

Evaluation practices and effects of indicator use—a literature review

Sarah de Rijcke^{1,*}, Paul F. Wouters¹, Alex D. Rushforth¹,
Thomas P. Franssen¹ and Björn Hammarfelt^{1,2}

¹Centre for Science and Technology Studies (CWTS), Faculty of Social Sciences, Leiden University, 2300 AX, The Netherlands, ²Swedish School of Library and Information Science (SSLIS), University of Borås, 501 90, Sweden

*Corresponding author. Email: s.de.rijcke@cwts.leidenuniv.nl

Abstract

This review of the international literature on evaluation systems, evaluation practices, and metrics (mis)uses was written as part of a larger review commissioned by the Higher Education Funding Council for England (HEFCE) to inform their independent assessment of the role of metrics in research evaluation (2014–5). The literature on evaluation systems, practices, and effects of indicator uses is extremely heterogeneous: it comprises hundreds of sources published in different media, spread over disciplines, and with considerable variation in the nature of the evidence. A condensation of the state-of-the-art in relevant research is therefore highly timely. Our review presents the main strands in the literature, with a focus on empirical materials about possible effects of evaluation exercises, ‘gaming’ of indicators, and strategic responses by scientific communities and others to requirements in research assessments. In order to increase visibility and availability, an adapted and updated review is presented here as a stand-alone—after authorization by HEFCE.

Key words: indicators; evaluation; metrics; assessments; strategic responses.

Introduction: main research strands

The use of metrics in academic evaluation and assessment systems has a range of constitutive effects (Dahler-Larsen 2014) that have only scarcely been documented and analyzed in empirical research. This knowledge gap may be due to the intimate nature of the interactions between evaluation (which is mostly confidential) and knowledge production (often in the realm of daily life of researchers) (Wouters 2014, 50). In addition, the research agenda that is exploring these interactions is relatively new (Gläser and Laudel 2001; Gläser et al. 2002; Gläser 2010). As a result, we know some of the implications of quantitative assessments, but the picture is still very incomplete (Gläser et al. 2002; Wouters 2014; de Rijcke et al. in press).

The present contribution is an updated version of a literature review that was written as part of a larger review commissioned by the Higher Education Funding Council for England to inform their independent assessment of the role of metrics in research evaluation (2014–5) (Wouters et al. 2015).¹ Below, we present the most relevant lines of work in the literature, with a focus on empirical materials about possible effects of evaluation exercises, ‘gaming’ of

indicators, and strategic responses by scientific communities and other actors to requirements in research assessments. Providing a complete overview of the literature is not feasible for a couple of reasons. First of all, the literature is very diverse. Studies on evaluation systems, evaluation practices, and effects of indicator uses are published in different media, and the preferred outlets are not necessarily always international journals that are covered well by web-based citation databases. The hundreds of sources are spread out over books, edited volumes, articles, reports, and other forms of gray literature that are sometimes relatively inaccessible. Secondly, the relevant literature is scattered over a large number of social science fields, including sociology of science, innovation studies, library and information science, higher education studies, sociology of evaluation, evaluation studies, economics and business studies, medical sociology, science policy studies, research management and innovation, political science, and governance studies. A third hampering factor in presenting a complete overview is the epistemic nature of the evidence presented in the literature. The studies range from surveys and interviews on researchers’ perceptions of evaluations and formal policy analysis of principal–agent relationships to

cultural critiques of the evaluation society and ethnographic studies of evaluation in action (an emerging body of work). The resulting heterogeneity of the evidence poses particular challenges in integrating the literature in a single review. We have therefore opted for the format of a qualitative literature review that aims at mapping the territory, identifies common elements and themes that lie in or across individual studies, and discusses the findings and conceptual arguments on the same subject. We do not aim for comprehensiveness and critical appraisal; the added value lies mainly in bringing together these sets of literature in a systematic manner.

We begin our review by briefly embedding the studies that focus on effects of evaluation practices and metrics uses in a broader context: the set of literatures that focuses on the governance of science at large. The main aim of this introduction is to delineate the main foci of the most important research strands.

A large body of work from higher education studies, new public management studies, organization studies, anthropology of science, philosophy of science, economics, and political science analyzes effects on academic institutions of shrinking governmental research funding and the emergence of new public management from the 1980s onward. These literatures characterize the rise of performance measurement in academic settings as part of a broader surge of accountability measures that have swept across public institutions over the past three decades (Thomas and Davies 2005; Feller 2009; Keevers et al. 2012; Nedeva, Boden, and Nugroho 2012)—which are in keeping with broader statements about the emergence of an ‘audit society’ (e.g. Power 1997; Strathern 2000; Dahler-Larsen 2012). The most critical voices portray researchers as becoming disciplined (often willingly) by predefined, measurable outcomes that crystallize an ideological attempt to extend market-like mechanisms into the sphere of higher education (Shore and Wright 2000; Shore 2010; Burrows 2012). As such, indicators are positioned as tools to steer academic institutions and researchers toward becoming more market oriented, by actively stimulating competition, instrumentality, and privatization strategies (Deem, Hillyard and Reed 2007; Willmott 2011; Leisyte and Dee 2012).

Science policy and higher education studies (e.g. De Boer, Enders, and Leisyte 2007; Rebora and Turri 2013; Reale and Seeber 2013) typically focus on formal characteristics of the assessment mechanisms adopted in national or regional evaluation systems (Fealing et al. 2011; Reale and Seeber 2013). Most research in this area provides the impetus for furthering science policy, and puts much effort into methods of research evaluation. These studies often respond relatively swiftly toward the introduction of new evaluation programs and methods (cf. Cozzens and Melkers 1997). A typical example is the study by Luukkonen (2012) that examines the peer review processes of the then-recently established European Research Council. Science policy studies tend to focus mainly on formalized national evaluation systems, in which indicators of funding success play a larger role than citation-based indicators. Even in indicator-based systems such as the Australian evaluation system, citation analysis does not figure prominently (Gläser et al. 2010: 169). The current Swedish system is an exception, as it uses citation counts from the Web of Science.² This does not mean that citation-based indicators are not used at all, but they are not visible in the formal systems of accounting (Wouters 2014: 49–50).

Sociology of science literature, on the other hand, focuses on institutional and organizational dynamics of science and innovation, including managerial and control mechanisms (Whitley 1984, 2000, 2011; Whitley and Gläser 2007; Whitley, Gläser and Engwall

2010). Organizational approaches to sociology of science have taught us that quality control is not exclusively a cognitive/epistemological phenomenon, but is also deeply social and organizational (cf. Whitley 2000; Hemlin 2006). This work intimates knowledge production is changing in light of transformations to relational systems in which academic researchers are broadly placed (Whitley, Gläser and Engwall 2010; Musselin 2013; Paradeise and Thoenig 2013). It has also begun to analyze how researchers handle the demand for accountability in their epistemic decision-making. However, most of these studies do not deal directly with effects of indicator uses. The studies that do engage with indicator effects will be discussed in the next section.

Implications on knowledge production

This section starts with an overview of studies that consider a number of implications of evaluative metrics on knowledge production (including strategic behavior and goal displacement, task reduction, and potential biases toward interdisciplinary research). This is followed by a discussion of empirical research on institutional responses to metrics-based assessments. Lastly, we describe emerging research on the relation between indicators and the daily practice of doing research.

Strategic behavior and goal displacement

In tandem with the development and first applications of performance indicators in the 1970s, discussions surfaced about strategic behavior and ‘gaming the system’ (cf. MacRoberts and MacRoberts 1989). Researchers are not passive recipients of research evaluation but play an active role in assessment contexts (cf. Aksnes and Rip 2009; Van Noorden 2010). Assessment systems that affect money or reputation, whether the systems are peer review or indicator based, will tend to influence researchers’ behavior in two ways (Butler 2007). The first is goal displacement: scoring high on the assessment criteria becomes the goal rather than a means of evaluating if certain objectives (or performance levels) have been met (Colwell et al. 2012: 27). The second is a more fundamental transformation of the scientific or scholarly process itself regarding the assessment criteria (for instance, by avoiding risk in selecting research topics), a transformation that may be harder to recognize (Butler 2007: 572).

Studies that focused on effects of funding and evaluation systems on scientific output have indeed demonstrated goal displacement. Butler (2003a,b, 2005), for instance, analyzed the introduction of performance metrics in Australian research funding allocation. Her study revealed a sharp rise in ISI-ranked publications in all university fields (but not in other branches of research where this type of funding allocation is not present) when funding becomes linked with publications (Butler 2005). As there was no differentiation between publications (besides being peer reviewed or not) in the Australian evaluation system, the amount of publications especially rose in the third quartile of journals that are more easily accessible. Butler earlier demonstrated how this strategy, while leading to a rise of relative share of Australian publications, has also contributed to a decline of scientific impact (measured in citations) during the same period (Butler 2003). Colwell et al. (2012) state that researchers’ quality considerations may be displaced by incentives to produce higher quantities of publications when funding is explicitly linked to research output (in terms of the number of publications) (Colwell et al. 2012: 26). Similar effects of the use of bibliometrics on the

amount of publications have been found in Spain (Jiménez-Contreras, Anegón and López-Cózar 2003; however, see Osuna, Cruz-Casto and Sanz-Menéndez, 2011), Denmark (Ingwersen and Larsen 2014), Flanders, and Norway (Schneider 2009; Ossenblok, Engels and Sivertsen 2012; Aagaard, Bloch and Schneider 2015).

Strategic response by the research community was also demonstrated in a longitudinal bibliometric study of UK science covering almost 20 years (Moed 2008). UK publication patterns between 1985 and 2003 suggest that specific publication patterns emerged in years before three Research Assessment Exercise (RAEs) that took place in that period (1992, 1996, 2001), depending on whether the RAE was aimed at quantity or quality of publications. In the UK, findings of 'playing the RAE game' (Harley 2002) in this way are numerous (e.g. Hare 2003; Keenoy 2005; Alldred and Miller 2007; Sousa and Brennan 2014). Another study of the UK RAE impact shows that the cumulative research productivity of individuals increased over time, but the effects differed across departments and individuals. Where researchers in higher-ranked programs increased their output in higher-quality journals, researchers in lower-ranked departments aimed at increasing their publications in other outlets (Moore et al. 2002, abstract). A survey among journal editors conducted at the end of the 1990s (Georghiou et al. 2000) also showed that the RAE influenced where authors published. Indeed, research shows that the status of a journal is crucial for researchers' submission decisions (Chew et al. 2007; Harley et al. 2010). However, claims that the RAE would lead toward salami publishing of 'least publishable units' (cf. Huth 1986) were not verified in a later study (Georghiou et al. 2000).

Osuna, Cruz-Casto, and Sanz-Menéndez (2011) are particularly critical of attempts to argue for a causal relation between the introduction of evaluation systems and the rise in the number of publications on a national level (see, for instance, Butler 2003b; Jiménez-Contreras, Anegón and López-Cózar 2003). Analyzing the Spanish case, they argue that there is a range of other explanations, such as the maturation of Spanish science and the rise of the R&D budget and number of researchers that could explain the rise in the number of publications. Moreover, their attempt to isolate the effect of the introduction of a new evaluation system in Spanish science in 1989 does not find any clear effects.

This suggests that the introduction of evaluation systems might be part of a broader development in national science contexts, and that the causal effect of evaluation systems itself is less clear-cut than is usually assumed. Osuna, Cruz-Casto, and Sanz-Menéndez (2011: 588) conclude by arguing that the introduction of the evaluation system might have had an effect in standardizing publication patterns and making explicit the norms that researchers should follow. As such, understanding the effects of the introduction of evaluation systems would benefit from both more advanced quantitative analysis as well as in-depth qualitative research into how these systems 'trickle down' into institutions and working practices.

Biases against interdisciplinarity

A related concern is the potential influence of disciplinary assessments like the REF on interdisciplinary research. An early survey of the impact of the 1996 UK RAE reported evidence of negative effects for interdisciplinary work. Almost half of those in management positions felt the RAE 'had hindered' interdisciplinarity (McNay 1998: 20). A worldwide survey among demographers showed contradictory results. It revealed that this community displayed no

tendency to focus on monodisciplinary research in terms of reading or publishing activity (Van Dalen and Henkens 2012). In economics and many departments in business studies, however, publication productivity has been strongly stimulated by the ubiquitous use of journal rankings as obligatory publication outlets for faculty (Mingers and Willmott 2013). Until recently, these lists were not based on citations but on a qualitative consensus in mainstream economics and business studies about the top journals (more recent lists use a version of the Journal Impact Factor (JIF)). In a first comparative analysis of the effect of these rankings in business and innovation studies, they were found to be biased against interdisciplinary work (Rafols et al. 2012). This study concludes that citation indicators may be more suitable than peer review for interdisciplinary work because criteria of excellence are essentially based on disciplinary standards.

Pontille and Torny (2010) analyzed the production of three journal lists in the humanities and social sciences by the Australian Research Council (ARC), the European Science Foundation, and the French Agency for Evaluation Research and Higher Education. They found that the production of journal ratings is a highly cognitive and political task, and not merely a matter of univocal inventory-making. The modalities that were selected produced very different effects in terms of which scientific communities were involved, how boundaries were drawn around disciplines, and the ways in which revision processes adopted criticism from the fields involved. According to the authors, these and other built-in tensions will not only keep feeding critique but also the need for permanent revisions of the ratings (see also Jensen 2011).

As soon as research managers started to set publication targets based on the two top categories on the list, the ARC decided to drop the established journal rankings in its assessment system (Colwell et al. 2012: 56).

Task reduction

An analysis of researchers' responses to funding criteria in Australia has analyzed the extent to which researchers are forced to focus their tasks and types of publication (Laudel and Gläser 2006). A tension was observed between the formalized journal ranking procedures and researchers' own valuations of their work. In 10 out of the 21 disciplines under study, the four publication types used in the Australian evaluation were not the same as the four types of output that researchers themselves found most important. This mismatch may result in the abandonment of particular types of work (Laudel and Gläser 2006: 294). Laudel and Gläser go on to conclude that arts and humanities are likely to suffer greatly in situations whereby evaluation systems have been modeled primarily around journal-oriented disciplines.

This conclusion is supported by Whitley's (2007) examination of the possible effects of evaluation in specific disciplines. He suggests that the consequences of indicator uses are dependent on both the design of the evaluation system and the social and intellectual structure of the field in which assessment is applied. Whitley makes a distinction between weak (results rarely published or directly tied to sources) and strong evaluation systems (results advertised and directly used for resource allocation). Not surprisingly, he finds that strong evaluation systems are most likely to influence research agendas and practices within any given field. This is especially the case if relatively few competing systems of evaluation and resource allocation exist. He continues to argue that research fields in the social

sciences and humanities are especially likely to be directly affected by strong evaluation systems, owing to little agreement on research goals, less collaboration, and relatively large freedom for individual researchers to develop their own research agendas. A recent study of departmental responses to the national indicator used in Norway supports these claims. The analysis shows how the Norwegian national indicator is used more frequently at individual levels in the humanities, and the indicator has more importance for judging scholarly status in these fields (Aagaard 2015). One possible consequence of strong evaluation systems, according to Whitley, is that disciplines of low 'centrality and prestige' will imitate research practices and norms of dominant fields, for example, by changing publication behavior. Hammarfelt's and de Rijcke's (2015) study of publication patterns at the faculty of Arts at Uppsala university (Sweden) seems to back this assertion, as an increase in the most common type of output in the natural sciences, life sciences, and medicine (i.e. English language journal articles) was observed among scholars in the humanities.

One could feasibly expect that increased drives toward journal publications will devalue other academic activities. A worldwide survey among demographers in developed and developing countries found that 'traditional' academic tasks—such as writing referee reports or translating research outcomes for policy—are negatively affected by a move toward rewarding individual productivity (Van Dalen and Henkens 2012). A study of the RAE exercise in 1996 found, for instance, that publication in professional journals was 'actively discouraged' by some research managers (McNay, 1998: 22). Hoecht (2006) warns that the increased audit-based quality control has replaced a trust-based form of control, and as such, autonomy has decreased in academic settings in the UK in recent decades. This form of one-way accountability that lacks a fostering of trust might negatively affect innovative teaching and research practices. Similarly, Willmott (2011) argues that audit-based control pushes researchers toward mainstream topics that have the highest chance of being published in the highest-ranked journals.

An exploratory study of the UK university education departments shows that the use of performance indicators in evaluation is perceived as an increasing recognition of specific forms of academic involvement—academic research and publications—that are often opposed to more public roles, including applied research, writing professional publications, or teaching (Wilson and Holligan 2013). The authors argue that the ways in which this development is perceived depends on the position of the researchers in question and whether or not it fits with their scholarly practices.

In a recent study, Laudel and Gläser (2014) find that the chances of 'exceptional' projects (in terms of planned innovations and answers to 'big questions') being funded appear to reduce across all disciplines in an increasingly standardized grant-funding landscape, even in those funding programs that enable this type of research (Laudel and Gläser 2014: 1208).

Effects on institutions

In addition to possible goal displacement, bias against interdisciplinarity and reduction of task complexity, metrics-informed assessments have also been thought to affect institutional arrangements as well as the relationship between higher education and government (Martin and Whitley 2010). According to Colwell et al. (2012), the RAE has led universities to develop 'transfer market' hiring strategies, some focusing on recruiting younger staff with research

potential, with others hiring only 'well-established researchers' in the run-up to the RAE (Colwell et al. 2012: 27–8). This second strategy was also observed in Australian university's responses toward formula-based funding (Gläser et al. 2002: 17). This type of strategic behavior may therefore have long-lasting effects on the position of universities and research institutes and thereby on the research agenda.

Alignment between demands of various formal evaluation agencies along more-or-less the same standardized criteria is likely to focus the attention of organizations and subgroups toward satisfying them (Pfeffer and Salancik 2003). Seeing that epistemic properties of knowledge and research funding and organization are strongly connected (Laudel and Gläser 2014), there is reason to assume that the development of formal performance criteria has grown stepwise, with local lock-in mechanisms dependent on specific systems of governance, rather than based on a consistent evaluation logic (cf. Van der Meulen 2007). Although the rise of formal evaluation may appear as an isomorphic phenomenon, the extent to which it has transformed scientific institutions is far from uniform (Whitley and Gläser 2007). So far, little is known of the capacities that governance mechanisms like evaluation programs have for actually controlling and steering loosely coupled work systems of academic knowledge (Gläser 2013). The strategic response of organizations to increasing demands of accountability takes many forms, and one of these is the development of evaluation systems on the university level. However, so far, there have been few studies on the effects of such initiatives. In a study of Norwegian universities, Aagaard (2015) finds several reasons for the local use of indicators: they are convenient and readily available, they reduce the anxiety associated with quality judgments, and are used as an instrument in battles for resources and prestige within departments. Hammarfelt and colleagues (n.d.), on the other hand, depict tensions between developing local systems and following national formulas among Swedish universities. The indicators used in Swedish academic settings are diverse; all universities have their own model; yet, the basic incentives given in these evaluation models are almost identical. Thus, local allocation models are highly influenced by the national system, but the universities under analysis also want to present a unique and independent indicator. Noteworthy in both the Norwegian and the Swedish case is that national systems, despite explicit advice against adoption on lower levels, are used at the university, departmental, and individual level. In fact, a handful of Swedish universities use bibliometric measures to reward individual researchers, and a couple of institutions allocate resources directly to authors based on recent publications. An article which is both listed as 'level 2' in the Norwegian list of journals and is indexed in the Web of Science could at one particular university generate as much as 70,000 SEK (approximately 7,000 EURO) to be used for research time and expenses (Hammarfelt et al. n.d.). Thus, it appears that in these Scandinavian cases bibliometric indicators do trickle down to the point that they actually permeate all organizational levels.

Knowledge about the performativity of numbers suggests that the availability of metrics for generating and ordering information hierarchically about performance creates a demand for such information (Porter 1995; Wouters 2014). Such information-generating functions could carry authority even if some of its first-order epistemic limits are known (Dahler-Larsen 2012). Organizations cannot resist the temptation to collect such information because it is considered as strategically useful in managing researchers and improving on measures of organizational performance. The fact that

competitors collect similar information for strategic purposes makes the prospect of opting-out or ignoring such information perilous (see also, effects of rankings, e.g. Espeland and Sauder 2007). The legitimacy of such indicators does not rest exclusively on their first-order accuracy, but also on the fact that they are assumed to carry authority within the institutional environments with which organizations strategically engage.

Conversely, emerging empirical work on rankings shows how in complex academic settings, performance metrics are not tightly coupled with actions across all sections of the organization (cf. de Rijcke et al. 2015). This goes against recent tendencies to assume that patterns of behavior toward indicators at one level of a public organization imply uniformity across the board (e.g. Sauder and Espeland 2009).

More fundamental changes in the character of academic settings as independent and critical institutions (Shore 2008, 2010; Craig et al. 2014) as well as increased levels of stress and anxiety are also reported in personal accounts and interview-based analyses (Chandler, Barry and Clark 2002; Gill 2009; Sá, Kretz and Sigurdson 2013). Burrows (2012) argues that performance-based control mechanisms have become autonomous entities that are increasingly used outside the original context of evaluations, and get a much more active role in shaping the everyday work of researchers. According to him, the rise of numbers into the fabric of university bureaucracies may even create feelings of powerlessness among researchers. Neoliberal universities provide fertile ground for a 'co-construction of statistical metrics and social practices within the academy' (Burrows 2012: 361). Among other things, Burrows contends that this leads to a reification of individual performance measures, such as the H-index (Burrows 2012: 361). Though hard to assess, such statements of discontent expressed by researchers are empirically important and are therefore included in this review. The emerging empirical evidence we discuss in the final section below suggests that the use of metrics in decision-making contexts cannot however be explained simply as explicit responses to top-down commands. To some extent, metrics also seem to transform more fundamentally 'what can be talked about' and how valuations of academic worth are being made (cf. Espeland and Stevens 1998; Lamont 2012).

Indicator use in research practice

The evidence discussed above indicates that performance-informed research assessment does indeed increase the pressure on researchers and institutions to meet the performance criteria, irrespective of whether the latter are based on peer review or on citations. That the research community and other stakeholders respond strategically toward interventions deliberately designed to align them with current science and innovation policy priorities (Whitley and Gläser 2007) may in turn have unintended effects. Examples include the mechanisms of goal displacement and the more structural changes to research priorities, publication activities, and research capacity-building and organization. However, the evidence to support these claims is at best partial. Whilst most existing studies focus on systems where funding decisions are based directly on performance, whether this link is necessary for unintended effects to occur is debatable. The effects are not primarily based on the amount of funding that is shifted due to performance differences, but mainly on the consequences for researchers' reputation (Hicks 2012). Systems

where performance is publicly reported but not directly linked to funding may therefore lead to comparable or identical effects. Lastly, the most visible types of strategic behavior may obscure more fundamental shifts in knowledge production (Wouters 2014).

Butler (2003a; 2005) notes conservatism of metric users as a long-standing problem, in displaying a preference for user-friendly measurements that trump other inclinations for adopting more state-of-the-art scientometrics. Many scientometric contributions take a normative stance regarding 'unanticipated effects' of quantitative performance measures on the science system and debate whether the field itself might play a more active role in promoting 'good practices' (Weingart 2005; Van Dalen and Henkens 2012). This concern can be seen in the journal *Scientometrics*, which recently published an issue dedicated solely to the JIF's uses and misuses (Braun 2012). Many studies have highlighted epistemic limitations in the JIF, particularly in the evaluation of individuals (Moed, van Leeuwen and Reedijk 1996; Buela-Casal and Zych 2012), whilst others cite 'gaming' by journal editors and publishers as a perverse effect of its rise to prominence in academic evaluation contexts (Archambault and Larivière 2009: 635).

However, emerging empirical evidence paints a more complex picture of how certain metrics, including the JIF, become reified in research management and decision-making contexts, as both formal and informal standards against which to assess the value and usefulness of research activities (e.g. Aksnes and Rip 2009; Buela-Casal and Zych 2012; Derrick and Gillespie 2013; Rushforth and de Rijcke 2015.; Sá, Kretz and Sigurdson 2013; Stephan 2012).

At least two potential explanations are provided in these literatures for the reification of evaluative indicators (both 'amateur' and 'advanced'). First of all, the metrics inform deep-seated, firmly established mechanisms to build reputation and to hire, select, and promote staff (including publishing in high JIF, peer-reviewed journals). Secondly, the responsibility for certain applications of bibliometric indicators is spread over many key stakeholders in the current 'citation infrastructure' (Wouters 2014)—including scientometricians, publishers, librarians (cf. Åström and Hansson 2013; Demšar and Južnić 2013; Petersohn 2014), policy makers, evaluators, research managers, consultancies, researchers, and other metrics users (de Rijcke and Rushforth 2015). Changing the current dominant 'order of worth' in research assessment (cf. Boltanski and Thevenot 2006; Linkova 2014; Stöckelová 2014) is therefore also a distributed responsibility.

In addition, recently, several constructivist studies have emerged, which challenge a number of received normative assumptions about 'proper uses' of indicators. Dahler-Larsen (2014) argues that commonly heard terms like indicator 'misuse' or 'unintended effects' perpetuate a fallacy that the 'right' use of an indicator can be *known* by some group of experts (and that deviation from such preconceived norms therefore constitutes perversion). He argues, in fact, that there can be no such external referent, as indicators can only acquire meaning through contexts of use. His encouragement to treat indicator uses as an empirical topic was taken-up by a recent ethnographic study that explored the constitutive effects of the JIF in Dutch biomedical research. Rushforth and de Rijcke (2015) reported that notions like 'misuses' or 'unintended consequences' did not match-up to the complexities of their empirical findings from the 'shop floor'. The authors found that prestige—an outcome of knowledge work which gets recycled as an 'exchange good' by academic researchers (Stephan 2012)—was tightly coupled with citation counts and indicators like the JIF (in targeting specific

publication outlets, referencing ‘hot’ papers, negotiation of authorship priority, etc.). Quantitative indicators were also observed to feed-into quite routine knowledge-producing activities (e.g. discussions over whom to collaborate with and when, how much—additional—time to spend in the laboratory producing data). The JIF in particular functioned on occasions as screening devices for selecting useful information from the overwhelming amounts of literature biomedical researchers could potentially read. These examples suggest that metric criteria feed into deliberations over where new scientific knowledge is likely to emerge or can be found. As such, the JIF can both denote a certain standing in a field and a particular type of scientific work (e.g. descriptive versus causal) (Rushforth and de Rijcke 2015). Generally speaking, the promise of bibliometric tools for ‘reducing complexity’ is a feature managers and policy-makers find most appealing (Woelert 2013; Cronin and Sugimoto 2014; Collini 2015). In sociological terms, journal ranking tools like JIF-lists help make ‘commensurable’ (Espeland and Stevens 1998) the levels of prestige acquired from publishing in one journal rather than another. This ‘shortcut’ is particularly attractive where there is a lack of substantive expertise in a field or of informal knowledge of the reputational standings of journals among peers. Given the extent of the JIF’s embeddedness in knowledge making activities, Rushforth and de Rijcke conclude that ‘misuse’ and ‘unintended effects’ are inadequate as analytical concepts for understanding deep-seated effects of indicators on research practices.

Another set of works has begun to explore researchers’ own perceptions and attitudes around metrics. This has tended to reveal a discrepancy between the importance of indicators in evaluation practices according to researchers and their own judgment of the accuracy of certain measures (Buela-Casal and Zych 2012). In addition, some researchers seem willing to wait longer on editorial decisions when the JIF is higher, which shows the clogging effects the JIF can have on scholarly communication (Rousseau and Rousseau 2012). Studies show researchers often have ambivalent attitudes about performance and citation indicators (Hargens and Schuman 1990 in Wouters 2014). Citations acquire value as part of the scientific reward system and are often mobilized by researchers in pursuit of scarce resources (Hessels et al. 2009). At the same time, researchers also criticize citations for inadequately representing actual scientific contribution and often preserve prior beliefs about the value of their work in interpreting citation scores. (Aksnes and Rip 2009: 895). This is explained in the literature by the concept of ‘folk citation theories’, which do not have to be consistent in order to be mobilized by researchers as explanatory devices in their competition for reputation. However, the sophistication and complexity of researchers’ interpretation of citation should not be underestimated. Aksnes and Rip (2009) observe that their respondents know the complexity and ambiguity in citations. ‘In other words, researchers have a sophisticated understanding of the citation process and its outcomes, and can explicate such understanding when there are no immediate stakes to be defended’ (Aksnes and Rip 2009: 904).

Lastly, the use of performance indicators and more advanced forms of bibliometric information may also influence the terms and conditions of the development of research agendas. Derrick and Pavone (2013) have reviewed government policies in three countries (the UK, Australia, and Spain). Each country has committed itself to increasing the societal role of researchers by opening up the setting of research problems to a larger group of stakeholders. The study characterizes this as a move toward more participatory approaches. The authors argue that the use of scientometrics for informing this

process has so far been largely overlooked. The question is too often formulated as the choice between peer review and metrics (Taylor 2011). Derrick and Pavone (2013: 573) conclude that the choice should be on the type of use that scientometric information is put to. ‘The simple information portrayed by correctly calculated and applied bibliometric indicators has the potential to engage a larger group of stakeholders than previous evaluation systems could.’ According to this study, future research policies should take advantage of bibliometrics to foster greater democratization of research to create more socially reflexive evaluation systems.

Despite the sheer volume, breadth, and heterogeneity of the studies discussed above, we still know relatively little about the actual effects of indicator use. Many studies are of tentative and theoretical nature, prophesizing on potential effects rather than documenting actual consequences. Larger empirical studies are often restricted to specific and limited aspects (e.g. publication patterns), and rarely touch on more general changes in academic work and knowledge production practices. Several evaluation mechanisms, such as the use of national indicators for resource allocation, are also of quite recent date. More time is needed to fully assess the impacts of these procedures. It is also quite challenging, if not virtually impossible, to separate the influence of specific evaluation procedures from the effects of other societal factors. Yet, these difficulties should stimulate rather than hamper further interdisciplinary, mixed-methods inquiry into effects of research assessment and evaluative metrics.

Notes

1. See Waltman (2015) for the part of the review that discusses the technical literature on citation impact indicators.
2. cf. Hicks (2012) for a few other countries that also use citation scores in their evaluation systems.

References

- Aagaard, K. (2015), ‘How Incentives Trickle Down: Local Use of a National Bibliometric Indicator System’, *Science and Public Policy*, 1–13.
- 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*, 1–12.
- Aksnes, D. W. and Rip, A. (2009), ‘Researchers’ Perceptions of Citations’, *Research Policy*, 38/6: 895–905.
- Allred, P. and Miller, T. (2007), ‘Measuring What’s Valued or Valuing What’s Measured? Knowledge Production and the Research Assessment Exercise’, in V. Gillies and H. Lucey (eds) *Power, Knowledge and the Academy: The Institutional Is Political*, pp. 147–67. Basingstoke: Palgrave.
- Archambault, É. and Larivière, V. (2009), ‘History of the Journal Impact Factor: Contingencies and Consequences’, *Scientometrics*, 79/3: 635–49.
- Åström, F., and Hansson, J. (2013), ‘How Implementation of Bibliometric Practice Affects the Role of Academic Libraries’, *Journal of Librarianship and Information Science*, 45/4: 316–22.
- Boltanski, L. and Thévenot, L. (2006), *On Justification: Economies of Worth*. Princeton, NJ: Princeton University Press.
- Braun, T. (2012), ‘Editorial’, *Scientometrics*, 92/2: 207–8.
- Buela-Casal, G. and Zych, I. (2012), ‘What do the Scientists Think About the Impact Factor?’, *Scientometrics*, 92/2: 281–92.
- Butler, L. (2003a), ‘Explaining Australia’s Increased Share of ISI Publications—the Effects of a Funding Formula Based on Publication Counts’, *Research Policy*, 32/1: 143–55.
- Butler, L. (2003b), ‘Modifying Publication Practices in Response to Funding Formulas’, *Research Evaluation*, 12/1: 39–46.

- Butler, L. (2005), 'What happens when funding is linked to publication counts?', in H. Moed, W. Glänzel and U. Schmoch (eds) *Handbook of Quantitative Science and Technology Research*, pp. 389–405. Dordrecht: Springer.
- Butler, L. (2007), 'Assessing University Research: A Plea for a Balanced Approach', *Science and Public Policy*, 34/8: 565–74.
- Butler, L. (2010), 'Impacts of Performance-based Research Funding Systems: A Review of the Concerns and the Evidence', in OECD, *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*. OECD Publishing.
- Burrows, R. (2012), 'Living with the h - index? Metric Assemblages in the Contemporary Academy', *The Sociological Review*, 60/2: 355–72.
- Chandler, J., Barry, J., and Clark, H. (2002), 'Stressing Academe: The Wear and Tear of the New Public Management', *Human Relations*, 55/9: 1051–69.
- Chew, M., Villanueva, E. V., and Van Der Weyden, M. B. (2007), 'Life and Times of the Impact Factor: Retrospective Analysis of Trends for Seven Medical Journals (1994–2005) and their Editors' Views', *Journal of the Royal Society of Medicine*, 100/3: 142–50.
- Collini, S. (2015), 'Defending Universities: Argument and Persuasion', *Power and Education*, 7/1: 29–33.
- Colwell, R. et al. (2012), *Informing Research Choices: Indicators and Judgment. The Expert Panel on Science Performance and Research Funding*. Ottawa: Council of Canadian Academics.
- Cozzens, S., and Melkers, J. (1997), 'Use and Usefulness of Performance Measurement in State Science and Technology Programs', *Policy Studies Journal*, 25/3: 425–35.
- Craig, R., Amernic, J., and Tourish, D. (2014), 'Perverse Audit Culture and Accountability of the Modern Public University', *Financial Accountability and Management*, 30/1: 1–24.
- Cronin, B. and Sugimoto, C. (eds) (2014), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. Cambridge, MA: MIT Press.
- Dahler-Larsen, P. (2012), *The Evaluation Society*. Stanford CA: Stanford Business Books.
- Dahler-Larsen, P. (2014), 'Constitutive Effects of Performance Indicator Systems', *Public Management Review*, 16/7: 969–86.
- De Boer, H. F., Enders, J., and Leislyte, L. (2007), 'Public Sector Reform in Dutch Higher Education: The Organisational Transformation of the University', *Public Administration*, 85/1: 27–46.
- De Boer, H., Enders, J. and Schimank, U. (2007), *On the Way Towards New Public Management? The Governance of University Systems in England, the Netherlands, Austria, and Germany*. Netherlands: Springer.
- Deem, R., Hillyard, S., and Reed, M. (2007), *Knowledge, Higher Education and the New Managerialism: The Changing Management of UK Universities*. Oxford: Oxford University Press.
- Demšar, F. and Južnić, P. (2013), 'Transparency of Research Policy and the Role of Librarians', *Journal of Librarianship and Information Science*, 46: 139–47.
- de Rijcke, S., and Rushforth, A. D. (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*, 66:1954–58. Article first published online: 13 May 2015, DOI: 10.1002/asi.23382.
- de Rijcke, S., Wallenburg, I., Bal, R., and Wouters, P. (in press), 'Comparing Comparisons. On Rankings and Accounting in Hospitals and Universities', in M. Guggenheim, J. Deville and Z. Hrdlickova (eds) *Practising Comparison: Logics, Relations, Collaborations*. Manchester: Mattering Press, doi:10.1093/reseval/rvv040.
- Derrick, G. E. and Gillespie, J. (2013), '"A Number You Just Can't Get Away from": Characteristics of Adoption and the Social Construction of Metrics Use by Researchers', in S. Hinze and A. Lottmann (eds) *Proceedings of the 18th International Conference on Science and Technology Indicators*, pp. 104–16. Berlin.
- Derrick, G. E. and Pavone, V. (2013), 'Democratising Research Evaluation: Achieving Greater Public Engagement with Bibliometrics-informed Peer Review', *Science and Public Policy*, 40/5: 563–75.
- Espeland, W. N. and Sauder, M. (2007), 'Rankings and Reactivity: How Public Measures Recreate Social Worlds', *American Journal of Sociology*, 113/1: 1–40.
- Espeland, W. N. and Stevens, M. L. (1998), 'Commensuration as a Social Process', *Annual review of sociology*, 24: 313–43.
- Fealing, K., Lane, J., Marburger J. III, and Shipp, S. (eds) (2011), *The Science of Science Policy. A Handbook*. Stanford, CA: Stanford University Press.
- Feller, I. (2009), 'Performance Measurement and the Governance of American Academic Science', *Minerva*, 47/3: 323–44.
- Garfield, E. (2006), 'The History and Meaning of the Journal Impact Factor', *JAMA*, 295/1: 90–3.
- Georgiou, L. et al. (2000), *Impact of the Research Assessment Exercise and the Future of Quality Assurance in the Light of Changes in the Research Landscape*. Manchester: Policy Research in Engineering, Science and Technology. <<https://research.mbs.ac.uk/INNOVATION/Portals/0/docs/raec.pdf>> accessed 29 November 2014.
- Gill, R. (2009), 'Breaking the Silence: The Hidden Injuries of Neo-liberal Academia', in *Secrecy and Silence in the Research Process: Feminist reflections*, pp. 228–44. London: Routledge.
- Gläser, J. (2010), 'Concluding Reflections: From Governance to Authority Relations?', in R. Whitley, J. Gläser and L. Engwall (eds) *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and Their Consequences for Intellectual Innovation*, pp. 357–69. Oxford: Oxford University Press.
- Gläser, J. (2013), *How Does Governance Change Research Content? On the Possibility of a Sociological Middle-range Theory Linking Science Policy Studies to the Sociology of Scientific Knowledge*. Berlin: The Technical University Technology Studies Working Papers.
- Gläser, J., Lange, S., Laudel, G., and Schimank, U. (2010), 'The Limits of Universality: How Field-specific Epistemic Conditions Affect Authority Relations and their Consequences', in R. Whitley, J. Gläser and L. Engwall (eds) *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation*, pp. 291–324. Oxford: Oxford University Press.
- Gläser, J. and Laudel, G. (2001), 'Integrating Scientometric Indicators into Sociological Studies: Methodical and Methodological Problems', *Scientometrics*, 52/3: 411–34.
- Gläser, J., Laudel, G., Hinze, S. and Butler, L. (2002), 'Impact of Evaluation-based Funding on the Production of Scientific Knowledge: What to Worry about, and How to Find Out', Report for the German Ministry for Education and Research. <<http://www.laudel.info/wp-content/uploads/2013/pdf/research%20papers/02expertiseglaehinbut.pdf>> accessed 29 November 2014.
- Hammarfelt, B. and De Rijcke, S. (2015), 'Accountability in Context: Effects of Research Evaluation Systems on Publication Practices, Disciplinary Norms and Individual Working Routines in the Faculty of Arts at Uppsala University', *Research Evaluation*, 24/1: 63–77.
- Hammarfelt, B., Nelhans, G., Eklund, P., and Åström, F. (n.d.) 'The Heterogeneous Landscape of Bibliometric Indicators: Evaluating Models for Allocating Resources at Swedish Universities', (Manuscript under review) doi:10.1093/reseval/rvv040.
- Hare, P. G. (2003), 'The United Kingdom's Research Assessment Exercise: Impact on Institutions, Departments, Individuals', *Higher Education Management and Policy*, 15: 43–62.
- Hargens, L. L. and Schuman, H. (1990), 'Citation Counts and Social Comparisons: Scientists' Use and Evaluation of Citation Index Data', *Social Science Research*, 19/3: 205–21.
- Harley, S. (2002), 'The Impact of Research Selectivity on Academic Work and Identity in UK Universities', *Studies in Higher Education*, 27/2: 187–205.
- Harley, D., Acord, S. K., Earl-Novell, S., Lawrence, S., and King, C. J. (2010), *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*. Center for

- Studies in Higher Education. <<https://escholarship.org/uc/item/15x7385g>> accessed 29 November 2014.
- Hemlin, S. (2006), 'Creative Knowledge Environments for Research Groups in Biotechnology. The Influence of Leadership and Organisational Support in Universities and Business Companies', *Scientometrics*, 67/1: 121–142.
- Hessels, L. K., van Lente, H. and Smits, R. E. H. M. (2009), 'In Search of Relevance: The Changing Contract between Science and Society', *Science and Public Policy*, 36: 387–401.
- Hicks, D. (2012), 'Performance-based University Research Funding Systems', *Research Policy*, 41/2: 251–61.
- Hoecht, A. (2006), 'Quality Assurance in UK Higher Education: Issues of Trust, Control, Professional Autonomy and Accountability', *Higher Education*, 51/4: 541–63.
- Huth, E. J. (1986), 'Irresponsible Authorship and Wasteful Publication', *Annals of Internal Medicine*, 104/2: 257–9.
- Ingwersen, P. and Larsen, B. (2014), 'Influence of a Performance Indicator on Danish Research Production and Citation Impact 2000–12', *Scientometrics*, 101/12: 1325–44.
- Jensen, C. B. (2011), 'Making Lists, Enlisting Scientists', *Science Studies*, 24/2: 64–84.
- Jiménez-Contreras, E., de Moya Anegón, F. and López-Cózar, E. D. (2003), 'The Evolution of Research Activity in Spain: The Impact of the National Commission for the Evaluation of Research Activity (CNEAI)', *Research Policy*, 32/1: 123–42.
- Keenoy, T. (2005), 'Facing Inwards and Outwards at Once: The liminal Temporalities of Academic Performativity', *Time and Society*, 14/2–3: 303–21.
- Keevers, L., Treleaven, L., Sykes, C. and Darcy, M. (2012), 'Made to Measure: Taming Practices with Results-based Accountability', *Organisation Studies*, 33/1: 97–120.
- Lamont, M. (2012), 'Toward a Comparative Sociology of Valuation and Evaluation', *Annual Review of Sociology*, 38: 201–21.
- Laudel, G. and Gläser, J. (2006), 'Tensions between Evaluations and Communication Practices', *Journal of Higher Education Policy and Management*, 28/3: 289–95.
- Laudel, G. and Gläser, J. (2014), 'Beyond Breakthrough Research: Epistemic Properties of Research and their Consequences for Research Funding', *Research Policy*, 43/7: 1204–16.
- Leisyte, L. and Dee, J. R. (2012), 'Understanding Academic Work in a Changing Institutional Environment', *Higher Education: Handbook of Theory and Research*, 27: 123–206.
- Linkova, M. (2014), 'Unable to Resist: Researchers' Responses to Research Assessment in the Czech Republic', *Human Affairs*, 24/1: 78–88.
- Luukkonen, T. (2012), 'Conservatism and Risk-taking in Peer Review: Emerging ERC Practices', *Research Evaluation* 21/1: 48–60.
- MacRoberts, M. H. and MacRoberts, B. R. (1989), 'Problems of Citation Analysis: A Critical Review', *Journal of the American Society for Information Science* 40/5: 342–9.
- Martin, B. and Whitley, R. (2010), 'The UK Research Assessment Exercise: A Case of Regulatory Capture', in R. Whitley, J. Gläser and L. Engwall (eds) *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation*, pp. 51–80. Oxford: Oxford University Press.
- McNay, I. (1998), 'The Research Assessment Exercise (RAE) and After: "You Never Know How it Will All Turn Out."', *Perspectives: Policy and Practice in Higher Education*, 2/1: 19–22.
- Mingers, J. and Willmott, H. (2013), 'Taylorizing Business School Research: On the "One Best Way" Performative Effects of Journal Ranking Lists', *Human Relations*, 66/8: 1051–73.
- Moed, H. F. (2008), 'UK Research Assessment Exercises: Informed Judgments on Research Quality or Quantity?', *Scientometrics*, 74/1: 153–61.
- Moed, H. F., Van Leeuwen, T. N. and Reedijk, J. (1996), 'A Critical Analysis of the Journal Impact Factors of Angewandte Chemie and the Journal of the American Chemical Society Inaccuracies in Published Impact Factors Based on Overall Citations Only', *Scientometrics*, 37/1: 105–16.
- Moore, W. J., Newman, R. J., Sloane, P. J. and Steely, J. D. (2002), *Productivity Effects of Research Assessment Exercises*. University of Aberdeen, Department of Economics.
- Musselin, C. (2013), 'How Peer Review Empowers the Academic Profession and University Managers: Changes in Relationships between the State, Universities and the Professoriate', *Research Policy*, 42/5: 1165–73.
- Nedeva, M., Boden, R. and Nugroho, Y. (2012), 'Rank and File: Managing Individual Performance in University Research', *Higher Education Policy*, 25/3: 335–60.
- Ossenblok, T. L., Engels, T. C. and Sivertsen, G. (2012), 'The Representation of the Social Sciences and Humanities in the Web of Science—A Comparison of Publication Patterns and Incentive Structures in Flanders and Norway (2005–9)', *Research Evaluation*, 21/4: 280–90.
- Osuna, C., Cruz-Castro, L. and Sanz-Menendez, L. (2011), 'Overturning Some Assumptions about the Effects of Evaluation Systems on Publication Performance', *Scientometrics*, 86: 575–92.
- Paradeise, C. and Thoenig, J. C. (2013), 'Academic Institutions in Search of Quality: Local Orders and Global Standards', *Organisation Studies*, 34/2: 189–218.
- Petersohn, S. (2014), 'Bibliometric Services in Research Evaluation: A New Task Area Strengthening the Jurisdiction of Academic Librarians', 35th IATUL Conference 2014, Espoo, 03.06.2014.
- Pfeffer, J. and Salancik, G. R. (2003), *The External Control of Organisations: A Resource Dependence Perspective*. Stanford, CA: Stanford University Press.
- Pontille, D. and Torny, D. (2010), 'The Controversial Policies of Journal Ratings: Evaluating Social Sciences and Humanities', *Research Evaluation*, 19/5: 347–60.
- Porter, T. (1995), *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NY: Princeton University Press.
- Power, M. (1997), *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P. and Stirling, A. (2012), 'How Journal Rankings Can Suppress Interdisciplinary Research: A Comparison between Innovation Studies and Business & Management', *Research Policy*, 41/7: 1262–82.
- Reale, E. and Seeber, M. (2013), 'Instruments as Empirical Evidence for the Analysis of Higher Education Policies', *Higher Education*, 65/1: 135–51.
- Rebora, G. and Turri, M. (2013), 'The UK and Italian Research Assessment Exercises Face to Face', *Research Policy*, 42/9: 1657–66.
- Rousseau, S. and Rousseau, R. (2012), 'Interactions between Journal Attributes and Authors' Willingness to Wait for Editorial Decisions', *Journal of the American Society for Information Science and Technology*, 63/6: 1213–25.
- Rushforth, A. and De Rijcke, S. (2015), 'Accounting for Impact? The Journal Impact Factor and the Making of Biomedical Research in the Netherlands', *Minerva*, 53: 117–39.
- Sá, C. M., Kretz, A., and Sigurdson, K. (2013), 'Accountability, Performance Assessment, and Evaluation: Policy Pressures and Responses from Research Councils', *Research Evaluation*, 22/2: 105–17.
- Sauder, M. and Espeland, W. N. (2009), 'The Discipline of Rankings: Tight Coupling and Organisational Change', *American Sociological Review*, 74/1: 63–82.
- Schneider, J. W. (2009), 'An Outline of the Bibliometric Indicator Used for Performance-based Funding of Research Institutions in Norway', *European Political Science*, 8/3: 364–78.
- Shore, C. (2008), 'Audit Culture and Illiberal Governance Universities and the Politics of Accountability', *Anthropological Theory*, 8/3: 278–98.
- Shore, C. (2010), 'Beyond the Multiversity: Neoliberalism and the Rise of the Schizophrenic University', *Social Anthropology*, 18/1: 15–29.
- Shore, C., and Wright, S. (2000), 'The Rise of Audit Culture in Higher Education. In: M., STRATHERN (ed.) *Audit Cultures*. London: Routledge.
- Sousa, S. B. and Brennan, J. L. (2014), 'The UK Research Excellence Framework and the Transformation of Research Production', in C. Musselin and P. Teixeira (eds) *Reforming Higher Education*, pp. 65–80. Dordrecht: Springer.

- Stephan, P. -E. (2012), *How Economics Shapes Science*. Cambridge, MA: Harvard University Press.
- Stöckelová, T. (2014), 'Power at the Interfaces: The Contested Orderings of Academic Presents and Futures in A Social Science Department', *Higher Education Policy*, 27/4: 435–51.
- Strathern, M. (ed.) (2000), *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy*. London: Routledge.
- Taylor, J. (2011), 'The Assessment of Research Quality in UK Universities: Peer Review or Metrics?', *British Journal of Management*, 22/2: 202–17.
- Thomas, R. and Davies, A. (2005), 'Theorizing the Micro-politics of Resistance: New Public Management and Managerial Identities in the UK Public Services', *Organisation Studies* 26/5: 683–706.
- Van Dalen, H. P. and Henkens, K. (2012), 'Intended and Unintended Consequences of a Publish-or-Perish Culture: A Worldwide Survey', *Journal of the American Society for Information Science and Technology*, 63/7: 1282–93.
- Van der Meulen, B. (2007), 'Interfering Governance and Emerging Centres of Control: University Research Evaluation in the Netherlands', in R. Whitley and J. Gläser (eds) *The Changing Governance of the Sciences*, pp. 191–203. Dordrecht: Springer.
- Van Noorden, R. (2010), 'Metrics: A Profusion of Measures', *Nature*, 465/7300: 864–6.
- Waltman, L. (2015), 'A Review of the Literature on Citation Impact Indicators', *Arxiv* 1507.02099. [v1] Wed, 8 Jul 2015 11:21:58 GMT.
- Weingart, P. (2005), 'Impact of Bibliometrics Upon the Science System: Inadvertent Consequences?', *Scientometrics*, 62/1: 117–31.
- Whitley, R. (1984), *The Intellectual and Social Organisation of the Sciences*. Oxford: Oxford University Press.
- Whitley, R. (2000), *The Intellectual and Social Organisation of the Sciences*, 2nd edn. Oxford: Oxford University Press.
- Whitley, R. (2007), 'Changing Governance of the Public Sciences', in R. Whitley and J. Gläser (eds) *The Changing Governance of the Sciences. Sociology of the Sciences Yearbook*, pp. 3–27. Dordrecht: Springer.
- Whitley, R. (2011), 'Changing Governance and Authority Relations in the Public Sciences', *Minerva* 49/4: 359–85.
- Whitley, R. and Gläser, J. (eds.) (2007), *The Changing Governance of the Sciences. Sociology of the Sciences Yearbook*. Dordrecht: Springer.
- Whitley, R., Gläser, J., and Engwall, L. (eds.) (2010), *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation*. Oxford: Oxford University Press.
- Willmott, H. (2011), 'Journal List Fetishism and the Perversion of Scholarship: Reactivity and the ABS List', *Organisation*, 18/4: 429–42.
- Wilson, M. and Holligan, C. (2013), 'Performativity, Work-related Emotions and Collective Research Identities in UK University Education Departments: An Exploratory Study', *Cambridge Journal of Education*, 43/2: 223–41.
- 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 B. Cronin and C. R. Sugimoto (eds) *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, pp. 47–66. Cambridge MA: MIT press.
- Wouters, P., Thelwall, M., Kousha, K., Waltman, L., de Rijcke, S., Rushforth, A., and Franssen, Th. (2015), *The Metric Tide: Literature Review (Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management)*. Bristol, UK. HEFCE. DOI: 10.13140/RG.2.1.5066.3520.