some from collaborations with researchers elsewhere in Norway or abroad, and some from co-authorship with PhD students. In the latter case—given what we know about the individuals and projects involved—this was not automatic inclusion of dissertation supervisors as co-authors, but rather PhD students engaged in projects together with their supervisors and/or getting more than simple supervisorial help with their first publication or with an international publication. But in spite of the trend toward more collaboration, the plurality of articles, books, and book chapters within this group remains single-authored. The next-largest category of titles is dual-authored. Rarely are there more than three authors per title within this group. Annual averages for the group range between 1 and 2 aside from 3 years that inched above two thanks to outlier articles.

It is understandable that heuristic work has fewer authors. Although textual data can be shared, it was long held to be unethical if not impossible to analyze interviews one had not oneself conducted or observation data one had not oneself produced. Furthermore, data interpretation is highly personal. While working together with another interpreter offers opportunities to discuss interpretations, the more researchers involved, the greater the effort in reaching consensus and the greater the everyday logistics costs. Such problems may be easier for correlationists to overcome. One data collection strategy they sometimes follow is for each member of a network to contribute a pre-agreed matrix of data to a shared data set. In that form of collaboration, the more participants the better, since the larger the data set, the greater its statistical power.[5]

Furthermore, researchers often seek assistance with statistical methods. Our department's correlationists have served in that capacity with authors from other fields (medical researchers, engineers, etc.) and with (heuristic and correlationist) colleagues. Thus, we were not surprised to find higher co-authorship numbers in the correlationist group. Nevertheless, the difference was striking. The correlationist group has never averaged below two authors per title so far this century, aside from 2002 when they registered no publications. Single-authored titles are rare in this group. The plurality of correlationist publications had from three to five authors per title. Peak annual averages are due to a handful of 'outlier' publications with from 10 to 23 authors, most of these stemming from the team analyzing health statistics. Outlier status may make them seem statistically irrelevant, but they are highly relevant in the hiring metric where they significantly raise this paradigm's number of job candidates and candidates' publication counts. It is neither the average numbers for this group that matter nor the median numbers (a central tendency that minimizes the effect of outliers). Rather, it is the outliers themselves that intra-act with hiring process practices to form a mechanism for paradigm change toward more emphasis on correlationist approaches. When hires are based largely on applicants' publication counts, a publication style yielding large numbers of authors (and thus large numbers of potential applicants), many with large numbers of publications (however small their share of the work on each), is favored. And, as we have seen above, the bibliometric reward system now also favors this publication style, further encouraging the hire metric's potential bias toward correlationism.

7. Empirical, theoretical, and practical conclusions

Taking Barad as our reference point, we avoided a simple causal relationship between stable entities (bibliometrics are A and yield X), stearing toward a detailed description of the recursive becoming of Academe as a phenomenon co-emergent with its (biblio-)metric assemblage. Our description of the history, parts, and workings of that assemblage, though necessarily incomplete in article format, has nevertheless shown us some of what the assemblage can co-produce.

We have seen that the metric assemblage was meant (top down) to have effects, not only on Academe's productivity but also on product content, for instance by prioritizing social and economic impacts. We have also seen that responses to the assemblage (bottom up) have been many and varied, with variable effects on the recursive becoming of the assemblage as a measurement apparatus. One response has been hyper-cooperation with the norms implied by the tools in the apparatus—making sure to register all publications, striving as many titles as possible and/or in the highest ranked

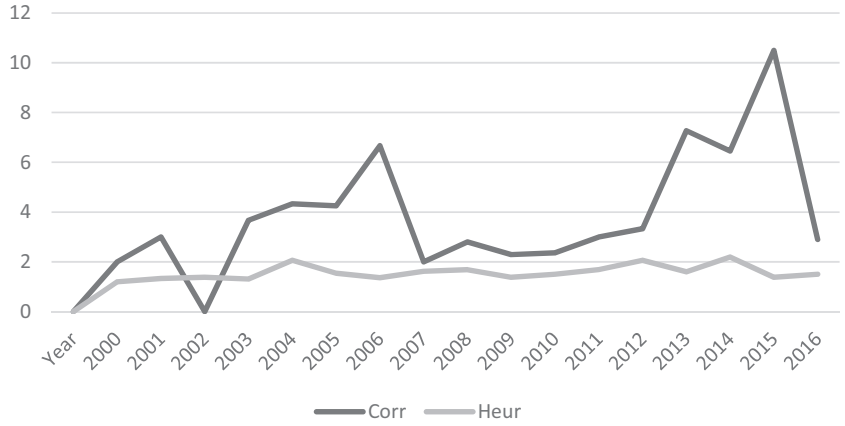

Figure 2. Average numbers of authors per article, book chapter, or book in the two paradigm clusters at ISS [Department], 2000–2016.

Source: Cristin. N = 405.

channels. Another has been to co-opt apparatus tools as means for task reduction. A third has been resistance—designing job announcements to counteract effects otherwise expected from the metrics, or demanding changes to the metrics themselves. Each response feeds recursively back into the assemblage, modifying its outcomes. We have also seen effects of the assemblage modified by intra-actions among the metric apparatuses that comprise it. Different metrics mutually reinforce the status and effects of publication counts and points, each contributing to a cumulative 'Matthew' effect.

We have also seen that the assemblage's and its constitutive metrics' effects are not uniform across Academe. The initial bibliometric algorithm was deemed to favor those who published alone over those who worked and published in large teams. Tenured professors from high-status team-based fields successfully demanded changes to the algorithm. Those changes resulted (at least for a time) in professors averaging the same number of publication points across fields with different publication styles. But, averages are not equal for early career academics subjected to publication counts in the hiring context. In that context, the new algorithm may negatively affect the job chances of those working in low-status fields and/or heuristic paradigms.

Finally, we have seen, at least in theory, a potential for effects beyond the immediate products (publication practices, budgets, rewards, rankings, and hires) the metrics explicitly target. For instance, if it is true that standardization and numeration commodify and disempower that which they numerate, and if the standardization and numeration of academic work does contribute to an increasing dominance of standardization and numeration methods in social science, then it is not only Academe that will be commodified and disempowered but also the society those social sciences study. Furthermore, the emphasis on international publication leads to an increasing distance between social scientists and the local society they are entrusted to study. This can happen through two mechanisms. Study results, even from local researchers, may become less available to local (non-English-speaking) populations. And local job candidates may find themselves out-published and therefore outcompeted by candidates from English-speaking countries.

We do not dispute that some questions are best answered through metrics-based methods. However, metrics are recursively modified through responses, complaints, resistances, and even compliance of the subjects/objects of measurement. There is no 'fair' metric, no perfect point where resistance ceases and all effects are as intended. As exemplified above, increased use of bibliometrics has produced an increase in the complexity and thickness of the 'web of representations' (Barad 2003: 811) that performance management depends on. As long as we 'stick to things and words, we can believe that we are speaking of what we see, that we see what we are speaking of, and that the two are linked' (Deleuze 1988: 65, as cited in Barad 2003: 811), improvements of metrics to better map the scientific world will tend to focus on those linkages. However, from an agential realism perspective, these efforts will fail. Managerial and governmental work targeted toward discovering and correcting distortions from interactions between metrical performance measures and strategic adjustments will likely increase the need for the same attention to the systems in use in the future. More representations of scientific activity, dissected into ever more fine-grained units, creating ever more of these small *things* in Academe, are no guarantee for better understanding of that which they are said to measure. Rather, each additional 'thing' may change the relations in the field, thus adding to the (re-)configuration of the discursive practices under the lens while they are measured.

One last message of caution: we have not concluded that (biblio-)metrics causally leads to hiring bias favoring one paradigm or another. What we claim to have seen is that publication metrics, in a given assemblage of metrics and practices, *can* have such an effect. We present this as a forewarning rather than a diagnostic. We have also seen that the actors involved constantly adapt the practices in the assemblage—sometimes reinforcing and sometimes counterbalancing the hiring bias. Our message is therefore not necessarily to discard (biblio-)metrics or to decry them as inherently biased, but rather to be ever watchful and proactive concerning how they are gathered, analysed, interpreted, and used.

Notes

1. See the official description of criteria for inclusion and levels of publishing channels at https://dbh.nsd.uib.no/publiseringskanaler/OmKriterier.action?request_locale=en.
2. A committee member explained this process at another meeting where one of us was present. It was then commented that the consequences for the assistants of the professors who would have all their contributions counted as a publication when applying for a position, was not calculated (Røeggen 2017).

3. Yet, as researchers know, the impression of competence is weakened if the articles are published in the 'wrong' journals. A publication list that consists of fewer articles published in high-rated journals might then matter more. Perhaps we can we talk of different fabrications or styles of presented competence within publication list practices.
4. Named for the bible verse Matthew 25:29. Merton (1968) identified this effect in science where people who occupy higher status tend to receive more rewards, including unearned ones.
5. The scholarship of collaboration has found that collaboration has a variety of salutary effects on research production, increasing the capacity of individuals and teams to produce formal publications (Lee and Bozeman 2005; Gaughan and Bozeman 2016).

Acknowledgements

The authors extend thanks to colleagues and anonymous reviewers for helpful discussions and suggestions, and to [NN colleague] for reading and commenting a revised version of the manuscript.

References

Aagaard, K., Bloch, C., and Schneider, J. W. (2015) 'Impacts of Performance-Based Research Funding Systems: The Case of the Norwegian Publication Indicator', *Research Evaluation*, 24: 106–17.

Aksnes, D. W., and Rip, A. (2009) 'Researchers' Perceptions of Citations', *Research Policy*, 38/6: 895–905.

Anderson, L. (2006) 'Analytical Authoethnography', *Journal of Contemporary Ethnography*, 35/4: 373–95.

Barad, K. (2003) 'Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter', *Signs: Journal of Women in Culture and Society*, 28/3: 801–31.

Barad, K. (2007) *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham, NC: Duke University Press.

Barker, K. (2007) 'The UK Research Assessment Exercise: The Evolution of a National Research Evaluation System', *Research Evaluation*, 16/1: 3–12.

Berg, M., and Seeber, B. (2016) *Slow Professor: Challenging the Culture of Speed in the Academy*. Toronto: University of Toronto Press.

Bevan, G., and Hood, C. (2006) 'What's Measured Is What Matters: Targets and Gaming in the English Public Health Care System', *Public Administration*, 84/3: 517–38.

Braverman, H. (1974) *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. New York, NY: Monthly Review Press.

Burrows, R. (2012) 'Living with the h-Index? Metric Assemblages in the Contemporary Academy', *The Sociological Review*, 60/2: 355–72.

Butler, L. (2003) 'Explaining Australia's Increased Share of ISI Publications – the Effects of a Funding Formula Based on Publication Counts', *Research Policy*, 32: 143–55.

Butler, L. (2007) 'Assessing University Research: A Plea for a Balanced Approach', *Science and Public Policy*, 34/8: 565–74.

Calhoun, C. (1992) *Sociology, Other Disciplines, and the Project of a General Understanding of Social Life*. Chicago, IL: University of Chicago Press.

Collins, H., and Pinch, T. (1993) *The Golem – What Everyone Should Know about Science*. Cambridge: Cambridge University Press.

Dahler-Larsen, P. (2014) 'Constitutive Effects of Performance Indicator Systems', *Public Management Review*, 16/7: 969–86.

Delamont, S. (2007) 'Arguments against Authoethnography', *Qualitative Researcher*, 4: 2–4.

Deleuze, G. (1988) *Foucault. Trans. Seán Hand*. Minneapolis: University of Minnesota Press.

De Rijcke, S., and Rushforth, A. (2015) 'To Intervene or Not to Intervene; Is That the Question? on the Role of Scientometrics in Research Evaluation', *Journal of the Association for Information Science and Technology*, 66/9: 1954–8.

De Rijcke, S. et al. (2016) 'Evaluation Practices and Effects of Indicator Use – a Literature Review', *Research Evaluation*, 25/2: 161–9.

Doloriert, C., and Sambrook, S. (2012) 'Organisational Authoethnography', *Journal of Organizational Ethnography*, 1/1: 83–95.

Gaughan, M., and Bozeman, B. (2016) 'Using the Prisms of Gender and Rank to Interpret Research Collaboration Power Dynamics', *Social Studies of Science*, 46/4: 536–58.

Hammarfelt, B., and Rushforth, A. (2017) 'Indicators as Judgment Devices. An Empirical Study of Citizen Bibliometrics in Research Evaluation', *Research Evaluation*, 3/1: 169–80.

Hicks, D. et al. (2015) 'The Leiden Manifesto for Research Metrics', *Nature*, 520: 429–31.

Hood, C. (2006) 'Gaming in Targetworld: The Targets Approach to Managing British Public Services', *Public Administration Review*, 66/4: 515–21.

Kuhn, T. (1970) 'Reflections on My Critics', in Lakatos I. and Musgrave A. (eds) *Criticism and the Growth of Knowledge*, pp. 231–78. Cambridge: Cambridge University Press.

Latour, B. (2005) *Reassembling the Social. An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Laudel, G., and Gläser, J. (2006) 'Tensions between Evaluations and Communication Practices', *Journal of Higher Education Policy and Management*, 28/3: 289–95.

Leahey, E., and Reikowsky, R. C. (2008) 'Research Specialization and Collaboration patterns in Sociology', *Social Studies of Science*, 38(3): 425–40.

Lee, S., and Bozeman, B. (2005) 'The Effects of Scientific Collaboration on Productivity', *Social Studies of Science*, 35/5: 673–702.

Lund, R. W.B. (2015) 'Doing the Ideal Academic. Gender, Excellence and Changing Academia', Doctoral dissertation 98/2015, Aalto University.

Merton, R. (1936) 'The Unanticipated Consequences of Purposive Social Action ', *American Sociological Review*, 1/6: 894–904.

Merton, R. K. (1968) 'The Matthew Effect in Science: The Reward and Communication Systems of Science Are Considered', *Science*, 159/3810: 56–63.

Miller, D. (2005) 'Bureaucracy, Virtualism, and Governance', in du Gay P. (ed.) *The Values of Bureaucracy*, pp. 233–55. Oxford: Oxford University Press.

Moed, H. F., and van Leeuwen, T. N. (1995) 'Improving the Accuracy of Institute for Scientific Information's Journal Impact Factors', *Journal of the American Society for Information Science*, 46/6579: 461–7.

OECD (2002) Distributed Public Governance: Agencies, Authorities and other Government Bodies. Paris: OECD Publishing.

OECD (2003) OECD Reviews of Regulatory Reform: Norway 2003: Preparing for the Future Now. OECD Reviews of Regulatory Reform. Paris: OECD Publishing.

Rasmussen, B. (2015) 'From Collegial organization to Strategic Management of Resources: Changes in Recruitment in a Norwegian University', SAGE Open: 1–11. DOI: 10.1177/2158244015603904.

Røeggen, V. (2017) 'Presentation of NPI [Norwegian Publication Indicator]', *Academic Council Meeting in Sociology*, Bergen, 14 November 2017.

Sætnan, A. R., Lomell, H. M., and Hammer, S. (2011) 'Introduction. By the Very Act of Counting: The Mutual Construction of Statistics and Society', in Sætnan, Lomell and Hammer (eds) *The Mutual Construction of Statistics and Society*, pp. 1–17. New York/London: Routledge.

Sivertsen, G. (2009) 'Publiseringsindikatoren. Formålet—Resultatet—Diskusjonen [the Publication Indiactor. Goal – Result – Discussion', in Østerud D. (ed.) *Hvordan Måles Vitenskap? Søkelys på Bibliometriske Metoder [How Is Science Measured? Investigating Bibliometric Methods]*, pp. 11–38. Oslo: Det norske vitenskapsakademi.

Sivertsen, G., and Schneider, J. (2012) Evaluering av den bibliometriske forskningsindikator. (NIFU Report No. 17). <> accessed 21 Jan 2018.

Sørensen, K. H. (2010a) 'Universitetsreformer i Vitenskapsformaterte Samfunn – Fra Ekspansjon Til Effektivitet? [University Reforms in Science-Formatted Societies – from Expansion to Effectivity?]', *Vardøger*, 32: 43–67.

Sørensen, K. H. (2010b) 'Fra Excel Til Exellence? Måling og Veiing av Regnearkuniversitetet. [from Excel to Exellence? Measuring and Weighing the Excel-Sheet University]', *Vardøger*, 32: 115–33.

Tarkang, E. E., Kweku, M., and Zotor, F. B. (2017) 'Publication Practices and Responsible Authorship: A Review Article', *Journal of Public Health in Africa*, 8/1: 723.

Tienari, J. (2012) 'Academia as Financial Markets? Metaphoric Reflections and Possible Responses', *Scandinavian Journal of Management*, 28: 250–6.

Tøndel, G. (2011) 'Constructing Mirrors, Constructing Patients–with High Stakes Statistics', in Sætnan, Lomell, and Hammer (eds) *The Mutual Construction of Statistics and Society*, pp. 245–62. New York/London: Routledge.

Tøndel, G. (2017) 'Omsorgens håndtrykk i scener rundt tall [The handshake of care in scenes about numbers]', in Larsen and Røyrvik (eds) *Trangen til å telle. Objektivering, mÍling og standardisering som samfunnspraksis [The Drive to Count. Objectification, Measurement and Standardization as Social Practice]*, pp. 91–121. Oslo: Scandinavian Academic Press.

Whitley, R. (2011) 'Changing Governance and Authority Relations in the Public Sciences', *Minerva*, 49: 359–85.

Woelert, P. (2013) 'The "Economy of Memory": Publications, Citations, and the Paradox of Effective Research Governance', *Minerva*, 51/3: 341–62.

Wouters, P. (2014) 'The Citation: From Culture to Infrastructure', in Cronin B. and Sugimoto C. (eds) *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, pp. 47–66. Cambridge: The MIT Press.

Wright, S. (2014) 'Knowledge That Counts: Point Systems and the Governance of Danish Universities', in Griffith A. I. and Smith D. E. (eds) *Under New Public Management. Institutional Ethnographies of Changing Front-Line Work*, pp. 294–337. Toronto: University of Toronto Press.

Wyatt, S. et al. (2017) 'Intellectual and Practical Contributions of Scientometrics to STS', in Felt U., Fouché R., Miller C. A. and Smith-Doerr L. (eds) *The Handbook of Science and Technology Studies*, pp. 88–112. Cambridge: The MIT Press.