Engaging Science, Technology, and Society 3 (2017), 21-40 DOI:10.17351/ests2017.108

Implicated in the Indicator Game? An Experimental Debate

MAXIMILIAN FOCHLER[1]
UNIVERSITY OF VIENNA

SARAH DE RIJCKE[2]
LEIDEN UNIVERSITY

Abstract

The rise of new modes evaluating academic work has substantially changed institutions and cultures of knowledge production. This has been reflected and criticized in the literature in STS and beyond. For STS scholars, these debates (should) however have an even more specific dimension. Many of us are experts on aspects of these changes. But at the same time, we too are part of the processes we are analyzing, and often criticizing. To put it slightly provocatively, often we cannot avoid playing the very same game that we scrutinize. This creates tensions that many of us reflect on, and it certainly has created many implicit and explicit normative stances on how to deal with them. Yet it seems that so far there has been little room in our field to reflect on and exchange this particular kind of experience-based knowledge. There are many different ways to engage with the dynamics of evaluation, measurement and competition in contemporary academia, or to play what we refer to colloquially here as the "indicator game." With this debate, we would like to give room to the expression and discussion of some of these ways. This text is the introduction and prompt to an experimental debate. We discuss the state of the academic discussion on the impact of indicator-based evaluation on academic organization, epistemic work and identities. We use insights from these debates to raise questions for how STS and STSers themselves deal with the indicator game. In conclusion, we summarize our contributors' arguments and propose the concept of "evaluative inquiry" as a new way of representing the quality of STS work in evaluative contexts.

Keywords

evaluation; metrics; indicators; STS; debate

[1] Maximilian Fochler, Email: maximilian.fochler@univie.ac.at
[2] Sarah de Rijcke, Email: s.de.rijcke@cwts.leidenuniv.nl

"Let me repeat that we are witnessing an effect that we (practitioners in higher education) have helped produce. Auditors are not aliens: they are a version of ourselves."
(Strathern 1997, 319)

Introduction

The rise of new modes of measuring and evaluating academic work has substantially changed scientific institutions and cultures of knowledge production. Work in STS and related fields has documented part of these changes, and has linked them to broader currents such as the *evaluation society* (Dahler-Larsen 2012, Power 1997), *new public management* (Burrows 2012, Shore 2008), or *academic capitalism* (Hoffman 2011, Slaughter and Rhoades 2004, Fochler 2016). Many feel that the ritual evaluation and surveillance of individuals, departments and entire institutions has become oppressively dense and are hindering rather than fostering what they see as good academic work (Sparkes 2007, Felt, Barben, et al. 2013, Mills and Ratcliffe 2012). Countless critical debates indicate that similar points are raised in other fields as well (Alberts et al. 2014, Hicks et al. 2015). Often, however, the pervasive impact of indicators is described as a form of coercion through an outside force, neglecting the ways in which academic actors *themselves* are implicated in indicator games and the new strategic possibilities they offer, in institutions, careers and beyond.

For scholars in STS, these debates (should) however have an even more specific dimension. Many of us are experts on important aspects of these changes, such as on the effects of quantification and standardization or on the change dynamics of cultures of knowledge production. But at the same time, we too are of course part of the changes we are analyzing, and often criticizing. Other "versions of ourselves" have to find ways to live within the world brought about by these changes. To put it slightly provocatively, often we cannot avoid playing the very same game that we scrutinize in our academic work. This creates tensions that many of us reflect on privately, and it certainly has created many individual implicit and explicit normative stances on how to deal with them. Yet it seems that so far there has been little room in our field to reflect on and exchange this particular kind of experience-based knowledge. There are many different ways to engage with the dynamics of evaluation, measurement and competition in contemporary academia, or to play what we refer to colloquially here as the "indicator game." With this debate, we would like to give room to the expression and discussion of some of these ways.

Our agenda is twofold. The first is political, at least in the sense of the politics of our own profession. We believe that scholars in STS should take a stance on how to deal with positive and problematic governance dynamics in academia, both individually and collectively. Over the past couple of years, a number of declarations, manifestos and reports on these issues were published, and though some of us were involved in these initiatives (Hicks et al. 2015, Wilsdon et al. 2015, DORA 2012), STS as a professional community has remained rather silent. There may be good reasons for this silence, and there may be a lot of back-channel activity that by definition remains invisible to others. We hope to generate discussion on exactly these issues. We feel that any position on these issues for STS should go beyond armchair reflection, and needs to be grounded

in a deep reflexivity of how we deal with these dynamics in our own practices. We hope that this debate will be a crystallization point for such reflections.

The second—related—dimension of our agenda is epistemic. To some extent, STS has already contributed to understanding the dynamics and problems of the indicator game. But many contributions are in classic academic outlets and formats, and derive from analyzing others rather than ourselves. We believe that, at least for the sake of a debate, putting ourselves in the spotlight and reflecting on our own dealings with these issues, can produce a different kind of knowledge. Maybe a knowledge that is more sensitive to all the ambivalences and contradictions inherent in the topic (Might it be that sometimes we also find joy in the indicator game, despite all its oppressive potential?), but hopefully also a kind of knowledge that can open perspectives on new ways of individual and collective action and of situated intervention (Zuiderent-Jerak and Bruun Jensen 2007, Zuiderent-Jerak 2015).

Not surprisingly, then, this debate is an experiment both in form and epistemic content. In developing it, we tried to stage a process in which interested scholars could shape their individual stance and perspective on these questions, in interaction with others. In epistemic terms, the challenge was how to develop interventions that draw on our analytical capacities and experiences as STSers, and potentially create new synergies across them. The authors have found different, though equally creative, solutions to this problem, both in terms of arguments and writing styles.

As a process, this debate started as a lunch session—organized by Maximilian Fochler, Martina Merz and Sarah de Rijcke at the 4S conference in Denver—focused both on new modes of measuring and evaluating academic work and on the ways in which STS scholars are implicated in analyzing, living, and intervening in "indicator games." We invited seven scholars to give brief input statements for the workshop, speaking from their different roles and responsibilities, as talented young researchers, as policy advisors, research managers or as editors and publishers—but above all: as individual scholars that live with and navigate performance indicators and other evaluative registers. Most contributors to the work published here took part in this vivid and reflexive session, which interestingly, though held in the US, attracted a mainly European audience. At the debate, there was a strong wish to continue the conversation, and to make it available to a wider STS audience. *Engaging Science, Technology and Society* provided us with the opportunity to do so in an experimental format.

To facilitate the transition from a spoken to a written format, as a next step we (Fochler/de Rijcke) wrote a prompt text, summarizing what we saw as some of the most important questions on the topic emerging from the existing literature and the Denver debate. All sections of this introduction, with the exception of the first and the last, are taken directly from this prompt text. Then, all contributors wrote a first version of their intervention, building on their Denver debate statements and reacting to our prompt as they saw fit. To facilitate cohesion in the debate, the contributions were then circulated, along with feedback from us conveners to the authors on potential improvements of the individual texts, but also on synergies and tensions with the other contributions. Building on this and their reading of the others' texts, the authors revised their texts. These revisions were then sent out by *ESTS* for anonymous peer review,

followed by a second round of revisions to address the reviewers' comments. As a final step, the entire collection of accepted papers was sent to invited commentators, Daniel Lee Kleinman, Maureen McNeil, and Marilyn Strathern, and their commentaries are published as part of this *ESTS* Thematic Collection.

The next three sections describe what we as conveners put forward as central impacts and dynamics of the indicator game, and as questions our debate contributors were invited to answer in relation to their own practices. In the last section, we summarize the different contributions to our experiment, and comment on their implications for how STS as field deals with the "indicator game."

Indicators, Competition and Organization

How have indicators changed the way academia and its practices are organized, and in doing so, how have they affected the way we live and work together as academics? Historically, the development of indicators has been driven by the agenda to provide transparency and orientation in processes that have always been a constitutive facet of science: the competition for resources and peer recognition (Wouters 2014). Indeed, one could argue that the appropriation of indicators was facilitated by the fact that they smoothly latched onto existing infrastructures for competition, reputation management, and research assessment (e.g. peer review). In addition, and not surprisingly from an STS perspective, indicators as new metric infrastructures have not only represented competitive positions, but also thoroughly changed how academic organizations and individual academics assess and compete (Espeland and Sauder 2007).

Current evaluation mechanisms and indicators have been inextricably connected with the rise of the new forms of institutional governance often conveniently summarized under the label *New Public Management* (Shore 2008, Sparkes 2007, Burrows 2012). These have introduced market-like structures in academia and other parts of the public sector, and have established accountability, competitive performance and performance measurement as *the* gold standard route to productivity and (cost-) efficiency for academic institutions and individuals alike. The first contours of this "move to metrics" were already visible in some countries in the late 1970s and early '80s (the Dutch and UK science systems are good examples). Following Neyland, Milyaeva, and Ehrenstein (2015), one could argue that the "problem" of academic research became constituted as an "absence" in this period: concretely, an absence of competition, assessment, and oversight. The "solution" to this deficiency was sought in the further institutionalization of evaluation mechanisms and performance metrics, which in turn broadened up the space for logics of competition and the accumulation of reputation measures through quantitative means.

That the appropriation of metric systems subsequently spread out on a larger scale was further stimulated, on the one hand, by a stronger emphasis on national economic goals such as innovation and growth (Beerkens 2003, Jacob and Meek 2013, Musselin 2014) and, on the other hand, by rather pervasive pressures to "internationalize" national science policy systems and evaluation procedures (Linková and Stöckelová 2012, Paradeise and Thoenig 2013). Increasingly,

states have tied institutional funding to performance in (international) competition, and higher education institutions are handing down this pressure to their faculties, departments and ultimately to the scientists they employ. This adds and institutionalizes a layer in the governance of science and scholarship concerned with gathering increasingly detailed numerical information about research. As a result, we are witnessing a shift in how quality is defined in academic work. Increasingly, "excellence" is seen in purely relational terms, as the ability to perform in a top small segment of those measured in the related infrastructures (Espeland and Sauder 2007, Felt, Barben, et al. 2013). Higher education institutions fight for top positions in rankings, individual academics strive to be among the tiny top percentage funded by specific sources or published in selective outlets, and even scientific journals compete to be listed in the top percentage of their respective fields.

In the last decade, these competitive dynamics have been exacerbated by a relatively sudden shift from an—unsustainable—growth of some segments of science to austerity as the new basic *modus operandi* in many higher education systems (Alberts et al. 2014, Stephan 2012). 20th century investments have increased both the number of people working or aspiring to work in science as well science's appetite for resources in an increasingly project- and technology-driven environment. Over the last years, austerity and stagnating state financing have dampened this unfettered growth, often creating considerable dynamics of competition for scarce resources, be they faculty positions or grant income.

But how are the organization of academic work and socio-epistemic practices affected by this competition? Some studies argue that the way different academic fields and institutions react to changes in their external governance and evaluative infrastructures, such as the introduction of indicator-driven competition, is complex and cannot be generalized (Gläser et al. 2010, Paradeise and Thoenig 2013). Others have argued that academic institutions find it particularly hard to buffer the impact of indicator-driven competition on their internal organizational and evaluation structures, and have largely begun to internalize the corresponding definition of quality in their own organizational cognition (Sauder and Espeland 2009). What has worth in a scientific institution hence tends ever more strongly to coincide with what can be expressed in evaluative metrics, and organizational structures are changed to foster the production of these particular forms of worth.

The logics and dynamics of early scientific careers are particularly salient examples of the impact of this changed organizational cognition on researchers' practices. This is first related to the extreme competition for tenured positions in a system that has grown largely by multiplying temporary project-funded, early-career forms of employment, without a matching increase in positions with a long-term perspective (Stephan 2012). Second, the temporal texture of academic work has also changed for young researchers, with short-term employment contracts and coercive mobility multiplying the temporal moments at which performance needs to be demonstrated in competition (Ylijoki 2010). Recent contributions have argued that this pushes junior researchers to strongly focus on individual productivity, and that this changes their perception of cooperation in research and of the places and groups in which they work. Particularly in competitive fields, young researchers may be structurally discouraged from

engaging in types of work they cannot directly validate for their own track record (Fochler, Felt, and Müller 2016, Müller 2014).

The resulting organizational dynamics affect junior researchers, but they also pose new challenges to senior researchers in positions with organizational responsibility (Kaltenbrunner and de Rijcke 2016). Sparkes (2007) describes the dilemmas and bodily discomfort experienced by academic mid-level management struggling to both care for those they are responsible for and comply with the imperatives of indicator-drive institutional competition and the evaluative infrastructures that come with it. Davies and Horst (2015) stress the importance of considering not only the entrepreneurial roles of contemporary senior academics, but also their caring craftwork in building and defending their groups as epistemic and social spaces.

But it may not only be group leaders, deans or heads of departments who increasingly struggle to both assure the success of their collectives in indicator driven competition and care for them as social and intellectual spaces. Other academic institutions, such as scientific journals, may be facing similar problems as their standing is ever more strongly coupled with indicators. Recent editorial reflections portray editorial work as, on the one hand, challenged by the consequences of the increasing importance of indexed publications. This leads to a rising number of submissions to top-tier journals and the resulting problems in the organization of high-quality peer-review processes (Bauer 2016). On the other hand, there is a nascent debate about some editors' strategies to boost the impact factor of their own journal in competition with others, such as for example by coercing authors to adapt their citation practices to the strategic needs of the journal (Martin 2013, 2016).

It is debatable whether contemporary academics can escape the organizational changes sketched above. As indicators facilitate certain orders of worth and not others, and have the capacity to actively shape what gets obscured and what gets highlighted, "thinking with indicators" becomes a matter of survival for some (Müller and de Rijcke n.d.). The ways in which indicators latch onto dominant evaluative infrastructures can be quite consequential for what is considered as a "doable" research project, an "excellent" resume, and a valuable form of output. Deep concerns about some of the more problematic differentiating effects of indicators have resulted in high profile statements such as the San Francisco Declaration for Research Assessment (DORA 2012) and the Leiden Manifesto for Research Metrics (Hicks et al. 2015)—in both cases containing recommendations that reflect visions for the future of research governance that can only be made real after making fundamental changes to current infrastructures for funding, producing, and evaluating research. The statements, for instance, decree "responsible uses of metrics" in peer review; open and transparent data collection processes in assessments; recognition of variation between fields in their publication and citation practices; and the "need to assess research on its own merits" instead of on the reputation of the journal. But despite the concerted and perhaps somewhat idealistic attempts of such manifestos to change the rules of the indicator game, much of the scholarly debate remains quite critical about the impact of metrics on the scientific system.

So what can we as STS scholars do about these issues? How do we experience them, particularly with our sensibility to the effects of quantification, standardization, and

temporalization? Do we comply, resist or attempt to find tactics that allow us to walk paths between these two extremes? Our training in STS may give us tools to make explicit and analyze how the multiple forms of infrastructure related to research assessment are at once normative and technical, social and methodological. Are we ourselves at risk of "falling through the cracks" when the impact of our research is not easily translated into quantifiable form? What do indicators offer us in the organization of science? Is it a more transparent way of competing, and hence also better orientation for both junior and senior researchers than older, potentially more nepotistic and provincial ways of attributing scientific reputation? Or do the competitive pressures increasingly alienate us from the substance of our own work and the communities we are part of?

How do these competitive dynamics challenge us not only as individual researchers, but also in our professional roles? Are scholarly journals increasingly a selective clearinghouse for career aspirations, rather than a space of intellectual exchange? How do those with managerial duties balance between assuring the competitiveness of the collective they are responsible for and caring for its intellectual and social development? Are they forced to hand down the pressure to comply with ever more rigid definitions of quality materialized in indicators, or are they able play the game in a way which also generates spaces of creativity, spaces under less competitive pressure, at least for a time? And if so, how? Cui bono? And can it be otherwise?

The Epistemic Impact of Indicators

The growing use of indicators and their "constitutive effects" (Dahler-Larsen 2014) on academic life and work have been highlighted for some time by way of analyses on ranking (Hazelkorn 2015) and indicator use (Butler 2003, Weingart 2005). Yet, it is only recently that considerable interest has been directed towards the more intricate effects that these measures might have on concrete practices of knowledge production (Derrick and Gillespie 2013, Hammarfelt and de Rijcke 2015, Müller 2014, de Rijcke, Wouters, Franssen, et al. 2016, Wouters 2014). A recent review of the literature regarding evaluation practices and effects of performance indicators outlines at least two possible consequences that these measures might have (de Rijcke, Wouters, Rushforth, et al. 2016).

First, indicator use might lead to *goal displacement* where scoring high according to the assessment criteria becomes a goal in itself rather than a way of measuring whether certain goals are met (Butler 2003, 2005, Colwell et al. 2012). In publishing, this may result both in a gaming of metrics, such as in the emergence of strategic "citation cartels," as well as in the attempt to "salami" the results of a specific of project into a maximum amount of "least publishable units" (Martin 2013). The epistemic effects of the latter practice have been vividly debated in many fields (Dupps and Randleman 2012, Owen 2004). It has both been criticized for fragmenting the academic literature and swamping the reader with de-contextualized findings rather than providing larger synthesis and more complex arguments. But it has also been defended as better suited to a rapid accumulation of knowledge.

Second, criteria for measuring research quality often only focus on activities that can be counted, which in turn might lead to *task reduction* among researchers. Certain types of work (e.g. book publishing, societal engagement) might be abandoned for activities that are recognized as valuable in evaluation systems (Hammarfelt and de Rijcke 2015, Laudel and Gläser 2006). Such task reduction is especially worrisome in research fields that are heterogeneous when it comes to methods and dissemination channels (e.g. many fields in the social sciences and humanities)(Whitley and Gläser 2007). Particularly in these fields, task reduction may also result in a reduction and standardization of the forms and formats of presenting a scholarly argument. STS work has shown that the literary form and the substance of scientific arguments are inextricably linked (Bazerman 1988). A reduction of the formats of scholarly communication, such as in the shift from books to journal articles that is observable in at least some evaluation logics, is thus bound to have epistemic effects.

Most epistemic effects of *goal displacement* and *task reduction* are related to choices researchers make in publishing and presenting their research work. Central to these issues are also questions of authorship, what it means to be an author both ethically and practically, and how an increasing focus on performance measurement is potentially affecting attribution of authorship across disciplines (Biagioli and Galison 2003).

But do indicators also have effects on researchers' epistemic assessments and choices in the actual conduct of research work or the planning of projects? The results of some recent studies suggest that this could indeed be the case. Based on a study of biomedical research in the Netherlands, Rushforth and de Rijcke (2015) argue that indicators, among other functions, also serve as *judgment devices* in reflexively assessing the worth of a particular line of experimentation, both in relation to how well it can be published and the career value of the resulting publication. In this, indicator scores of journals are linked to epistemic properties of experiments, such as whether a study is purely descriptive or may be argued to also analyze biological mechanisms. Sigl (2015) has argued that in junior scientists' biographies in the contemporary academic life sciences epistemic risk is inextricably linked to career risk, as even temporary experimental failure and the resulting gap in productivity is likely to endanger a person's career. This may lead postdocs to choose less risky and more conservative research questions than PhD students, who are yet less fully exposed to the logics of contemporary academic careers, of which the indicator game is part and parcel (Fochler, Felt, and Müller 2016). In a similar vein, Felt, Igelsböck, et al. (2013) describe how PhD students in trans-disciplinary sustainability research struggle to align the wider agenda of their research with disciplinary career structures and the accountability regimes they are connected to.

So, results of STS work show that indicators indeed may have epistemological effects on how research is planned, done and communicated. And, as is often the case, we derive much of our knowledge about changes in contemporary research cultures from observing the natural sciences, and the life sciences in particular.

But how are our *own* practices of knowledge production influenced by metrics, if at all? In what ways are indicators having an influence on our own possible career paths as STS scholars? Does the fact that STS is an interdisciplinary field, and that many scholars thus (have

to) pursue placement in other disciplinary contexts, make a difference in this? Do we ourselves notice consequences of a shift toward metric-based research assessment for our own epistemic cultures and authorship practices? Do indicators affect how we choose to structure our research and write up our results? Is there indeed a tendency to salami results into least publishable results for publication, and, if so, how might this affect the epistemic development of our field? Are journal articles displacing books because career requirements and institutional evaluations privilege the former? If so, what are the epistemic consequences of such a shift? Do considerations of future career value guide the epistemic choices of junior scholars in our field, and, if so, in which ways? Does the indicator game foster or hinder innovative research in STS, and in which ways? And finally, again, which personal or institutional strategies do we have or might we develop to align our epistemic agendas and the demands of accountability in better ways?

Indicators and Identities

How we and our work are represented and valued influences how we perceive our own professional identity. It may lead us to change the way we think about being an academic, and also how we plan, organize and do research and teaching. Some of this change happens through institutional nudging or at times coercion, such as when research assessment exercises and their foreshadowed metrics influence our publication strategies. But perhaps the potentially more important and pervasive changes are not directly imposed from the outside. As has been argued in much detail, the most potent governance effects of the New Public Management come from a shift in the way we perceive ourselves and our work (Shore 2008, Power 1997, Strathern 1997). This shift happens as we internalize the rules of the indicator game.

Over the past decades, performance indicators have increasingly spread out beyond the narrow confines of output assessment, to be absorbed by other parts of academic life. Think for instance of how metrics are increasingly infiltrating teaching contexts: nowadays, many universities and faculties calculate the student/staff ratio in their courses in order to assess cost-effectiveness of their institutional teaching activities (Burrows 2012). Another example is how institutions now use elaborate survey-based student evaluation forms to assess the performance of lecturers, after which these assessments are not only used to improve the classes in terms of their content, but are also fed into the yearly appraisals of academics to assess their worth. But indicators have also emerged to assess more ephemeral academic virtues such as leadership potential, "professor-ability" and "international outlook."

The rhizomatic growth of the "evaluation machine" (Dahler-Larsen 2012) has led some to argue that academics are increasingly at risk of becoming overworked and overwhelmed by a "deep, affective, somatic crisis" (Burrows 2012, 355). According to Gill (2010), academics tend to justify working so hard because they possess a passionate drive for self-expression and experience pleasure in intellectual work. Paradoxically, Gill argues, it is this drive that feeds a whole range of disciplinary mechanisms and lets academics internalize a neoliberal subjectivity.

We play "the numbers game," as Burrows (2012) calls it, because of "a deep love for the 'myth' of what we thought being an intellectual would be like." (Gill 2010, 15)

So it seems that the rules of the indicator game are quite fundamentally affecting how we live, work, and feel as academics. Metrics and indicators may even become woven into the very socio-material fabric that shapes new academic selves, a process that is further encouraged by the many "little tools" (Asdal 2008) we use to represent, valuate and valorize our work along the way. In a recent analysis, Hammarfelt, de Rijcke, and Rushforth (2016) have explored how social networking platforms like Research Gate and academia.edu serve as infrastructures for self-quantification on the part of academics. They argue that these sites, and the metric infrastructures they tie into, are increasingly used to gauge and display the market standing of researchers, as well as provide them with incentives to work on this representation. Hammarfelt et al. tentatively diagnose a "gamification" in how users are presented with means and motivation to work on and increase their profile, and are rewarded with achievements as they do so. We may easily dismiss this as nerdy and vain, and criticize the marketing ideology that is seemingly introduced into academia through these portals. But maybe there is something else at work here. What if researchers actually find joy in working on their score, or in watching their h-index increase on their Google scholar profile? What if, in a system in which positive feedback is otherwise ever harder to acquire, this is a simple and seemingly controllable form of gratification? Also, the authors argue, researchers build profiles to track their own work in order to retain or regain control over their own metric representation. In doing so, they can decide how their work is being presented, and which indicators are used to this end.

Of course, not all of us are engaged in gaming our metrics every day. But these examples nevertheless point to important questions. How do our identities as academics change as the quality of academic work is increasingly expressed in metric indicators? How do we cope with metrics in our work, lives, identities, and careers, and what kinds of "hidden injuries" emerge (Gill 2010, 2014, Knights and Clarke 2014)? Burrows (2012, 355-357) concludes that we are all implicated in the "autonomization of metric assemblages" in the academy, and directly addresses us, his fellow-academics: "we need to obtain critical distance"; "we need to better understand ourselves as academics"; "why do we feel as we do?" Ironically, though, this focus on individual academics' affective states fits rather well with neoliberal policies that applaud personal autonomy and the individual's responsibility for one's own well-being and professional success. And it begs the question of our own agency in choosing, shaping and maybe resisting these indicators and the ways they represent our work. How do we feel about being lovingly coerced into working long hours, and spending our weekends writing grant proposals and articles, and grading students' exams? Does engaging in the indicator game also change the way we actually do our research work? If so, how does this affect our joy in being an academic, and the way we are able to convey the potential attractiveness of this occupation to our students? Could it be otherwise? Could we then be other, better, versions of ourselves?

Reflecting the Debate

As is due for an experimental format, our contributors reacted to our prompt in very different stylistic and argumentative ways. Some put their individual experience front and center, and used it as a lens to think through wider issues. Others took a more distanced and analytical writing stance, and enriched it with examples from their experience. The result, we hope, is a thought-provoking collection of interventions that open up a diversity of perspectives on the individual and collective challenges of dealing with the "indicator game." In this last section, we give a brief summary of the individual contributions, before moving on to comment on lines of argument that run across them and pointing to implications of these for STS.

Roland Bal (2017) analyzes the strategies his research group developed to deal with the interlinked dynamics of research evaluation in a multi-disciplinary institutional environment characteristic for much STS research. Scholars in such environments constantly need to navigate and negotiate the standards of evaluation, in complex choreographies of cooperation and competition with other disciplinary groups. Bal describes strategies his group has successfully used, and how these strategies have both shifted the way research quality is assessed within the department as well as changed the way his group works and publishes. In conclusion, he describes performance management systems and research practices as co-constituted and calls for a debate on which forms of evaluation infrastructures allow for better ways of doing research in STS.

Ulrike Felt's (2017) essay relates the growth of indicators to the shifting temporalities of academic work. Drawing on her own research but also on her professional experiences, she develops the notion of *chronopolitics* to analyze the politics of time governing academic knowledge production, work and evaluation. Drawing on a range of examples, she argues that the focus of academic work is increasingly shifting from a logic of discovery to one of delivery. In her conclusion, she suggests to move beyond a debate of how to best play the indicator game to a more fundamental critique of the entanglement of indicators and time, and to a re-timing of research as a political project.

Alan Irwin (2017) begins his contribution by arguing that indicators in themselves do not have effects, but that it is crucial to understand which questions in the wider culture and governance of academia their use speaks to. Drawing partly on his own experience in university management but also as an STS scholar, he argues that understanding the wider institutional logics is crucial if "better tunes" are to be developed in response to the indicator game: both to strengthen the standing and identity of the field and to support emerging scholars in this challenging context. His paper closes with an optimistic outlook, inviting STSers—and particularly our professional institutions—to celebrate and, at the same time, take a stand for our diverse values, practices and notions of quality.

Julien McHardy (2017) draws on experiences in the alternative publishing collective *Mattering Press*, and stresses the importance of opening up and maintaining "bewildering spaces within" the increasingly normative and clear-cut regimes of evaluation. Drawing on a range of examples from alternative publishing, McHardy illustrates how STS sensibilities can become productive in experimenting with new forms of academic communication. His essay embraces

the worth of practices that work around logics of audit and accountability to create space for otherness. At the same time, he also warns us that caring for difference is a struggle that includes making radical cuts and disconnections, and that in practice such caring often may build on precarious and invisible work.

Drawing on her own biography, Ruth Müller (2017) opens her contribution by reminding us of the strong interdisciplinary roots and entanglements of STS as a field, as well as of its emancipatory political agenda. This rich history creates multiple registers for defining the worth of work in our field, ranging from scholarly excellence over successful interdisciplinary collaborative engagement to assembling matters of care and concern. Building on this, she analyzes how current indicatorized career dynamics render it increasingly difficult to derive a sense of meaning from one's own work, which builds on a successful integration of these multiple registers of worth. She calls for developing "standards of our own," that is, ways of defining quality and processes of accountability that do better justice to the multiple valences of our field.

Drawing on her experiences as managing editor for two STS journals, Katie Vann (2017) looks at the rise of "journal impact factor (JIF) centrism" in academic practices as an instance of a more generalized form of "data deferral." She nests data deferral within the broader problematic of the increasing reliance of public-sector scholarly communities on (oligopolistic) commercial academic publishing houses, and shows how centrism for the JIF indicator leads to two different kinds of surplus. For scholarly communities, surplus emerges in the form of layers of knowledge effects and labor that remain foreclosed to formal recognition and are thereby reconstructed as dispensable (waste). For private sector publishing companies—which contribute to the publishing process primarily by providing content management/distribution platforms and transforming manuscripts into standardized digital objects that are amenable to indexing, aggregation, and comparative calculation—surplus emerges in the form of publishers' monetary surplus (profit).

Paul Wouters' (2017) essay is concerned with how to bridge the "evaluation gap" that many scholars experience as the criteria used in assessment do not seem to adequately represent what they themselves value in their work. He discusses the recent evaluation of his own center as an attempt to bridge this very gap, and reminds us that it is productive to see evaluations not as the (obviously impossible) attempt to produce a true representation of past work, but as explorations of "*w*ho one wants to be." Reflecting on this experience, he calls for more STS-inspired experimentation in the formats and processes of research evaluation.

We are very grateful to the contributors of this experimental project for their willingness to share and reflect their own strategies of dealing with the "indicator game." Let us give a brief overview of the strategies present in our collection.

For our more junior contributors in either precarious or pre-tenured tiered positions this seemed to involve a lot of *expectations management*. They saw themselves as quite literally not in a position to radically reform the indicator game. Instead, they undertake careful trust-building efforts with "othered" auditors at multiple levels (heads of departments, deans, grant evaluation panel members), aimed, for instance, at articulating the relevance of time spent on activities that

are notoriously hard to get credit for, such as administrative and civic duties. Other contributors make clear that playing the indicator game sometimes involves *circumventing the logics of indicator-driven evaluation* altogether. We witness this strategy in the contributions of some of our more seasoned colleagues who have become experienced in recognizing the rare opportunities for workarounds in otherwise highly routinized moments of audit (for instance by inviting evaluators to *read* instead of merely *count* output by providing PDF's of papers and book chapters where "only" a publication list is required). A third strategy, the *adapt-to-(make)-fit* strategy, is similarly used by contributors in an attempt to make good use of the co-constitution of evaluation systems, indicators, and scholarly work. In these cases, audit is re-labeled in positive terms as both organizational and individual learning, as development instead of judgment, as care and support instead of competition and opposition. Publication strategies are tweaked, goals are re-formulated, or calculation is used in such a way that it does not dominate but "keeps the edges tidy" (McHardy 2017). A fourth and final strategy could perhaps best be characterized as an attempt at *collective progressivism*. This response propagates a pro-active, collective, and also highly political use of indicators. It aims at counteracting some of the inhibiting and individualizing modes of metric-oriented accountability. The goal is to simultaneously play the game, expand the board, and diversify the pool of potential players. It is visible in attempts to turn institutional assessments into experimental, intellectually motivated future-making exercises instead of conservative adjudications of past performance (Wouters 2017). It is also visible in attempts to enhance the brand of open access, non-profit journals in our field through JIF-qualification (Vann 2017). Such attempts put the focus on the collective interest of, for instance, institutions or fields. They involve creatively and collectively imagining and practicing new ways of presenting and defining quality in evaluative contexts—as we suggest with the concept of *evaluative inquiry* in the next section. Existing indicators are re-defined as progressive technologies of intervention, to emancipate from instead of reinforce increasingly narrow scholarly publishing and assessment practices.

Re-configuring Evaluation

Interestingly, with the exception of perhaps the strategy of *collective progressivism*, the other strategies do not fundamentally question the basic playing field of the indicator game. However, many individual contributions do suggest that, collectively, STS has the capacity to change how the game is played, and they aim to shift it in a direction that has a better fit with our own values and concepts of quality.

So how can we develop tools and concepts that would enable ourselves, as auditors and researchers in STS, to assess and be assessed on appropriate terms, with criteria specific to STS? Here, we would like to pick up on our contributors' attempts to think through some foundational yet implicit characteristics of STS in relation to often similarly implicit notions of "quality" in the field. Interestingly, most contributors relate the tensions arising from assessment logics in their own work practices to larger existential questions about what STS is about, about the *raison d'etre* of STS, about what STS is and should be doing. We did not elicit this response in our prompt, but

the move makes sense in a discussion on criteria for assessing value in and of work practices in the field. There is a deep-felt commitment in all contributions to the disruptive, ideographic, political, and multivalent character of STS. All contributors celebrate how STS enables interdisciplinary, inductive and experimental (in the sense of inventive and unruly) "thinking against the grain." We share this commitment, and we believe that what it warrants now are more explicit, partly open-ended and reflexive accounts of quality in STS. Strong but non-essentializing accounts of the multiple "doings" of STS such as the ones shared in this collection could serve as excellent starting points for more productive evaluations of our work. Accounts of quality that give space to the "critical multidimensionality" (Rabinow 2011, 139) of STS, of the ecologies we are sustaining, of the objects we are analyzing, and of the corpus we are building. We would like to propose the concept of "evaluative inquiry" to capture this potentially more enabling, instead of reductive, approach to building accounts for assessment. "Evaluative inquiries" are not solely structured along the lines of externalizing explanations and metrics. They are also capable of representing the heterogeneous associations and practices that constitute our work. They draw in and (temporarily) hold together elements of STS practices that would remain neatly separated in standard evaluation accounts. They may represent STS work numerically, verbally, and/or visually – but aim to do so in ways which represent the complexity of actual practice and its engagements, rather than to reduce for the sake of standardization. Rather than alienation, they aim to produce and represent the meaning and purpose of STS work (Müller 2017), both in relation to external audiences and to and for ourselves. At the heart of the concept lies a different evaluative principle—based on more exploratory forms of engaging with STS. Evaluative inquiries perform a shift from a predominantly bureaucratic to more substantive modes of assessment. In this, a standardization of indicators and methods is less relevant than "staying with the trouble" (Haraway 2016); staying closer to the epistemic missions, frictions and resonances of the work under scrutiny.

For instance, let's consider what an evaluative inquiry into the recent work of an STS department would look like. Such an inquiry could treat the different activities and outputs of the department not primarily as calculative units, but as punctualizations of ongoing processes of meaningful work, as nodes in actor-networks. An inquiry inspired by such relational concepts would not be grounded in the vocabulary of "impact," but would try to visualize the relations to actors enrolled in the work (e.g. the NIH, the pharmaceutical industry, academic biomedicine), trace the issues that emerged in the assemblage (e.g. the role of third-party funding in steering research priorities), and describe how the department's research came to matter in the network (e.g. by how it has made visible the work of patient activist groups). This, of course, is just one possibility. Many more styles and formats of good evaluative STS inquiries are imaginable—of inquiries that do not reduce to fit with standardizing logics, but valorize STS work in its own terms in presenting it for evaluation.

Importantly, we do *not* think that evaluative inquiries should ban indicators altogether. But as a field we may want to put our knowledge of the world-making capacities of numbers (cf. using indicators as "tin openers," de Rijcke et al. 2016) to more productive uses, and we should

employ our collective expertise and authority to point to why and how numbers can never tell the whole story.

For the sake of our students and ourselves, we think it is time for clearer institutional commitments to developing explicit, yet multi-faceted and non-essentialist notions of quality in and for STS. As we see it, this would require a) institutional commitment to systematically experiment with, share, and build evaluative inquiries in STS; b) the acumen to synchronize these inquiries with existing evaluative repertoires; and c) institutional memory practices to stabilize the resulting accounts, celebrating quality in our field. More than just a move in the indicator game, and as an ambitious hope, this could also help to build an even richer collective identity for our field, foster stronger forms of its institutionalization, and help consolidate successful strands and places of STS research beyond individual achievements. Care for a game-changing move, anyone?

Acknowledgements

We thank all authors involved in this collection for their dedicated involvement in what was a highly pleasant intellectual process. We are particularly grateful to *ESTS*, Daniel Kleinman and Katie Vann for their willingness to experiment with formats beyond established conventions, and for their superb and caring editorial attention. The quality of the papers in this collection, including this one, has significantly benefited from the feedback of anonymous reviewers. Further, we would like to express our gratitude to the commentators invited by ESTS—Daniel Lee Kleinman, Maureen McNeil, and Marilyn Strathern—for their readiness to engage with the issues we are attempting to raise in this collection. And finally, we thank all participants present at the 4S Denver 2015 lunch meeting, where the enormously productive and enthusiastic discussion motivated us to take this into writing, and Martina Merz, who organized this initial lunch meeting with us.

Author Biographies

Maximilian Fochler is assistant professor and head of the Department of Science and Technology Studies of the University of Vienna, Austria. His main current research interests are forms of knowledge production at the interface of science and other societal domains (such as the economy), as well as the impact of new forms of governing science on academic knowledge production. He has also published on the relations between technosciences and their publics as well as on publics' engagement with science.

Sarah de Rijcke is associate professor and deputy director at the Centre for Science and Technology Studies (CWTS) of Leiden University, the Netherlands. She leads a research group that focuses on a) developing a theoretical framework on the politics of contemporary research governance; b) gaining a deep empirical understanding of how formal and informal evaluation practices are re-shaping academic knowledge production; c) contributing to shaping

contemporary debates on responsible research evaluation and metrics uses (including policy implications).

References

Alberts, B., M. W. Kirschner, S. Tilghman, and H. Varmus. 2014. "Rescuing US biomedical research from its systemic flaws." *Proceedings of the National Academy of Sciences* 111 (16):5773-5777. doi: 10.1073/pnas.1404402111.

Asdal, K. 2008. "On politics and the little tools of democracy: A down-to-earth approach." *Distinktion: Scandinavian Journal of Social Theory* 9 (1):11-26.

Bal, R. 2017. "Playing the Indicator Game: Reflections on Strategies to Position a Group in a Multidisciplinary Environment." *Engaging Science, Technology, and Society* 3: 41-52. DOI:10.17351/ests2016.111.

Bauer, M. W. 2016. "A word from the parting editor, 4760 pages down the line." *Public Understanding of Science* 25 (2):130-134. doi: 10.1177/0963662516629275.

Bazerman, C. 1988. *Shaping written knowledge : the genre and activity of the experimental article in science*, *Rhetoric of the human sciences*. Madison, Wis.: University of Wisconsin Press.

Beerkens, E. 2003. "Globalisation and higher education research." *Journal of Studies in International Education* 7 (2):128-148.

Biagioli, M., and P. Galison. 2003. *Scientific authorship: Credit and intellectual property in science*: Routledge.

Burrows, R. 2012. "Living with the h-index? Metric assemblages in the contemporary academy." *The Sociological Review* 60 (2):355-372. doi: 10.1111/j.1467-954X.2012.02077.x.

Butler, L. 2003. "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 *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*, edited by Henk F. Moed, Wolfgang Glänzel and Ulrich Schmoch, 389-405. Dordrecht: Springer Netherlands.

Colwell, R., M. Blouw, L. Butler, S. E. Cozzens, I. Feller, Y. Gingras, J. Hurtubise, G. Jordan, J. S. MacDonald, and M. Makarow. 2012. *Informing research choices: Indicators and judgment. The Expert Panel on Science Performance and Research Funding*. Ottawa: Council of Canadian Academies.

Dahler-Larsen, P. 2012. *The evaluation society*. Stanford, California: Stanford Business Books, an imprint of Stanford University Press.

Dahler-Larsen, P. 2014. "Constitutive effects of performance indicators: Getting beyond unintended consequences." *Public Management Review* 16 (7):969-986.

Davies, S. R., and M. Horst. 2015. "Crafting the group: Care in research management." *Social Studies of Science* 45 (3):371-393. doi: 10.1177/0306312715585820.

Derrick, G., and J. Gillespie. 2013. "“A number you just can’t get away from”: Characteristics of adoption and the social construction of metrics use by researchers’." Proceedings of the 18 th International Conference on Science and Technology Indicators.

DORA. 2012. "San Francisco Declaration on Research Assessment: Putting science into the assessment of research." Accessed 24.03.2016. http://www.ascb.org/files/SFDeclarationFINAL.pdf.

Dupps, W. J., Jr., and J. B. Randleman. 2012. "The perils of the least publishable unit." *Journal of Cataract & Refractive Surgery* 38 (9):1517-1518. doi: 10.1016/j.jcrs.2012.07.020.

Espeland, W. N., and M. Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113 (1):1-40. doi: 10.1086/517897.

Felt, U. 2017. "Under the Shadow of Time: Where Indicators and Academic Values Meet." *Engaging Science, Technology, and Society* 3: 53-63. DOI:10.17351/ests2016.109.

Felt, U., D. Barben, A. Irwin, P.-B. Joly, A. Rip, A. Stirling, and T. Stockelova. 2013. *Science in Society: caring for our futures in turbulent times*, *ESF policy brief*. Strasbourg: European Science Foundation.

Felt, U., J. Igelsböck, A. Schikowitz, and T. Völker. 2013. "Growing into what? The (un-) disciplined socialisation of early stage researchers in transdisciplinary research." *Higher Education* 65 (4):511-524. doi: 10.1007/s10734-012-9560-1.

Fochler, M. 2016. "Variants of Epistemic Capitalism: Knowledge Production and the Accumulation of Worth in Commercial Biotechnology and the Academic Life Sciences." *Science, Technology & Human Values* 41 (5):922-948. doi: 10.1177/0162243916652224.

Fochler, M., U. Felt, and R. Müller. 2016. "Unsustainable Growth, Hyper-Competition, and Worth in Life Science Research: Narrowing Evaluative Repertoires in Doctoral and Postdoctoral Scientists’ Work and Lives." *Minerva* 54 (2):175-200. doi: 10.1007/s11024-016-9292-y.

Gill, R. 2010. "Breaking the silence: The hidden injuries of neo-liberal academia." In *Secrecy and silence in the research process: Feminist reflections*, edited by Róisín Ryan-Flood and Rosalind Gill, 228-244. London: Routledge.

Gill, R. 2014. "Academics, Cultural Workers and Critical Labour Studies." *Journal of Cultural Economy* 7 (1):12-30. doi: 10.1080/17530350.2013.861763.

Gläser, J., S. Lange, G. Laudel, and U. Schimank. 2010. "The Limits of Universality: How Field-Specific Epistemic Conditions Affect Authority Relations and their Consequences." In *Reconfiguring Knowledge Production. Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation*, edited by Richard Whitley, Jochen Gläser and Lars Engwall, 291-324. Oxford: Oxford University Press.

Hammarfelt, B., and S. de Rijcke. 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. doi: 10.1093/reseval/rvu029.

Hammarfelt, B., S. de Rijcke, and A. D. Rushforth. 2016. "Quantified Academic Selves: The Gamification of Science through Social Networking Services." *Information Research* 21 (2).

Haraway, D. J. 2016. *Staying with the trouble: making kin in the Chthulucene, Experimental futures: technological lives, scientific arts, anthropological voices*. Durham: Duke University Press.

Hazelkorn, E. 2015. *Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence*. London: Palgrave Macmillan UK.

Hicks, D., P. Wouters, L. Waltman, S. de Rijcke, and I. Rafols. 2015. "Bibliometrics: The Leiden manifesto for research metrics." *Nature* 520:429-431.

Hoffman, S. G. 2011. "The new tools of the science trade: contested knowledge production and the conceptual vocabularies of academic capitalism." *Social Anthropology* 19 (4):439-462. doi: 10.1111/j.1469-8676.2011.00180.x.

Irwin, A. 2017. "If the Indicator Game is the Answer, then What is the Question?" *Engaging Science, Technology, and Society* 3: 64-72. DOI:10.17351/ests2017.110.

Jacob, M., and V. L. Meek. 2013. "Scientific mobility and international research networks: trends and policy tools for promoting research excellence and capacity building." *Studies in Higher Education* 38 (3):331-344. doi: 10.1080/03075079.2013.773789.

Kaltenbrunner, W., and S. de Rijcke. 2016. "Quantifying 'Output' for Evaluation: Administrative Knowledge Politics and Changing Epistemic Cultures in Dutch Law Faculties." *Science and Public Policy*. doi: 10.1093/scipol/scw064.

Knights, D., and C. A. Clarke. 2014. "It's a Bittersweet Symphony, this Life: Fragile Academic Selves and Insecure Identities at Work." *Organization Studies* 35 (3):335-357. doi: 10.1177/0170840613508396.

Laudel, G., and J. Gläser. 2006. "Tensions between evaluations and communication practices." *Journal of Higher Education Policy and Management* 28 (3):289-295. doi: 10.1080/13600800600980130.

Linková, M., and T. Stöckelová. 2012. "Public accountability and the politicization of science: The peculiar journey of Czech research assessment." *Science and Public Policy* 39 (5):618-629. doi: 10.1093/scipol/scs039.

Martin, B. R. 2013. "Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment." *Research Policy* 42 (5):1005-1014. doi: http://dx.doi.org/10.1016/j.respol.2013.03.011.

Martin, B. R. 2016. "Editors' JIF-boosting stratagems – Which are appropriate and which not?" *Research Policy* 45 (1):1-7. doi: http://dx.doi.org/10.1016/j.respol.2015.09.001.

McHardy, J. 2017. "Like Cream: Valuing the Invaluable." *Engaging Science, Technology, and Society* 3: 73-83. DOI:10.17351/ests2017.116.

Mills, D., and R. Ratcliffe. 2012. "After method? Ethnography in the knowledge economy." *Qualitative Research* 12 (2):147-164. doi: 10.1177/1468794111420902.

Müller, R. 2014. "Postdoctoral Life Scientists and Supervision Work in the Contemporary University: A Case Study of Changes in the Cultural Norms of Science." *Minerva* 52 (3):329-349. doi: 10.1007/s11024-014-9257-y.

Müller, R. 2017. "Crafting a Career in STS: Meaning Making, Assessment, and Interdisciplinary Engagement." *Engaging Science, Technology, and Society* 3: 84-91. DOI:10.17351/ests2017.112.

Müller, R., and S. de Rijcke. n.d. "Thinking with Indicators. Exploring the Epistemic Impacts of Academic Performance Indicators in the Life Sciences." *Submitted to Research Evaluation*.

Musselin, C. 2014. "Research issues and institutional prospects for higher education studies." *Studies in Higher Education* 39 (8):1369-1380. doi: 10.1080/03075079.2014.950449.

Neyland, D., S. Milyaeva, and V. Ehrenstein. 2015. "Market-Based Initiatives as Solutions to Techno-Scientific Problems: MISTS." *EASST Review* 34 (3):9-12.

Owen, W. 2004. "First person: in defense of the least publishable unit." *The Chronicle of Higher Education* 9.

Paradeise, C., and J.-C. Thoenig. 2013. "Academic Institutions in Search of Quality: Local Orders and Global Standards." *Organization Studies* 34 (2):189-218. doi: 10.1177/0170840612473550.

Power, M. 1997. *The audit society : rituals of verification*. Oxford, England ; New York: Oxford University Press.

Rabinow, P. 2011. *The accompaniment : assembling the contemporary*. Chicago: University of Chicago Press.

de Rijcke, S., I. Wallenburg, R. Bal and P. Wouters. 2016. "Comparing Comparisons. On Rankings and Accounting in Hospitals and Universities." In *Practising Comparison: Logics, Relations, Collaborations*, edited by Joe Deville, Michael Guggenheim and Zuzana Hrdličková. Manchester: Mattering Press.

de Rijcke, S., P. F. Wouters, A. D. Rushforth, T. P. Franssen, and B. Hammarfelt. 2016. "Evaluation practices and effects of indicator use—a literature review." *Research Evaluation* 25 (2):161-169. doi: 10.1093/reseval/rvv038.

Rushforth, A., and S. de Rijcke. 2015. "Accounting for Impact? The Journal Impact Factor and the Making of Biomedical Research in the Netherlands." *Minerva* 53 (2):117-139. doi: 10.1007/s11024-015-9274-5.

Sauder, M., and W. N. Espeland. 2009. "The Discipline of Rankings: Tight Coupling and Organizational Change." *American Sociological Review* 74 (1):63-82. doi: 10.1177/000312240907400104.

Shore, C. 2008. "Audit culture and Illiberal governance." *Anthropological Theory* 8 (3):278-298. doi: 10.1177/1463499608093815.

Sigl, L. 2015. "On the Tacit Governance of Research by Uncertainty: How Early Stage Researchers Contribute to the Governance of Life Science Research." *Science, Technology & Human Values* 41 (3):347-374. doi: 10.1177/0162243915599069.

Slaughter, S., and G. Rhoades. 2004. *Academic capitalism and the new economy. markets, state, and higher education*. Baltimore: Johns Hopkins University Press.

Sparkes, A. C. 2007. "Embodiment, academics, and the audit culture: a story seeking consideration." *Qualitative Research* 7 (4):521-550. doi: 10.1177/1468794107082306.

Stephan, P. E. 2012. *How Economics Shapes Science*. Cambridge, Mass.: Harvard University Press.

Strathern, M. 1997. "'Improving ratings': audit in the British University system." *European Review* 5 (03):305-321. doi: doi:10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4.

Vann, K. 2017. "Surplus and Indicator." *Engaging Science, Technology, and Society* 3: 92-107. DOI:10.17351/ests2016.113.

Weingart, P. 2005. "Impact of bibliometrics upon the science system: Inadvertent consequences?" *Scientometrics* 62 (1):117-131. doi: 10.1007/s11192-005-0007-7.

Whitley, R., and J. Gläser. 2007. *The changing governance of the sciences : the advent of research evaluation systems*, *Sociology of the sciences yearbook*. Dordrecht, the Netherlands: Springer.

Wilsdon, J., L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, R. Kain, S. Kerridge, M. Thelwall, J. Tinkler, I. Viney, P. Wouters, J. Hill, and B. Johnson. 2015. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*.

Wouters, P. 2014. "The Citation: From Culture to Infrastructure." In *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, edited by B. Cronin and C. R. Sugimoto, 47-66. Cambridge: Mit Press.

Wouters, P. 2017 "Bridging the Evaluation Gap." *Engaging Science, Technology, and Society* 3: 108-118. DOI:10.17351/ests2017.115.

Ylijoki, O.-H. 2010. "Future orientations in episodic labour: Short-term academics as a case in point." *Time & Society* 19 (3):365-386. doi: 10.1177/0961463x10356220.

Zuiderent-Jerak, T. 2015. *Situated intervention : sociological experiments in health care*, *Inside technology*. Cambridge, Massachusetts: The MIT Press.

Zuiderent-Jerak, T., and C. Bruun Jensen. 2007. "Editorial Introduction: Unpacking 'Intervention' in Science and Technology Studies." *Science as Culture* 16 (3):227-235. doi: 10.1080/09505430701568552.