

Valuation regimes in academia: Researchers' attitudes towards their diversity of activities and academic performance

Nicolas Robinson-García^{1,2*}, Rodrigo Costas^{3,4}, Gabriela F. Nane² and Thed N. van Leeuwen³

¹EC3 Research Group, Departamento de Información y Comunicación, Universidad de Granada, Granada, Spain

²Delft Institute of Applied Mathematics (DIAM), TU Delft, Delft, Netherlands

³Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, Netherlands

⁴Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch University, Stellenbosch, South Africa

* Corresponding author: elrobin@ugr.es

Abstract

Evaluation systems have been long criticised for abusing and misusing bibliometric indicators. This has created a culture by which academics are constantly exposing their daily work to the standards they are expected to perform. In this study we investigate whether researchers' own values and expectations are in line with the expectations of the evaluation system. We conduct a multiple case-study of five departments in two Dutch universities to examine how they balance between their own valuation regimes and the evaluation schemes. For this we combine curriculum analysis with a series of semi-structured interviews. We propose a model to study diversity of academic activities and apply it to the multiple-case study to understand how such diversity is shaped by discipline and career stage. We conclude that the observed misalignment is not only resulting from an abuse of metrics, but also by a lack of tools to evaluate performance in a contextualised and adaptable way.

Keywords: evaluation culture; research evaluation; scientometrics; research careers; academia

Introduction

Bibliometric indicators were incorporated in research evaluation exercises already in the early 1970s (De Bellis 2014). Based on the notion of citation and productivity as predictors of recognition and academic success (Merton 1968; Reskin 1977), their use soon expanded from the United States to Europe in the 1980s (Luukkonen 2002), and to the rest of the world. Citation- and publication-based indicators were seen as objective and measurable ways to assess academic performativity and impact. Merton's normative framework of recognition served as a means to justify the use of citations as proxies for research quality (Merton 1968; Merton 1973), despite his ideas being seriously criticized by sociologists and substituted by a constructivist framework (Latour and Woolgar 1979; Knorr-Cetina and Mulkay 1983; Luukkonen 1997).

Bibliometric indicators have become a *de facto* standard in evaluation practices, being extensively used in academic hiring and promotion, increasing the role scientific publishing plays in determining academic success and career advancement (Schimanski and Alperin 2018; Alperin et al. 2019). Many have claimed that their influence has become excessive (de Rijcke et al. 2016; Hammarfelt and Rushforth 2017; McKiernan et al. 2019), leading to a situation in which

“we risk damaging the system with the very tools designed to improve it” (Hicks et al. 2015, 429).

The goal of this paper is to understand how academics deal with the wide array of tasks they are confronted with in an evaluative context which prioritises certain activities over others. We hypothesize that, explicitly or implicitly, bibliometric indicators influence the decision-making process by introducing an *availability bias* which reduces the role played by less visible activities. That is, “people assess the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind” (Tversky and Kahneman 1974, 1127). In the context of research evaluation, bibliometric indicators have been held responsible for pushing researchers to focus exclusively on scientific publishing and leaving aside other tasks which may be as important for academia (de Rijcke et al. 2016).

We use the concept of *valuation regimes* (Fochler et al. 2016; Bigger and Robertson 2017) to understand the mechanisms by which researchers embrace diverse activities while confronting such evaluative pressure that pushes them away from this diversity. By employing a multiple case study research design, we profile five research departments based on the diversity of academic activities in which their members are involved. We combine data from publication records, curricula, online presence, and personal interviews to explore how scientists value and perceive others’ valuation of such diversity and the mechanisms employed to navigate an evaluative context that constrains it. Building on our findings, we propose a new research agenda to advance on the development of quantitative methods that can improve current academic evaluation systems in promotion and hiring activities.

This paper is structured as follows. We start by briefly reviewing the most recent literature discussing effects of (mis)use of bibliometric indicators in research careers. Criticisms have either focused on the abuse of certain indicators (e.g., Journal Impact Factor, H-Index) or on recent initiatives to change the current academic evaluation culture. Second, we present a conceptual framework in which we build from the notions of *diversity and valuation regimes* as key concepts from which we propose a *valuation model* to study diversity of academic activities. Third, we describe the design of a multiple case study approach. Finally, we present and discuss the results of our study and conclude with some remarks on the main contributions of the paper.

Literature review

The abuse of metrics in the assessment of academic careers

In the past few years many studies have been devoted to the effects of the misuse of metrics in research assessment in academia (cf., Dalen and Henkens 2012; Fochler et al. 2016; de Rijcke et al. 2016; Cañibano et al. 2018). These studies can be organised into three main groups: 1) those related to changes in academics’ behaviour and on their preferences, 2) those related to disciplinary and geographic inequalities arising from the technical and conceptual limitations of the metrics used, and 3) those derived from the overexposure of academics to an increasingly competitive environment.

Studies focused on academics’ behaviour warn of a reduction of tasks as a consequence of an overemphasis on publications (cf., de Rijcke et al. 2016). As goals are placed on research output, scientists shift their focus from research contents and quality to journal venue and quantity. Publishable topics potentially attracting a high number of citations are prioritised (Moher et al. 2016; McKiernan et al. 2019), modifying also research agendas to better accommodate to evaluation agencies’ expectations (Cañibano et al. 2018). Priority changes do not affect all

researchers evenly, with men being more prone to adopt profiles which are better rewarded in terms of citation impact (Robinson-Garcia et al. 2020; Zhang et al. 2021).

Academic freedom is mediated by the expectations the community has based on what it considers as scientific merit (Polanyi 1962). As evaluation schemes shape such expectations, scientists' goals may deviate from what is considered relevant, to what pays off best. By using journal rankings or rewarding high citation rates, evaluation schemes pressure scientists to adapt their strategies to maximize their efforts (Cañibano et al. 2018). For instance, an emphasis on novelty is a contributing factor for the lack of replication studies, under-reporting of false positives or testing of prior published results (Nosek et al. 2012). As bibliometric indicators are mostly focused on scientific impact, they reinforce the sense of isolation of scientists in their 'ivory tower', alien from societal challenges and demands (Sarewitz and Pielke Jr 2007; Sarewitz 2016a).

The negative effects of the abuse of metrics in research evaluation varies across fields and regions. Local and non-mainstream journals in peripheral countries are relegated, despite the role they play in national science systems, bridging between local communities and international scientific networks (Chavarro et al. 2017). Policies promoting publications in high impact journals penalise locally-oriented (López Piñero and Hicks 2015) and interdisciplinary research (Rafols et al. 2012). This is due to the well-known biases mainstream bibliometric databases have in regard to language (van Leeuwen et al. 2001), disciplinary coverage (Archambault et al. 2006) and peripheral or regional knowledge (Rafols et al. 2015). A non-discriminatory use of metrics in different contexts undermines the performance of scientists in fields and regions which do not adjust to the expectations in which bibliometric indicators are built (Gläser and Laudel 2007; Ràfols et al. 2016; Robinson-Garcia and Ràfols 2020).

Finally, scientists working in highly competitive environments perceive an external pressure to publish, which makes them pay excessive attention to "tasks that benefit or glorify the individual (publishing internationally, being cited by other scholars)" rather than those "tasks that benefit larger groups" (Dalen and Henkens 2012, 1291). The notion of performance becomes entangled with the indicators used to measure it. A suspicion comes then from anything that deviates from the expected performance, not being able to interpret correctly those cases, as "there may be favoritism in some competitions and smart hiring in others" (Abramo et al. 2015, 772).

Aligning scientists' motivations and evaluation criteria

Scientists' motivations have been extensively studied in the field of sociology of science. From Merton's normative theory (Merton 1973) or Latour's credibility cycle (Latour and Woolgar 1979) to more recent frameworks on motivations for research commercialisation (Lam 2011) or collaborations with non-academic partners (Ramos-Vielba et al. 2016). Recent studies suggest increasing attrition in academia, which is especially felt among early career researchers (Fochler et al. 2016; van der Weijden et al. 2016; Waaijer et al. 2018; Woolston 2020). The cause of attrition is both the misuse of bibliometric indicators and the misalignment between what researchers perceive as important and what evaluation schemes prioritise (Bonn and Pinxten 2021). More exasperating is the fact that researchers feel that "despite increasing criticism, alternative assessments are difficult to find" (Bonn and Pinxten 2020, 4).

Such misalignment damages the credibility of the whole scientific system (Ioannidis 2005). Researchers are confronted with a conflict of interest, as they are pushed to prioritise publication over accuracy and transparency (Nosek et al. 2015). A claim shared by many scientists (Baker 2016), who point at selective reporting and pressure to publish as the two most

common factors contributing to the lack of reproducibility in science. While these claims are still under scrutiny (Fanelli 2018; Fanelli et al. 2019), it seems evident that the incentive structure in the scientific system disrupts the way science is conducted (Stürmer et al. 2017) and threatens the quality and reliability of the knowledge produced (Bouter et al. 2016; Sarewitz 2016b).

New initiatives to reform evaluative schemes

The Declaration of San Francisco on Research Assessment (DORA)¹ seems to have stirred action from institutions, funding agencies and researchers. DORA is the result of a meeting held by a group of scientific editors, publishers and journals. The declaration was a call for action directed to researchers, funders, publishers, institutions and bibliometricians to abandon the use of the Journal Impact Factor in promotion, hiring or funding decisions. Although the recommendations and warnings included in the document had been raised before, DORA had a cascading effect in the scientific system. A year after its publication, a group of Dutch researchers published a position paper calling for a reform of the science system (Dijstelbloem et al. 2013). The paper was the result of a series of workshops and meetings in which participants reflected on the health of the scientific system. In 2015, a group of bibliometricians published the Leiden Manifesto, which calls for the promotion of good practices in the use of research metrics (Hicks et al. 2015). More recently, the Hong Kong Manifesto was published, calling for deeper changes in the evaluative culture of the scientific system (Moher et al. 2020). These initiatives added to an already existing stream of literature proposing alternative ways in which individual researchers could be assessed (i.e., ACUMEN 2014; Ioannidis and Khoury 2014).

The results of such actions are still under-going, although some reactions can already be observed. For instance, LIS-Bibliometrics, a forum for professionals working with or interested on the use of bibliometrics based mainly in the United Kingdom, launched an annual survey in which professionals are asked about the implementation of responsible metrics policies at their institutions (Gadd 2015). Institutions have also taken notice and many now issue statements regarding the way metrics are implemented in hiring and promotion decisions. An example of such institutional action is the Responsible Metrics Policy from the University of Loughborough², which sets 10 statements on how metrics must be used. Another example is the Imperial College London which issues a statement in which it commits to abide by the recommendations set by DORA³.

More radical is the reform currently being undertaken in The Netherlands, which has recently presented its Strategy Evaluation Protocol (SEP) for 2021-2027⁴. The protocol explicitly forbids the use of the Journal Impact Factor and discusses the need to assess other aspects of the academic activity such as societal relevance or viability. Although SEP refers specifically to research units (i.e., institutes, departments), its consequences are expected to be felt at all levels. This has led to some controversy as to the role metrics should play in hiring and promotion decisions (Chawla 2021). Two positions seem to emerge, on the one hand, those advocating for more qualitative forms of evaluation by introducing narrative CVs, and on the other hand those arguing that metrics should still have room as measures for benchmarking and comparison, incl. JIF and h-index.

¹ <https://sfdora.org/read/>

² <https://www.lboro.ac.uk/research/support/publishing/responsible-use-of-metrics/>

³ <https://www.imperial.ac.uk/research-and-innovation/about-imperial-research/research-evaluation/>

⁴ https://www.nwo.nl/sites/nwo/files/documents/SEP_2021-2027.pdf

Those in favour of qualitative forms of evaluation argue that it would introduce broader perspectives on academic performance which encourage greater diversity of profiles and a healthier and more sustainable research environment. An example of such are Utrecht and Tilburg universities. Both have introduced new models for evaluation inspired by the *Room for Everyone's Talent* position paper (VSNU and ZonMw 2019) in which domains of academic work such as education, teamwork, professional performance or managerial tasks are also assessed. Those raising concerns against these new approaches argue that the removal of metrics would introduce arbitrariness, would reduce transparency in evaluation exercises, and would discourage researchers and reviewers who would have no clear idea of the goals of the assessment and the expectations to be met (Chawla, 2021).

Conceptual framework

These debates reflect the need to rethink assessment and realign incentives, values and expectations of the science system and of all its members in order to redefine the conditions under which academics work. While we observe some consensus on the criticisms made, the alternative does not seem to convince everyone. We observe that the criticism to metrics goes beyond technical limitations but has to do with the fact that these metrics are focused on one single type of activity: academic publishing.

To better understand how diverse academic activities are, and how they are perceived and valued by researchers, we explore the concepts of diversity and valuation regimes in science. First, we introduce *valuation regimes* to clarify what we refer to by value, which is different from the concept of public value or social value of research (Bozeman and Sarewitz 2011). Second, we review how diversity has been addressed when studying science, especially within the fields of research evaluation and research policy. Finally, we propose a *valuation model* to understand how diversity of activities is perceived in academia.

Valuation regimes in academia

The worth attributed to each of the academic activities undertaken will depend on researchers, institutions and funding agencies' valuation systems. Hence, priority or worth will be given to certain activities over others. To do so, the academic system must go through a critical examination of its goals and objectives. This must be done at all levels and in a transparent way to find some agreement between valuation regimes. These are defined as social rules or modes of comparison on what worth is. In the academic context, worth can be defined as the set of principles by which researchers abide and conduct their academic work (Fochler et al. 2016). Valuation regimes can take the form of assumed principles as well as imperative rules and can help explain behaviours and attitudes towards academia.

When these valuation regimes differ between groups in academia, frustration and controversy may result (van der Weijden et al. 2016; Curry 2018; Curry et al. 2020; Chawla 2021). To confront differing valuation regimes, researchers may use different strategies, such as diversifying their portfolio of activities (Rushforth et al. 2019) or dealing only with those activities and goals which minimize the risk of failure (Fochler and Sigl 2018).

By framing research evaluation of individuals in terms of evaluation regimes, we can approach the issue as a matter of priorities between what academics considered worthy and what is indicated as valuable in evaluations. In this way it is possible to understand how misalignments are produced and contribute to discussions on how to assess individuals' performance shifting the focus from which criteria should be used to which principles should be guiding such criteria.

Diversity in academic work

Diversity in science, understood as “an attribute of any system whose elements may be apportioned into categories” (Stirling 2007, 708), has been studied from many lenses. It has been used to define and measure interdisciplinarity (Gómez et al. 1996; Zitt 2005; Rafols and Meyer 2009; Zhou et al. 2012), to identify novelty in science (Wagner et al. 2019), or to distinguish forms of collaboration in terms of team dynamics (Bone et al. 2019) and team formation (Freeman and Huang 2014; Murray et al. 2019; Ding et al. 2021). Here, we approach diversity based on Bourdieu’s belief about fundamental differences in the ways in which scholars conduct science (Bourdieu 1975). We refer to diversity in science in terms of the breadth of activities and outputs academics produce. We do not circumscribe our study solely to scientific tasks devoted to the production of new knowledge (Latour and Woolgar 1979), but refer to academia in its broadest sense, as a community of scholars embarked in the role of teaching new generations, making new discoveries in science and translating scientific knowledge into public value (Bozeman and Sarewitz 2011).

Academics conduct a breadth of activities either in a distributed manner or in isolation. They adopt different roles based on the activities they perform. These roles shape the academic profile that will define them during their career trajectory. The roles will also change as they acquire experience and seniority. Some of these activities involve coordinating, designing and leading research agendas or curricula; producing new tools and software for data collection and processing; building new conceptual and theoretical frameworks; or engaging with non-academic stakeholders and bridge with societal demands, among others.

But evaluative processes struggle with the notion of diversity at the individual level (Walsh et al. 2019). The introduction of New Public Management practices in academia have promoted the use of quantitative measures to ensure “objectivity and transparency” in decision-making processes. Bibliometric indicators offer such measures which can be applied universally by placing responsibility on numbers (Porter 1996). By placing an inexistent relation between the different missions of university (Sánchez-Barrioluengo 2014), a representativeness bias (Kahneman and Tversky 1972) is introduced by which academics excelling in research performance will excel in all facets of the academic work.

Secondly, contemporary science deals with many additional activities derived from the profound transformation modern societies have undergone in the 21st century (e.g., social media and fake news, economic crisis, COVID-19 pandemic). As a result, researchers are expected to respond to societal demands (Frodeman and Parker 2009; Holbrook 2012), involve citizens in research processes (Irwin 1995; Trencher et al. 2014), and be transparent and open when conducting, reporting and disseminating research (Nosek et al. 2015). These new expectations involve a reconfiguration on the criteria applied in evaluative processes to accommodate research processes as well as impacts (Spaapen and Drooge 2011). The societal shift to the online environment and the increase in computational power have also given birth to a plethora of sources through which tasks beyond scientific publishing are now visible. They provide a unique opportunity to further advance our understanding of how science is socially constructed.

Diversity of academic activities has been explored mostly in the context of team science and task distribution. Seminal works in this regard are the studies conducted by Latour and Woolgar (1976) or Richard Whitley (2000) on researchers’ internal organization, introducing the notions of task uncertainty and mutual dependence. More recently, quantitative studies empirically looked into diversity of activities based on contribution statements included in publications

(Larivière et al. 2016; Larivière et al. 2020; Robinson-Garcia et al. 2020). These studies, as well as others (cf. Jackson et al. 2003) mention the direct and indirect influence personal attributes may have on the types of activities individuals undertake. However, they are still limited to research publishing activities.

A broader framework is introduced by the Science & Technology Human Capital model (Bozeman et al. 2001). Although it does not address activities, it discusses researchers' cognitive, science and technological skills. From an evaluative stand, diversity of activities has also been addressed, although most of the work has been limited to proposals which still remain to be tested. For instance, the ACUMEN project (2014) undertaken by a consortium of European research institutions elaborated a CV format designed for assessing individual performance. It combines qualitative and quantitative information offering space for introducing narratives. It distinguishes between three aspects of an academic career: 1) expertise (methods, areas of theory, etc.), 2) outputs (publications, patents, etc.) and 3) impacts (citations, awards, etc.). These three domains are expected to cover all types of activities academics may undertake. More recently, Utrecht University presented the TRIPLE model which stands for team spirit, research, impact, professional performance, leadership and education. These six components seek to comprise all academic activities surrounding research, education and professional performance.

A valuation model of academic activities

Here we aim to understand how academics deal with diversity of activities in an evaluative context in which scientific publishing is prioritized over the rest of activities. To do so, we propose a valuation model by which we can categorise activities and profile academics building on our conceptual framework. The model differs from evaluation models such as the ones mentioned above as much as it does not consider the success or impact of each of the activities.

The model consists of three components. The central component is what we call *evaluative dimensions* of an academic's performance. That is, the set of activities which are considered when assessing an academic's CV. Namely, researchers' trajectory or background, scientific work, social engagement, teaching, and capacity to attract resources. The evaluative dimensions are detailed in the next section. The other two components refer to attributes and factors which may directly or indirectly affect performance or trigger biases in the assessment process. These are the academic's *personal features* and the *external factors* shaping their work environment. Figure 1 provides a complete overview of the model. Next, we describe each of the components.

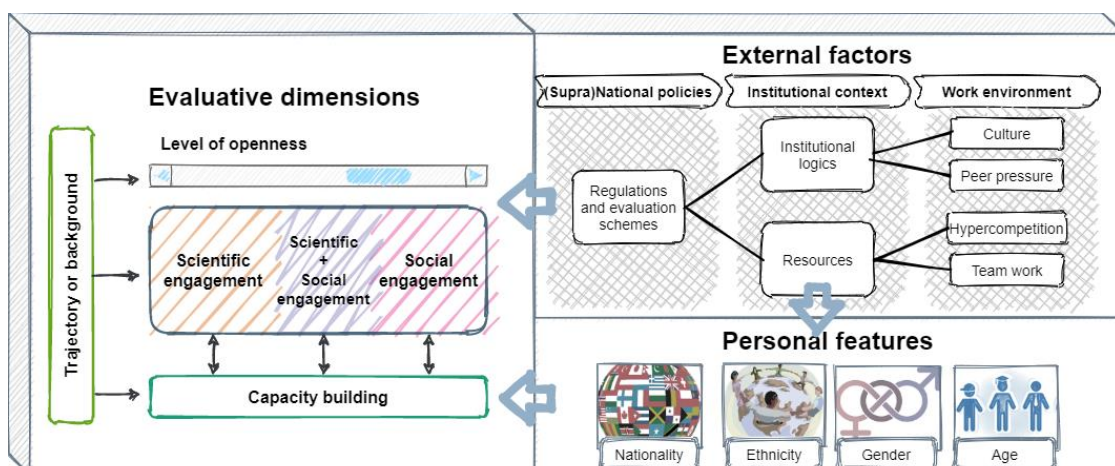


Figure 1. Overview of the proposed valuation model for studying individuals' profiles based on their diversity of academic activities

Evaluative dimensions

We define five dimensions of activity. The purpose of each dimension is to group activities in a tractable way that can help later discern what and how academics value their achievements. In the following we describe them.

Scientific engagement. Defined as those activities related with conducting research and engaging in science communication. Activities in this group would include scientific publishing, editorial activities, or organizing and participating in congress and seminars. This goes beyond what is usually considered in evaluation schemes (namely publications and their citation impact), as the dimension includes tasks such as reviewing manuscripts or organising scientific events which are normally invisible in evaluation schemes.

Social engagement. It refers to those activities in which there is a clear intention to interact with or to benefit society. This is different from the concept of *productive interactions*, which is broader and is defined as “exchanges between researchers and stakeholders in which knowledge is produced and valued that is both scientifically robust and socially relevant” (Spaapen and Drooge 2011, 212). Such distinction allows to distinguish, between e.g., scientific publications as scientific engagement from popular science as social engagement. In cases where such distinction is not possible, - e.g., publications from the Humanities fields directed at multiple academic and non-academic audiences (Nederhof 2006), - these activities are included in both dimensions. Other examples of activities of social engagement are social outreach activities, consultancy, social dissemination of research outcomes or knowledge transfer activities.

Trajectory. Understood as those characteristics of an individual's experience which can be of value for their potential career development. Although this dimension does not include activities *per se*, it is an aspect of consideration in evaluation schemes, as “individuals' careers constitute an important resource for evaluation of their claims” (Latour and Woolgar 1986, 172). Non-academic experience, international experience, mobility or training may constitute important elements when assessing an individual's potential for future performance, especially in earlier stages of academic career, when other achievements may be still lacking. It also gives room to specific factors which may be affecting individuals' trajectories, such as institutional reputation (Morgan et al. 2018; Nielsen et al. 2021).

Capacity building. Defined as the process by which individuals acquire and retain resources and skills in order to advance in their career. We build from the concept of *science and technology human capital* defined by Bozeman and colleagues (2001). Capacity building refers to individuals' ability to succeed on gaining funding and building infrastructure in order to advance on their *community career* (Laudel and Glaser 2011) and expand their social capital.

Level of openness. Understood as the integration of open scholarship practices within academics' daily routines, ensuring transparency and reusability of their outputs. Open practices do not refer to specific activities but on how these are conducted. Hence, it is a transversal dimension which affects academics' social and scientific engagements. For instance, open peer review or open access to publications would reflect open scientific engagement, while public and universal access to teaching contents, use of social media or publishing open software and code would reflect and open social engagement. On the contrary, commercialisation of products or patenting would reflect closed social engagement.

Each dimension encompasses a variety of activities academics may conduct, value or experience differently. As observed in Figure 1, each dimension is related to the rest differently. Social and scientific engagement may overlap when conducting activities which are aligned both with academics and non-academics. Trajectory affects how academics engage as well as their levels of openness. Capacity building has a circular relation with scientific and social engagement. Academics conduct activities by which they accrue *credit*, this credit allows them to build capacity and consequently, undertake more ambitious activities. Here we understand *credit* as defined by Latour and Woolgar (1986), who move away from the notion of symbolic capital (Bourdieu 1986) by acknowledging that credit will not always translate into reward. Individuals may accumulate credit, - e.g., in the form of citations to their publications or reputation within their former PhD students, - which they may use to attract funding, new PhD students and postdocs, build new training programmes, etc. In the same vein, there may be activities which, despite being of importance, may not facilitate capacity building (e.g., teaching activities).

Personal features

This component refers to individual characteristics such as gender, nationality, ethnicity, or age among others. These are features which, in principle, would not be considered for an assessment, but may bias evaluations or may respond to policies introduced to correct such biases.

For instance, in the case of gender, there is compelling evidence of barriers against women in science worldwide (Larivière et al. 2013). Such barriers do not only affect women's research performance, but all areas of their academic work, being more associated with administrative work (Heijstra et al. 2017), technical roles (Macaluso et al. 2016; Robinson-Garcia et al. 2020) or suffering a greater cost from their work-life balance when considering temporal mobility stays (Cañibano et al. 2016). Gender barriers are especially evident in longitudinal analyses on career trajectories, where a higher rate of dropouts is observed for women than for men (Lerchenmueller and Sorenson 2018; Huang et al. 2020). Life balance and parenting affect scientists, and especially women scientists' decision and opportunity to perform (Azoulay et al. 2016; Derrick et al. 2019), with recent evidence suggesting that the gap is widening (Viglione 2020).

In the case of nationality, social scientists in the United States have devoted a large proportion of studies on the analysis of foreign-born scientists' productivity (Levin and Stephan 1999; Stephan and Levin 2001), mobility (Rumbley et al. 2012), or job satisfaction (Lin et al. 2009; Mamiseishvili and Rosser 2010) among others. Understanding cultural barriers, integration issues or push-pull factors for attracting foreign scientists is key to maintaining an internationalised scientific system.

Age is another feature which has a clear effect on academic performance. It is normally accounted in evaluation schemes when promoting younger generations of scholars or identifying productivity peaks (Levin and Stephan 1991; Costas et al. 2010). This ageing effect is especially relevant at the institutional level, where policy measures on hiring are crucial on the maintenance of a dynamic scientific workforce (Bonaccorsi and Daraio 2003).

Finally, ethnicity has been found to be a feature related with collaboration; those from the same ethnicities will be more likely to work together (Freeman and Huang 2014). Ethnicity has been used also as a proxy to identify academics with a common cultural background (Karaulova et al. 2019).

External factors

External factors refer to confounders external to academics' range of action which may lead to misinterpretations in evaluative procedures, especially in terms of comparability (Vandenbroucke 2002). Three levels of external factors can be determined, 1) a (supra)national, 2) an institutional and 3) a departmental or unit level. In the upper-right side of Figure 1 the main factors affecting academics' performance are listed for each level. These reflect a hierarchical structure of interconnected entities which can jointly shape academics' decisions about performed activities.

To illustrate how external factors may affect the types of activities academics undertake we refer to the extensive literature discussing intentional or unintentional effects of national policies on publishing practices. These studies aim at establishing causal links between the introduction of policies and effects on the performance of an institution or country. An example is that by which an increase on the production of Spanish authored papers was attributed to the introduction of the *sexenios* in Spain (Jiménez-Contreras et al. 2003), - a policy by which academics are assessed every six years based on their five most important contributions (see Osuna et al. 2010 for a refutation of the study). But it has also been criticised for modifying and damaging the type of research conducted (López Piñero and Hicks 2015; Ràfols et al. 2016; Cañibano et al. 2018). Another notable example is that of the Australian evaluation system which was criticised for quantitatively incentivising an increase on the number of publications (Butler 2003). A controversial case which years later was still under debate (van den Besselaar et al. 2017; Waltman 2017).

Additionally, factors such as the internal distribution of tasks in research teams (Latour and Woolgar 1979; Larivière et al. 2016), lack of resources or peer pressure may also have confounding effects on academics' performance.

Research design

We performed a multiple-case study analysis to investigate the diversity of activities researchers conduct. For this purpose, five research units were selected from two Dutch universities and three fields of research. The two universities selected reflect very different institutional structures and missions, one being a classical university with a strong focus on the classical activities, such as the natural sciences, biomedicine and the social sciences, law and humanities, while the other one is a university of technology with a greater focus on engineering, technology and innovation. Table 1 includes a brief description of the number of individuals by field and university under study.

Heads of each research unit were contacted and informed on the purpose of the study and all of them gave consent⁵. The selection of the five case studies took place in July 2019. For each unit we gathered from their website the complete list of their staff, along with their gender, and academic and organizational status when available.

Table 1. Basic information on the case studies selected

Unit	Field	No. of individuals
Physics A	Astrophysics and Astronomy	177

⁵ A sixth research unit was selected for the field of Physics but was later removed after several attempts at contacting them.

Biomedicine A	Medical Statistics	16
Biomedicine B	Biomaterials and Biomechanics	27
Social Sciences A	Governance and Public Affairs	29
Social Sciences B	Technology and Policy Management	121
Total		369

Two types of analysis were conducted. First, we performed a curriculum analysis (Cañibano and Bozeman 2009), enriched with data from online academic websites, social media platforms, and publication databases. The purpose of such analysis was to identify the diversity of tasks researchers performed. We focus on public data as we assume that researchers will make information public if they attach worth to it (Wildgaard et al. 2014). Hence, we acknowledge that the diversity of activities presented here will likely underrepresent the true breadth of tasks. The goal was twofold. First, to identify diversity in a systematic way. Second, to analyse differing valuation regimes. We then conducted a series of in-depth interview to 8 staff members of the selected research units. Next, we describe in detail the procedure followed.

Curriculum analysis

Between December 2019 and March 2020, we gathered public data for all staff members in our case studies. This was a total of 369 individuals, including supporting staff. We gathered data from two types of sources: scientific publication databases (namely Web of Science and Google Scholar Profiles) and online websites and social media platforms. Only data related to their academic and professional activity was gathered.

We retrieved individuals' publication history from Web of Science by linking it to the researcher identifier used by the Centre for Science and Technology Studies (CWTS) at Leiden University. This identifier results from applying an author name disambiguation algorithm based on rule-scoring (Caron and van Eck 2014). In the case of Google Scholar, profiles were identified manually and publication data was extracted using the Publish or Perish software (Harzing 2007).

We retrieved data from online websites and social media platforms by searching for external links in their institutional website. We then searched for each individual's name in the search engine DuckDuckGo to find out if they had a personal website or online CV which could be gathered. Table 2 includes the most common sources of information identified from which individual's information was retrieved.

Out of the 323 staff members for which a source of information was found, we were finally able to gather information for 307 individuals (83%). We were not able to retrieve any information for 32 PhD students, 26 supporting staff (i.e., secretariat, financial administrators, student affairs), two postdoctoral researchers, one lecturer and one guest researcher.

Table 2. Information sources used, and number of researchers identified per source. The total refers to individuals for which we gathered information from at least one of the sources.

Source	Type of data gathered	No. of individuals
Web of Science	Publication data	235
LinkedIn	Curriculum	160
Google Scholar	Publication data	155
Institutional Web	Curriculum	136
Personal web	Curriculum and social media accounts	102
Twitter	Social media presence	53

CV (pdf format)	Curriculum	52
ORCID	Curriculum	15
Wikipedia	Curriculum	11
Facebook	Social media presence	11
Total		323

Information on individuals was gathered in the following way. We scanned each source for activities reported by the individual and labelled them according to our valuation model. That is, we noted which dimension each activity belongs to. We also coded personal features when mentioned such as PhD year, age, marital status, parenthood or nationality. This information was later used for the selection of interviewees. Information regarding teaching experience was only noted when emphasised in the data sources, and mentoring was only included when advising PhD students or postdocs. Also, conference papers or invitations to scientific conferences were tagged as publishing activities. Each activity was assigned to one dimension exclusively except for scientific engagement and social engagement. Finally, we normalised the different activities into a total of 31 activities. These are listed in Table 3 along with the acronyms used to refer to them in the results section.

Table 3. List of activities identified grouped by dimension.

Dimension	Acronym	Academic activity	Example
CAPACITY BUILDING	FU	Funding distribution	Member of evaluation panels
	GR	Grants / Awards	Acquisition of funding
	ME	Mentoring	Supervision of PhD students
	OT	Other	Honorary doctorates
	PH	PhD committees	External examiner
	PM	Project management	Coordinator of research consortia
	TE	Teaching management	Committee member in master
	TL	Team leader	Head of research team
	UN	University management	Dean
SCIENTIFIC ENGAGEMENT	AW	Awards committees	Member of awards and prizes committee
	CO	Conference organization	Conference organiser
	ED	Editorial activities	Editor, reviewer, associate editor
	FI	Fieldwork	Observing experience
	OD	Open data/code	Code sharing, GitHub account
	SC	Scientific organizations	Member of academic society
	PU	Scientific publications	Articles, books, chapters
	SO	Software	Software developer
SCI. & SOCIAL ENGAGEMENT	OP	Open teaching	Design of MOOCs, sharing slides
	ST	Student organizations	Member of student committees
SOCIAL ENGAGEMENT	AS	Academic Service	Contract research, consultancy
	AR	Artwork	Animations, short stories
	MC	Media coverage	Interviews, mentions in media
	NM	Non academic membership	Member of local organisations
	OT	Other	Patenting
	PR	Private Sector	Board member of company
	PO	Public outreach	Intervention in parliament, Op-Ed
	PS	Public sector	Director of museum

Dimension	Acronym	Academic activity	Example
TRAJECTORY	SM	Social media	Blogging, Twitter, podcasting
	VO	Volunteering	NGO experience
	MO	Mobility	Working abroad, temporary visits
	NE	Non academic experience	Work experience beyond academia

Interview protocol

We interviewed a total of 8 staff members from the five research units between July 2020 and January 2021. These individuals were selected according to their academic status, gender, nationality and range of activities fulfilled. Participants were all informed of the purposes of the study and how their data would be treated, and all provided written consent. The aim was to cover extreme cases of diversity as well as more standard academic profiles. We also included an interview with an individual who no longer worked in academia, who at the moment of collecting data appeared was listed as a staff member in one of the five research units.

Interviews were semi-structured (Given 2008), with a set of questions related to the researcher's trajectory, international experience, and self-assessment. Explicit mentions were made to specific activities undertaken by researchers to understand their motivation to conduct them. Also, researchers were asked on their mentoring practices, as well as on their values and preferences in recruitment. Specifically, researchers were asked to reflect on the aspects they perceived others valued the most versus what they valued the most. Furthermore, they were asked to name the top three things they would look at in a CV when recruiting a PhD student, a postdoctoral researcher and a tenure tracker. After the formal interview finalised, an open discussion based on the answers provided would follow deepening in some of the aspects raised during the interview.

Interviews were conducted online and lasted between 45 and 60 minutes with at least two interviewers present in each of them. They were recorded and fully transcribed. Transcriptions were later analysed and coded. We conducted a qualitative content analysis (Gläser and Laudel 2013) to provide rich narratives on the motivations and values of researchers when conducting academic work, and recruiting and mentoring others. Within the paper we provide quotes to illustrate the results. Individual quotes are confidential in nature, therefore some of these quotes have been masked to ensure the anonymity and privacy of participants.

Results and discussion

Following we present our findings from both the curriculum analysis and the conducted interviews. These are grouped into 5 sub-sections dealing with different themes (scientific engagement, social engagement, scientific & social engagement, capacity building, and trajectory) which emerged from the interviews and the following discussions.

Diversity between and within research units

Figure 2 reports the number of individuals for which at least one activity in each of the evaluative dimensions of our model was identified overall and for each of the five research units under analysis. As observed, scientific engagement seems to be the most valued dimension, followed by academics' trajectory. Activities which reflect both social and scientific engagement are rare, while social engagement and capacity building activities are somewhat in between.

However, there are some notable differences between research units. For instance, while activities on scientific engagement were the most common (71% of academics), only 41.3% of the individuals in Social Sciences A reported at least one activity in this area. Similarly, roughly half of the academics included information on their trajectory, but these ranged from 69% of academics in Social Sciences B to 37% in Physics A.

When deepening into the activities undertaken in each dimension, we observe that scientific publishing is at the core of their scientific engagement activities (Figure 3). While this may surface an alignment on valuation regimes, the fact is that there are many reasons why researchers would highlight their publications. For instance, a researcher in the biomedical field talked about how they felt about the open software they produced. He responded indicating that their ambition was “scientific achievement. And I think those are documented in publications, so that would be also my personal ambition, and the software would be a means to an end. But in the research I do, software is indispensable”. An alternative view was offered by a researcher from Physics A, who referred to publications as “part of the job” and pointed towards other types of activities such as teaching or social outreach.

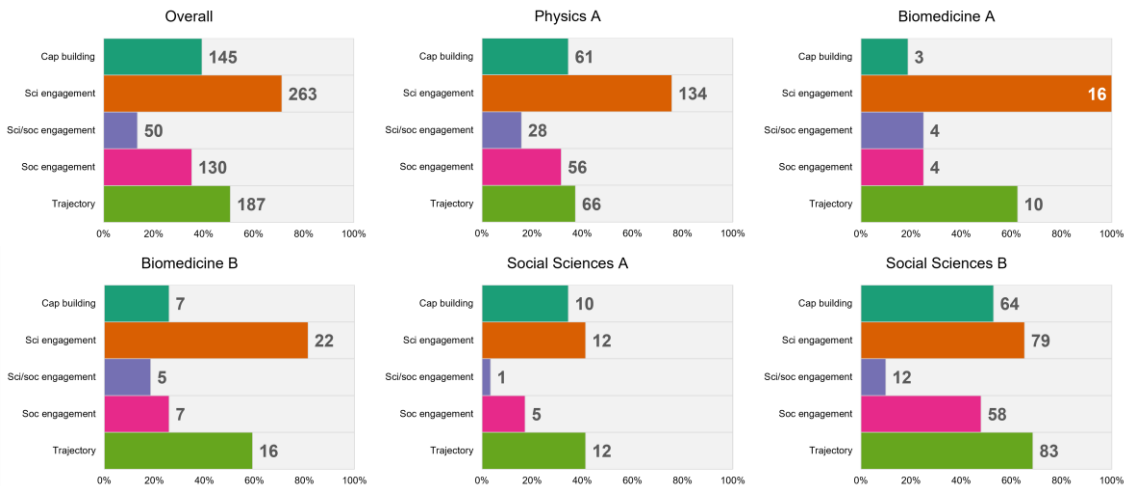


Figure 2. Number of individuals reporting at least one activity in each of the evaluative dimensions overall and by research unit.

We all say at all levels that teaching is important. But at the end of the day, officially we have something like a teaching career, but in practice we don't.

An assistant professor reflected on how teaching is normally assessed negatively, "it's like you have to teach these classes and if there is no problem then that's it. It doesn't matter if you do it in an excellent way or if you just do it like barely there." (Physics A).

This is not the case for mentoring (included in capacity building). Still, both were mentioned often by our interviewees as one of the activities they were most proud of. A professor from Biomedicine B reflected on this in the following terms:

Looking back at my career now, probably this changes over the years as you grow older... Probably, if you asked me the question 25 years ago, I would probably be [proud] mostly about research achievements and publications. But now I think the training of future generations every year, I consider that far more important. (Biomedicine B)

This is an aspect most of the interviewees agreed upon, regardless of their career stage. A professor from Social Sciences B indicated that what made them the proudest were "my PhD students who were able to pass their PhDs in a very successful way".

In the case of social outreach, we observed mixed opinions within our interviewees. Some considered as something for which "collectively you want to make significant outreach" (Physics A). In other cases, they find that their research can align well with societal demands, and it helps them shape their agenda:

The reason I am happy to do it is because it gives you insight into the research needed, societal relevance research and where we are heading at. That makes it also easier for you to come up with research proposals. (Social Sciences B)

While others consider it part of their duty as academics, regardless of how other value their efforts: "I am mostly doing outreach because I find it very relevant, but not everyone finds it very relevant" (Physics A).

Priorities change depending on the career stage

The types of activities academics employ seem to be aligned with their career stage, and hence, each research unit's profile is largely influenced by how its workforce is composed in terms of size and academic status. Eleven different figures were identified within each research unit, the most common being professor, associate professor, assistant professor, researcher, postdoctoral researcher and PhD student. Figure 4 offers a breakdown of individuals reporting at least one activity per dimension according to their academic status (showing the number of individuals behind each activity). Activities related to scientific engagement are again the most common, but there are some notable differences between profiles. While most professors (above 80%) report information related to their trajectory or capacity building activities, these shares drop as we look at more junior positions, except for PhD students who report in similar shares scientific engagement activities and information on their trajectory.

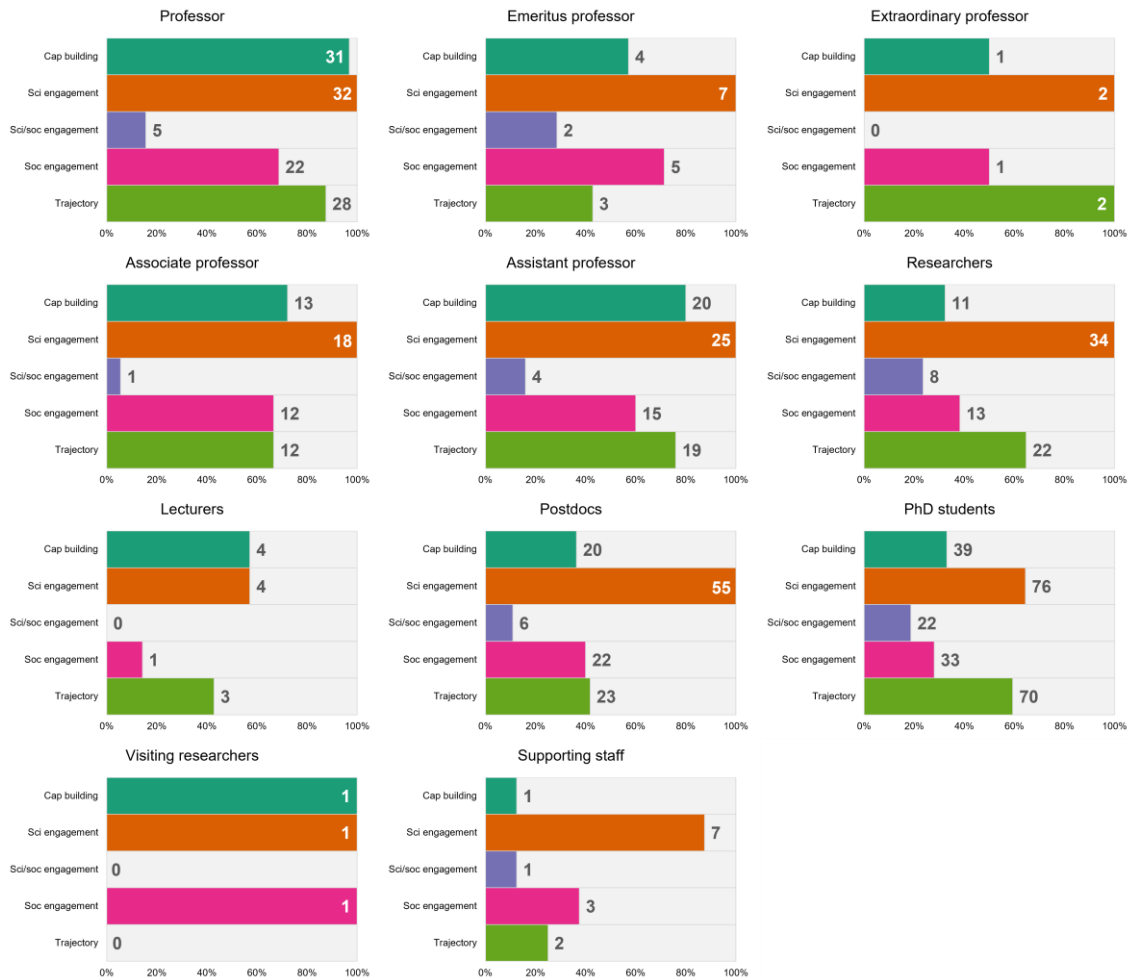


Figure 4. Number of individuals reporting activities per dimension for the different academic statuses.

There are also differences on the number of individuals reporting social engagement activities. Between 60 and 70% of assistant, associate and full professors report some activity in this dimension. This share drops to 40% for postdoctoral researchers and less than 30% for PhD students.

Figure 5 breaks down the different activities by dimension according to academics' status or career stage. Again, publications are the most reported activities, followed by mobility and grant acquisition. The third activity identified for associate professors is presence in social media. When asked, an interviewee who was very active on Twitter, they confirmed that they used it "100% professional". Queried by what type of things they would tweet about, they replied "mostly papers that I have or other papers that I found interesting or something happening at the institute, things like that".

		prof.	emeritus	extr. prof.	assoc. prof.	assist. prof.	researchers	lecturers	postdocs	PhDs	visiting	supp. staff
Cap building												
FU	funding distr.	3	0	0	1	0	1	0	3	0	0	0
GR	grants/awards	21	1	0	6	12	5	0	19	25	0	1
ME	mentoring	17	1	1	6	9	3	0	8	0	0	0
PH	PhD committees	2	0	0	0	4	1	0	3	0	0	0
PM	project manag.	2	1	0	3	4	1	1	4	0	0	0
TE	teaching manag.	5	0	0	2	8	1	3	2	1	0	1
TL	team leader	5	0	0	6	3	1	0	2	0	0	0
UN	univ. manag.	4	2	0	2	2	1	1	1	0	0	0
OT	other	1	0	0	0	0	0	0	1	0	0	0
Sci engagement												
AW	award committees	2	0	0	0	0	0	0	0	0	0	0
CO	conference org.	9	0	0	4	9	3	0	8	4	0	1
ED	editorial activities	14	1	0	7	8	6	0	13	4	0	0
FI	fieldwork	0	0	0	0	0	0	0	6	4	0	1
OD	open data/code	1	1	0	2	3	2	0	14	8	0	0
SC	sci memberships	19	2	0	5	5	3	0	9	1	0	0
PU	sci. pubs	31	7	2	18	25	34	6	54	70	1	7
SO	software	2	2	0	1	0	2	0	1	0	0	0
Sci/soc engagement												
OP	open teaching	5	2	0	0	3	4	0	3	0	0	0
ST	student org.	0	0	0	1	0	0	0	1	1	0	0
Soc engagement												
AS	acad. service	6	0	0	4	4	6	0	3	1	0	0
AR	artwork	0	1	0	0	1	0	0	2	1	0	0
MC	media coverage	6	1	0	0	3	1	0	5	2	0	0
NM	non-acad. member.	4	0	0	4	3	2	0	0	2	0	0
OT	other	1	0	0	0	0	0	0	0	0	0	0
PR	private sector	3	0	0	3	1	1	0	1	1	0	0
PO	public outreach	6	3	1	3	10	1	0	14	7	0	1
PS	public sector	2	1	1	0	0	2	1	0	0	0	0
SM	social media	6	1	0	8	9	3	2	11	23	1	1
VO	volunteering	0	1	0	0	2	1	0	1	10	0	1
Trajectory												
MO	mobility	22	3	0	11	12	12	1	22	36	0	1
NE	non-acad. experience	5	1	2	4	10	12	4	1	48	0	1
TOTAL		32	7	2	18	25	34	7	55	118	1	8

Figure 5. Overall distribution of individuals by activity, dimension and academic status. Colour grading by academic status and dimension. The darker the colour grading the higher the number of individuals conduct a given activity within the same academic status.

Although not many of the individuals analysed had an account on general social media platforms (e.g., Twitter), having an academic website presence or a profile in sites like LinkedIn, ResearchGate or Google Scholar was quite common within all cohorts of academics. But they were valued differently. A professor from Physics A who had a personal website, thought that “especially for young researchers it is important”, but that in their case, they already had a reputation and hence, “for me to push what I am doing in a website is slightly less important”. This was shared by a younger academic who had a website by stating that they first built it for “finding jobs”, although in other cases they reported that it was more for the sake of experimenting. A researcher from Social Sciences B indicated that “sometimes people got in touch with me from LinkedIn”, finding it “quite nice to see people [using] these medium for interaction”.

The type of activities undertaken or reported have also much to do with how academics believe they will be valued by others, especially when still in the market. Hence, we observe how postdoctoral researchers will emphasize their publications and international experience, while PhD students will include volunteering or non-academic experience.

But the evaluative focus on publications in many cases, and especially with non-traditional profiles, can become a burden. The few lecturers included in our analysis would still emphasize their publication record, despite “the main aspect we are looking for is teaching”. A lecturer in charge of the design and management of several education programmes highlighted how the fact that they were not active in research and had not undertaken a PhD was preventing them from advancing in their career in educational management.

Perception of others' assessment	Self-assessment	
1 st International experience;	1 st International experience;	
2 nd non-academic experience;	2 nd non-academic experience;	
3 rd research trajectory	3 rd research trajectory	
1 st Most cited publications;	1 st mentoring;	
2 nd grants	2 nd research trajectory;	
	3 rd societal impact	
1 st Publications;	1 st Publications;	
2 nd funding	2 nd software development	
1 st Collaboration;	1 st Mentoring;	
2 nd Productivity	2 nd teaching	
1 st Grants;	Interconnections between activities towards solving a societal challenge	
2 nd publications;		
3 rd mentoring		
1 st Publications;	1 st Teaching;	
2 nd first authored papers;	2 nd Social outreach	
3 rd mentoring		
1 st Grants;	1 st Grants;	
2 nd publications;	2 nd publications;	
3 rd funding	3 rd funding	
1 st Research trajectory	1 st Organisational skills;	
	2 nd vision and capacity	

Dimension

Capacity Building Sci engagement Sci/Soc engagement Soc engagement

Trajectory

Figure 6. Interviewees' perception on activities valued by others versus their own self-assessment. Each activity has been linked to its evaluative dimension.

Confronted valuation regimes between and within academics

7 of the 8 academics interviewed clearly viewed a conflict between their valuation principles and how they perceived they were being assessed. And even the one who did not acknowledge it openly, would still point at the difficulty of maintaining a balance when evaluating the overall performance of an academic:

The institute is evaluated, my performance, my job is evaluated, there is the education evaluation, there is the evaluation of teaching, we look at the evaluation of outreach... It is not separate, but all of those five or six different things have different balance and different ways of evaluating. (Physics A)

To further dig into these conflicting valuation regimes, 7 of the respondents were asked to indicate which are the characteristics of their profile they believed others valued and which were the ones they valued themselves. Figure 6 summarises their responses and links each of those aspects (when possible) with an evaluative dimension. While the activities highlighted partly

reflected their own profiles, we observe some common patterns., such as the omnipresence of publications for the former, and their lower presence in the latter case. Interviewees consider mainly scientific engagement and capacity building activities as the most valued by their colleagues (except for one interviewee who also highlighted career trajectory). But when asked about their own valuations, they give room to overlapping activities of social and scientific engagement.

To explain these differences, several interviewees would refer to academia as “playing the scientific game”, where “you have to apply certain rules” (Social Sciences B). Hence, differences between self-assessment and valuation regimes would be explained in terms of external principles to which one has to adhere. This belief makes change difficult, as “you have to play the game in order to be able to change it” (Biomedicine B).

That goes for the young scientists as well, they have to play the game in the sense that they have to acquire a high h-index, have to do good publications, etc., apply for a VENI⁶, go through all the loops even if they think that it is not a good system. The only way to change it is when you are already in the system. But then the incentive to change is very low. (Biomedicine B)

Such situation affects mostly those playing non-traditional roles or in educational managerial positions, as they are perceived as outsiders. In this sense, a lecturer indicated that “the reward system in academia is not for generalists, is for specialists on some field”, reflecting on their own difficulties to advance their career in academia.

But these difficulties are also present for those academics who are good in the so-called *game*. Indeed, a highly prolific professor discussed how the excessive focus on citation impact could negatively affect certain types of research. They gave the following example:

I am really proud of this special issue with papers on observational studies in Iran, in Yemen, in Indonesia, Brazil, Kenya... But people don't cite those papers unless they are doing something on culture specifically on that country, because as soon as it says Peru in the title when you are doing a simple search on [Research field] it's going to get discarded by western researchers, because they don't think it applies to them. (Social Sciences A)

Two respondents indicated that this excessive focus on publications as a proxy for academic performance had further consequences on future generations of scholars. A professor in social sciences saw the academic system as becoming “a bit of a factory”, where “an actual deep intellectual development doesn't always happen”. This affected directly mentoring and teaching activities as they saw that new PhDs were individuals who could “competently carry out this type of research and write papers in this”, but who did not adhere to “what a PhD I think, originally was intended to be”, in terms of delivering intellectuals.

A professor in biomedicine even went further and suggested a radical change in the reward system:

Probably, especially in academia, all the generations of the staff, that's also quite a high percentage of full professors, they should be... especially they⁷ should be valued

⁶ Grant for talented young researchers awarded by the Dutch Research Council (NWO). The overall success rate between 2010 and 2012 was of 16.5% (Lee and Ellemers 2015)

⁷ Referring to full professors

differently. And I would think that, maybe at some point in time we shouldn't give high and large grants programme finance for let's say professors over the age of 55. Let them go and do what they should be doing at that time in their career, and that is coaching. Whether it is coaching in science or coaching in education I don't care. But let's save all the grant money for the young and eager, to give them a head start, to make sure that they can use all their energy in an effective and positive way. (Biomedicine B)

PhD student		Postdoctoral researcher		Tenure tracker	
1 st Field of expertise;		1 st International experience;		1 st Non-academic experience;	
2 nd research experience;		2 nd research topics;		2 nd social network;	
3 rd grades		3 rd theoretical or applied		3 rd social skills	
1 st Background;		1 st Background;		1 st Publication record;	
2 nd diversity of activities;		2 nd research topics;		2 nd leadership qualities	
3 rd Versatility		3 rd school of thought			
1 st Diverse background;		1 st Publications;		1 st Teaching;	
2 nd grades;		2 nd PhD thesis;		2 nd publications;	
3 rd master thesis		3 rd trajectory		3 rd co-authorships	
1 st project requirements;		1 st Social skills;		1 st Social skills	
2 social skills;		2 nd scientific and academic goals			
3 rd grades					
1 st Cultural and geographical background;		1 st Cultural and geographical background;		1 st Teaching experience	
2 nd diversity of activities;		2 nd diversity of activities;			
3 rd open-minded		3 rd open-minded			
1 st Recommendation letters;		1 st Publications;		1 st Management skills;	
2 nd social skills		2 nd independence;		2 nd social skills	
		3 rd scientific goals			
1 st Grades;		1 st Grades;		1 st Scientific excellence;	
2 nd number and difficulty of activities;		2 nd number and difficulty of activities;		2 nd cutting edge;	
3 rd research experience		3 rd research experience		3 rd social skills	
Dimension					
<div>Capacity Building</div> <div>Sci engagement</div> <div>Sci/Soc engagement</div> <div>Soc engagement</div> <div>Trajectory</div>					

Figure 7. Top three characteristics one would look in a CV if they were to recruit a PhD student, postdoctoral researcher and tenure tracker. Each activity is linked to its evaluative dimension when possible.

In the evaluators' shoes

During the interviews, respondents were asked to adopt an evaluator's stand and discuss the types of activities they would value if they were to recruit new scholars. Specifically, they were asked to indicate which were the three top characteristics they would look for in a CV when recruiting a PhD student, a postdoctoral researcher and a tenure tracker. One of their common reactions was to indicate the difficulty of the task, especially with the PhD student. The second reaction was to mention the things they would look for in an interview. Figure 7 includes a list for 7 of the respondents⁸.

It was common for interviewees to emphasize the need of a successful candidate to have "some kind of collaborative spirit", showing their capability to "fit in" within the team, department and institution. For PhD students, as well as having the right background and "reasonable grades" they emphasized the inclusion of "different" activities which reflected "not a boring CV", and a diversity in terms of subjects. Criteria for postdoctoral researchers tended to be more normative, in the sense that respondents would highlight the need to have the right profile for the project at hand, a good scientific profile as well as trajectory. In one case, a social scientist

⁸ The eighth interviewee was not asked this question due to this person's profile in teaching management.

highlighted the importance of being of the same school of thought as their research. They explained the need for this in the following way:

There are different research philosophies and if you don't have the right research philosophy, if you are completely educated in a [school of thought contrary to the interviewee's] ... You know, in my research there is interpretivism, constructivism, etc. with that there is often, there will be a tension. So that's not smart doing, because somebody will not be happy. I think at the end it should fit and it should benefit both situations.

Lastly, in the case of tenure trackers, they emphasized social skills as well as management and teaching skills, characteristics linked with the social and scientific engagement, and capacity building dimensions. An assistant professor indicated that "for tenure track you also look for how they are at managing groups, so they might be super smart but then they are not good with people and if you are going to have a PI then you definitely want someone who knows how to manage a group". Also their potential as scientists was highly regarded, as "tenure track is about potential of becoming a full professor" (Social Sciences B). A research unit director indicated that they looked for individuals who were "top-notch and top of the field", people who were not only excellent but who "know where the field is going and then move on". However, this ideal of excellent researcher was not shared by all. One of the interviewees argued that a high focus on scientific productivity is what converts the scientific career into a *scientific game*. In their words:

What I see now is that it is a selection process, a very strict selection process, and some people are good enough, they are just good enough and they reach it. But then there is a majority that is basically just competing, and they are roughly the same, I am probably also in this group, and then other components come into, things like who is more aggressive, who is more capable of playing the game. (Biomedicine A)

Personal features and external factors affect perceptions

During the interviews also personal features such as gender or nationality, as well as external factors such as departmental evaluations or national research programmes, would come into the conversation, in most cases by the interviewees themselves. For instance, when discussing the importance of academic titles in order to progress in academia, a respondent added "especially if you are not old, grey, white and male", highlighting the importance that age and gender have in academic career development. Another interviewee who indicated that they would value diversity when recruiting someone, added the following when later asked to expand:

You should have diversity in all aspects. If you think about gender balance, you should have male and female. I see some parts of our university are only male, I think you miss something because you need other type of people. Also, the type of people you need is also based on diversity independent of gender. You shouldn't have only Dutch people. I think that is not healthy. We should have people from abroad and diverse, preferably from everywhere, with experience in other universities.

Diversity of personal features is not only thought of in terms of having a healthy institution, but also as a matter of principles. In this sense, another respondent indicated that had recently decided recruiting, when possible, only PhD students from the Global South, arguing the following:

Most professors when they get an email from Yemen or from Indonesia or from Kenya, they will delete it. That's what I know from colleagues, they will delete it, they won't even read it. Those people don't have an easy opportunity to do a PhD or to be supervised properly in developing themselves. So I am probably most needed there, plus those people always have something to teach me.

Regarding external factors, programmes such as the Dutch Veni, Vidi and Vici personal grants were recurrently mentioned as well as departmental annual evaluations⁹. Regarding the former, a respondent summarised its arbitrariness by stating “I'm just on the lucky side of the Matthew Effect”, while others were highly critical with this system. Regarding the annual evaluations, a respondent described them as a “moving target”, signalling the difficulty to plan one's own career based on the institution's expectations on their performance.

Concluding remarks

In this study we questioned how scientists conduct a diversity of tasks in an evaluative context overly focused on scientific publishing. We conducted a multiple case-study analysis in which we examined CVs and other public information related to academics from five research units at two Dutch universities. We then performed semi-structured interviews to 8 of the academics in these departments. To study diversity of activities in academia, we proposed the use of a valuation model consisting of five evaluative dimensions of activities, external factors affecting performance and personal features. Our findings yield the following conclusions.

First, the model proposed seems to be suitable to this type of analyses, as activities were easily assigned to each dimension, while at the same time providing a clear cut by which we could rapidly grasp differences between profiles and valuation regimes. Our curriculum analysis explores mainly differences between evaluative dimensions, and we delve into how these are influenced by personal features and external factors in the interviews. These analyses show a highly diverse scientific workforce between and within research units, with no distinct pattern by fields, but rather by individual's career stage. However, we did not thoroughly examine relations between the different elements of the model, nor deepened each of them thoroughly. We believe this model can be further exploited to study specific activities or aspects of academic activity, as well as to analyse the relation between the type of activities academics conduct and their features or the environment in which they work. In this paper, all academics were placed within the same national evaluation system. It would be of interest to conduct a similar analysis, including for example scholars from the humanities and law domains, but also conduct cross-country comparisons to better understand differences between countries, cultures, etc.

Second, we do observe a misalignment between valuation regimes. While this is not surprising given previous findings (e.g., Fochler et al. 2016), this study has served to disentangle how these mismatches are produced. In our interviews we discuss such diversity of profiles as well as the value they attach to each of them, and the role evaluation plays on their development. Their responses reflect a diverse and critical workforce, who find many times evaluation principles working against their own valuation principles. Much of their criticisms seems to be directed at a system which targets excessively research performance, rewarding short-term (e.g., publications) rather than long-term goals (e.g., societal impact, mentoring, teaching). The interviewees were selected on the basis of the diversity of profiles observed and hence are not

⁹ These are annual interviews with the heads of department to assess yearly academics' performance. More information here: <https://www.staff.universiteitleid.nl/human-resources/pd-interviews>

a random sample. But they all reveal a common effort to juggle between their valuation regimes and what they believe is expected from them, in many cases with very strong opinions on how the current evaluation system works. In most cases they do manage to maintain this balance successfully but feel frustrated by it. This juggling has not only to do with a lack of diversity on the activities valued in research evaluation systems, but also with the inability of the metrics to contextualize and adapt to their needs. On the other hand, they seem to lack the tools when they are the ones evaluating and sometimes referred to “going with their guts” when deciding who was the best candidate for a given position.

Finally, we observe how activities are valued differently depending on academics’ career stage. While this may be reasonable, it questions to what extent researchers in each stage are provided with the opportunity to prepare for the next stage and acquire the required experience and expertise. This paper aims at contributing to the field of research evaluation by going one step back and reflecting on the activities academics conduct, the worth they place versus the worth placed in evaluation schemes, and the struggles academics face to meet the expectations set by these schemes. We consider this type of studies a crucial step before moving forward to new alternative proposals to current schemes. While the limitations of metrics-based evaluation systems have been thoroughly studied and pointed out in the literature, an alternative may not necessarily be better just because it is different. In order to develop responsible evaluative measures to nurture and foster a sustainable academic system, we must learn from past experiences and understand the specific problems and errors that these have created.

Acknowledgments

The authors are grateful to all participants of this study and particularly to the heads of the research units analysed as well as the interviewees. This work was supported by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 707404 and Rodrigo Costas is partly supported by the South African DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP). Nicolas Robinson-Garcia is currently supported by a Ramón y Cajal grant from the Spanish Ministry of Science (RYC2019-027886-I). The authors declare that they have no competing interests.

References

- Abramo, Giovanni, Ciriaco Andrea D’Angelo, and Francesco Rosati. 2015. The determinants of academic career advancement: Evidence from Italy. *Science and Public Policy* 42: 761–774. <https://doi.org/10.1093/scipol/scu086>.
- ACUMEN. 2014. *Guidelines for Good Evaluation Practice with the ACUMEN Portfolio*. Deliverable D6.14.
- Alperin, Juan P, Carol Muñoz Nieves, Lesley A Schimanski, Gustavo E Fischman, Meredith T Niles, and Erin C McKiernan. 2019. How significant are the public dimensions of faculty work in review, promotion and tenure documents? Edited by Emma Pewsey, Peter A Rodgers, Emily Janke, and Heather Coates. *eLife* 8: e42254. <https://doi.org/10.7554/eLife.42254>.
- Archambault, Éric, Étienne Vignola-Gagne, Grégoire Côté, Vincent Larivière, and Yves Gingrasb. 2006. Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics* 68: 329–342.

- Azoulay, Pierre, Ina Ganguli, and Joshua S. Graff Zivin. 2016. *The Mobility of Elite Life Scientists: Professional and Personal Determinants*. Working Paper 21995. National Bureau of Economic Research. <https://doi.org/10.3386/w21995>.
- Baker, Monya. 2016. 1,500 scientists lift the lid on reproducibility. *Nature News* 533: 452. <https://doi.org/10/gdgzjx>.
- van den Besselaar, Peter, Ulf Heyman, and Ulf Sandström. 2017. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* 11: 905–918. <https://doi.org/10.1016/j.joi.2017.05.016>.
- Bigger, Patrick, and Morgan Robertson. 2017. Value is Simple. Valuation is Complex. *Capitalism Nature Socialism* 28. Routledge: 68–77. <https://doi.org/10.1080/10455752.2016.1273962>.
- Bonaccorsi, Andrea, and Cinzia Daraio. 2003. Age effects in scientific productivity. *Scientometrics* 58: 49–90. <https://doi.org/10/c68b6w>.
- Bone, Frédérique, Michael M. Hopkins, Ismael Ràfols, Jordi Molas-Gallart, Puay Tang, Gail Davey, and Antony M. Carr. 2019. *DARE to be different? Applying diversity indicators to the evaluation of collaborative research projects*. 2019–09. SPRU Working Paper Series. SPRU - Science Policy Research Unit, University of Sussex Business School.
- Bonn, Noémie Aubert, and Wim Pinxten. 2021. Advancing science or advancing careers? Researchers' opinions on success indicators. *PLOS ONE* 16. Public Library of Science: e0243664. <https://doi.org/10.1371/journal.pone.0243664>.
- Bourdieu, Pierre. 1975. The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information* 14. Sage Publications Sage CA: Thousand Oaks, CA: 19–47.
- Bourdieu, Pierre. 1986. The forms of capital. In *Handbook of theory and research for the sociology of education*, ed. J. Richardson, 241–258:258. Westport, CT: Greenwood.
- Bouter, Lex M., Joeri Tjink, Nils Axelsen, Brian C. Martinson, and Gerben Ter Riet. 2016. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. *Research Integrity and Peer Review* 1: 17.
- Bozeman, Barry, James S. Dietz, and Monica Gaughan. 2001. Scientific and technical human capital: an alternative model for research evaluation. *International Journal of Technology Management* 22: 716–740. <https://doi.org/10.1504/IJTM.2001.002988>.
- Bozeman, Barry, and Daniel Sarewitz. 2011. Public Value Mapping and Science Policy Evaluation. *Minerva* 49: 1–23. <https://doi.org/10.1007/s11024-011-9161-7>.
- Butler, Linda. 2003. Explaining Australia's increased share of ISI publications—the effects of a funding formula based on publication counts. *Research Policy* 32: 143–155. [https://doi.org/10.1016/S0048-7333\(02\)00007-0](https://doi.org/10.1016/S0048-7333(02)00007-0).
- Cañibano, Carolina, and Barry Bozeman. 2009. Curriculum vitae method in science policy and research evaluation: the state-of-the-art. *Research Evaluation* 18: 86–94. <https://doi.org/10.3152/095820209X441754>.

- Cañibano, Carolina, Mary Frank Fox, and F. Javier Otamendi. 2016. Gender and patterns of temporary mobility among researchers. *Science and Public Policy* 43: 320–331. <https://doi.org/10.1093/scipol/scv042>.
- Cañibano, Carolina, Immaculada Vilardell, Carmen Corona, and Carlos Benito-Amat. 2018. The evaluation of research excellence and the dynamics of knowledge production in the humanities: The case of History in Spain. *Science and Public Policy* 45: 775–789. <https://doi.org/10.1093/scipol/scy025>.
- Caron, Emiel, and Nees Jan van Eck. 2014. Large scale author name disambiguation using rule-based scoring and clustering. In *19th International Conference on Science and Technology Indicators. "Context counts: Pathways to master big data and little data,"* 79–86. CWTS-Leiden University Leiden.
- Chavarro, Diego, Puay Tang, and Ismael Ràfols. 2017. Why researchers publish in non-mainstream journals: Training, knowledge bridging, and gap filling. *Research Policy* 46: 1666–1680. <https://doi.org/10.1016/j.respol.2017.08.002>.
- Chawla, Dalmeet Singh. 2021. Scientists at odds on Utrecht University reforms to hiring and promotion criteria. *Nature Index*.
- Costas, Rodrigo, Thed N. van Leeuwen, and María Bordons. 2010. A bibliometric classificatory approach for the study and assessment of research performance at the individual level: The effects of age on productivity and impact. *Journal of the American Society for Information Science and Technology* 61: 1564–1581. <https://doi.org/10.1002/asi.21348>.
- Curry, Stephen. 2018. Let's move beyond the rhetoric: it's time to change how we judge research. *Nature* 554: 147–147. <https://doi.org/10.1038/d41586-018-01642-w>.
- Curry, Stephen, Sarah de Rijcke, Anna Hatch, Dorsamy (Gansen) Pillay, Inge van der Weijden, and James Wilsdon. 2020. *The changing role of funders in responsible research assessment: progress, obstacles and the way ahead*. Report. Research on Research Institute. <https://doi.org/10.6084/m9.figshare.13227914.v1>.
- Dalen, Hendrik P. van, and Kène Henkens. 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: 1282–1293. <https://doi.org/10.1002/asi.22636>.
- De Bellis, Nicola. 2014. History and Evolution of (Biblio)Metrics. In *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, ed. Blaise Cronin and Cassidy R. Sugimoto, 23–44. Cambridge, MA: MIT Press.
- Derrick, G. E., A. Jaeger, P.-Y. Chen, C. R. Sugimoto, T. Van Leeuwen, and V. Larivière. 2019. Models of parenting and its effect on academic productivity: Preliminary results from an international survey. In , 1670–1676. ITA: International Society for Informetrics and Scientometrics.
- Dijstelbloem, Huub, Frank Huisman, Frank Miedema, and Wijnand Mijnhardt. 2013. Why science does not work as it should. And what to do about it. *Science in Transition Position Paper*.

- Ding, Jielan, Zhesi Shen, Per Ahlgren, Tobias Jeppsson, David Minguillo, and Johan Lyhagen. 2021. The link between ethnic diversity and scientific impact: the mediating effect of novelty and audience diversity. *Scientometrics*. <https://doi.org/10.1007/s11192-021-04071-1>.
- Fanelli, Daniele. 2018. Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences* 115: 2628–2631. <https://doi.org/10/gc5ppc>.
- Fanelli, Daniele, Rodrigo Costas, Ferric C. Fang, Arturo Casadevall, and Elisabeth M. Bik. 2019. Testing Hypotheses on Risk Factors for Scientific Misconduct via Matched-Control Analysis of Papers Containing Problematic Image Duplications. *Science and Engineering Ethics* 25: 771–789. <https://doi.org/10/c2kx>.
- Fochler, Maximilian, Ulrike Felt, and Ruth 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: 175–200. <https://doi.org/10.1007/s11024-016-9292-y>.
- Fochler, Maximilian, and Lisa Sigl. 2018. Anticipatory Uncertainty: How Academic and Industry Researchers in the Life Sciences Experience and Manage the Uncertainties of the Research Process Differently. *Science as Culture* 27. Routledge: 349–374. <https://doi.org/10.1080/09505431.2018.1485640>.
- Freeman, Richard B., and Wei Huang. 2014. Collaboration: Strength in diversity. *Nature* 513: 305–305. <https://doi.org/10.1038/513305a>.
- Frodeman, Robert, and Jonathan Parker. 2009. Intellectual merit and broader impact: The National Science Foundation's broader impacts criterion and the question of peer review. *Social Epistemology* 23: 337–345.
- Gadd, Elizabeth. 2015. When are journal metrics useful? A balanced call for the contextualized and transparent use of all publication metrics. *Impact of Social Sciences*.
- Given, Lisa M. 2008. *Semi-structured interview*. *The SAGE encyclopedia of qualitative research methods*. Vol. 2. Sage Publications Thousand Oaks, CA.
- Gläser, Jochen, and Grit Laudel. 2007. Evaluation Without Evaluators. In *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, ed. Richard Whitley and Jochen Gläser, 127–151. Sociology of the Sciences Yearbook. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-6746-4_6.
- Gläser, Jochen, and Grit Laudel. 2013. Life With and Without Coding: Two Methods for Early-Stage Data Analysis in Qualitative Research Aiming at Causal Explanations. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 14. <https://doi.org/10.17169/fqs-14.2.1886>.
- Gómez, Isabel, Maria Bordons, M. T. Fernández, and Aida Méndez. 1996. Coping with the problem of subject classification diversity. *Scientometrics* 35: 223–235. <https://doi.org/10.1007/BF02018480>.
- Hammarfelt, Björn, and Alexander D. Rushforth. 2017. Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation* 26: 169–180. <https://doi.org/10.1093/reseval/rvx018>.

Harzing, Anne-Wil. 2007. *Publish or Perish*.

Heijstra, Tamar M, Þorgerður Einarsdóttir, Gyða M Pétursdóttir, and Finnborg S Steinþórsdóttir. 2017. Testing the concept of academic housework in a European setting: Part of academic career-making or gendered barrier to the top? *European Educational Research Journal* 16: 200–214. <https://doi.org/10.1177/1474904116668884>.

Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. 2015. The Leiden Manifesto for research metrics. *Nature* 520: 429–431. <https://doi.org/10.1038/520429a>.

Holbrook, J. Britt. 2012. Re-assessing the science-society relation: The case of the US National Science Foundation's broader impacts merit review criterion (1997-2011). *Technology in Society* 27: 437–451.

Huang, Junming, Alexander J. Gates, Roberta Sinatra, and Albert-László Barabási. 2020. Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences* 117. National Academy of Sciences: 4609–4616. <https://doi.org/10.1073/pnas.1914221117>.

Ioannidis, John P. A. 2005. Why Most Published Research Findings Are False. *PLOS Medicine* 2: e124. <https://doi.org/10/chhf6b>.

Ioannidis, John P. A., and Muin J. Khoury. 2014. Assessing Value in Biomedical Research: The PQRS of Appraisal and Reward. *JAMA* 312: 483–484. <https://doi.org/10.1001/jama.2014.6932>.

Irwin, Alan. 1995. *Citizen Science. A Study of People, Expertise and Sustainable Development*. Environment and Society. London and New York: Routledge.

Jackson, Susan E., Aparna Joshi, and Niclas L. Erhardt. 2003. Recent Research on Team and Organizational Diversity: SWOT Analysis and Implications. *Journal of Management* 29: 801–830. [https://doi.org/10.1016/S0149-2063\(03\)00080-1](https://doi.org/10.1016/S0149-2063(03)00080-1).

Jiménez-Contreras, Evaristo, Félix de Moya Anegón, and Emilio Delgado López-Cózar. 2003. The evolution of research activity in Spain: The impact of the National Commission for the Evaluation of Research Activity (CNEAI). *Research policy* 32: 123–142.

Kahneman, Daniel, and Amos Tversky. 1972. Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3: 430–454. [https://doi.org/10.1016/0010-0285\(72\)90016-3](https://doi.org/10.1016/0010-0285(72)90016-3).

Karaulova, Maria, Abdullah Gök, and Philip Shapira. 2019. Identifying author heritage using surname data: An application for Russian surnames. *Journal of the Association for Information Science and Technology* 70: 488–498. <https://doi.org/10.1002/asi.24104>.

Knorr-Cetina, Karin D., and Michael Mulkay. 1983. Introduction: Emerging principles in social studies of science. In *Science Observed, Perspectives on the Social Study of Science*, ed. Karin D. Knorr-Cetina and Michael Mulkay, 1–17. London: SAGE.

Lam, Alice. 2011. What motivates academic scientists to engage in research commercialization: 'Gold', 'ribbon' or 'puzzle'? *Research Policy* 40: 1354–1368. <https://doi.org/10.1016/j.respol.2011.09.002>.

Larivière, Vincent, Nadine Desrochers, Benoît Macaluso, Philippe Mongeon, Adèle Paul-Hus, and Cassidy R Sugimoto. 2016. Contributorship and division of labor in knowledge production. *Social Studies of Science* 46: 417–435. <https://doi.org/10.1177/0306312716650046>.

Larivière, Vincent, Chaoqun Ni, Yves Gingras, Blaise Cronin, and Cassidy R. Sugimoto. 2013. Bibliometrics: Global gender disparities in science. *Nature* 504: 211–213. <https://doi.org/10/qgf>.

Larivière, Vincent, David Pontille, and Cassidy R. Sugimoto. 2020. Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies*. MIT Press: 1–18. https://doi.org/10.1162/qss_a_00097.

Latour, Bruno, and Steve Woolgar. 1979. *Laboratory life: The construction of scientific facts*. Princeton University Press.

Laudel, Grit, and Jochen Glaser. 2011. *Academic careers and how to find research excellence*. Vol. 2011.

Lee, Romy van der, and Naomi Ellemers. 2015. Gender contributes to personal research funding success in The Netherlands. *Proceedings of the National Academy of Sciences* 112. National Academy of Sciences: 12349–12353. <https://doi.org/10.1073/pnas.1510159112>.

van Leeuwen, Thed N., Henk F. Moed, Robert J. W. Tijssen, Martijn S. Visser, and Anthony F. J. Van Raan. 2001. Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics* 51: 335–346. <https://doi.org/10.1023/A:1010549719484>.

Lerchenmueller, Marc J., and Olav Sorenson. 2018. The gender gap in early career transitions in the life sciences. *Research Policy* 47: 1007–1017. <https://doi.org/10.1016/j.respol.2018.02.009>.

Levin, Sharon G., and Paula E. Stephan. 1991. Research Productivity Over the Life Cycle: Evidence for Academic Scientists. *The American Economic Review* 81: 114–132.

Levin, Sharon G., and Paula E. Stephan. 1999. Are the Foreign Born a Source of Strength for U.S. Science? *Science* 285: 1213–1214. <https://doi.org/10.1126/science.285.5431.1213>.

Lin, Zeng, Richard Pearce, and Weirong Wang. 2009. Imported talents: Demographic characteristics, achievement and job satisfaction of foreign born full time faculty in four-year American colleges. *Higher Education* 57: 703–721. <https://doi.org/10.1007/s10734-008-9171-z>.

López Piñeiro, Carla, and Diana Hicks. 2015. Reception of Spanish sociology by domestic and foreign audiences differs and has consequences for evaluation. *Research Evaluation* 24: 78–89. <https://doi.org/10.1093/reseval/rvu030>.

Luukkonen, Terttu. 1997. Why has Latour's theory of citations been ignored by the bibliometric community? discussion of sociological interpretations of citation analysis. *Scientometrics* 38: 27–37. <https://doi.org/10.1007/BF02461121>.

Luukkonen, Terttu. 2002. Research evaluation in Europe: state of the art. *Research Evaluation* 11: 81–84. <https://doi.org/10.3152/147154402781776871>.

- Macaluso, Benoit, Vincent Larivière, Thomas Sugimoto, and Cassidy R. Sugimoto. 2016. Is Science Built on the Shoulders of Women? A Study of Gender Differences in Contributorship. *Academic Medicine* 91: 1136–1142. <https://doi.org/10.1097/ACM.0000000000001261>.
- Mamiseishvili, Ketevan, and Vicki J. Rosser. 2010. International and citizen faculty in the United States: An examination of their productivity at research universities. *Research in Higher Education* 51: 88–107. <https://doi.org/10.1007/s11162-009-9145-8>.
- McKiernan, Erin C., Lesley A. Schimanski, Carol Muñoz Nieves, Lisa Matthias, Meredith T. Niles, and Juan Pablo Alperin. 2019. *Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations*. e27638v2. PeerJ Inc. <https://doi.org/10.7287/peerj.preprints.27638v2>.
- Merton, Robert K. 1968. The Matthew Effect in Science The reward and communication systems of science are considered. *Science* 159: 56–63. <https://doi.org/10.1126/science.159.3810.56>.
- Merton, Robert K. 1973. The normative structure of science. The sociology of science. *University of Chicago Press, Chicago*.
- Moher, David, Lex Bouter, Sabine Kleinert, Paul Glasziou, Mai Har Sham, Virginia Barbour, Anne-Marie Coriat, Nicole Foeger, and Ulrich Dirnagl. 2020. The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLOS Biology* 18. Public Library of Science: e3000737. <https://doi.org/10.1371/journal.pbio.3000737>.
- Moher, David, Steven N. Goodman, and John P. A. Ioannidis. 2016. Academic criteria for appointment, promotion and rewards in medical research: where's the evidence? *European Journal of Clinical Investigation* 46: 383–385. <https://doi.org/10.1111/eci.12612>.
- Morgan, Allison C., Dimitrios J. Economou, Samuel F. Way, and Aaron Clauset. 2018. Prestige drives epistemic inequality in the diffusion of scientific ideas. *EPJ Data Science* 7. SpringerOpen: 1–16. <https://doi.org/10.1140/epjds/s13688-018-0166-4>.
- Murray, Dakota, Kyle Siler, Vincent Larivière, Wei Mun Chan, Andrew M. Collings, Jennifer Raymond, and Cassidy R. Sugimoto. 2019. Gender and international diversity improves equity in peer review. *bioRxiv*. Cold Spring Harbor Laboratory: 400515. <https://doi.org/10.1101/400515>.
- Nederhof, Anton J. 2006. Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review. *Scientometrics* 66: 81–100. <https://doi.org/10.1007/s11192-006-0007-2>.
- Nielsen, Mathias Wullum, Christine Friis Baker, Emer Brady, Michael Bang Petersen, and Jens Peter Andersen. 2021. Meta-Research: Weak evidence of country- and institution-related status bias in the peer review of abstracts. Edited by Peter Rodgers. *eLife* 10. eLife Sciences Publications, Ltd: e64561. <https://doi.org/10.7554/eLife.64561>.
- Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, et al. 2015. Promoting an open research culture. *Science* 348: 1422–1425. <https://doi.org/10.1126/science.aab2374>.
- Nosek, Brian A., Jeffrey R. Spies, and Matt Motyl. 2012. Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science* 7: 615–631. <https://doi.org/10.1177/1745691612459058>.

- Osuna, Carmen, Laura Cruz-Castro, and Luis Sanz-Menéndez. 2010. Overturning some assumptions about the effects of evaluation systems on publication performance. *Scientometrics* 86: 575–592. <https://doi.org/10.1007/s11192-010-0312-7>.
- Polanyi, Michael. 1962. The republic of science. *Minerva* 1. Springer: 54–73.
- Porter, Theodore M. 1996. *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press.
- Rafols, Ismael, Tommaso Ciarli, and Diego Chavarro. 2015. Under-reporting research relevant to local needs in the global south. Database biases in the representation of knowledge on rice.
- Rafols, Ismael, Loet Leydesdorff, Alice O'Hare, Paul Nightingale, and Andy Stirling. 2012. How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. *Research Policy* 41. Exploring the Emerging Knowledge Base of "The Knowledge Society": 1262–1282. <https://doi.org/10.1016/j.respol.2012.03.015>.
- Rafols, Ismael, and Martin Meyer. 2009. Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics* 82: 263–287. <https://doi.org/10.1007/s11192-009-0041-y>.
- Ràfols, Ismael, Jordi Molas-Gallart, Diego Andrés Chavarro, and Nicolas Robinson-Garcia. 2016. *On the Dominance of Quantitative Evaluation in 'Peripheral' Countries: Auditing Research with Technologies of Distance*. SSRN Scholarly Paper ID 2818335. Rochester, NY: Social Science Research Network.
- Ramos-Vielba, Irene, Mabel Sánchez-Barrioluengo, and Richard Woolley. 2016. Scientific research groups' cooperation with firms and government agencies: motivations and barriers. *The Journal of Technology Transfer* 41: 558–585. <https://doi.org/10.1007/s10961-015-9429-4>.
- Reskin, Barbara F. 1977. Scientific Productivity and the Reward Structure of Science. *American Sociological Review* 42: 491–504. <https://doi.org/10.2307/2094753>.
- de Rijcke, Sarah, Paul F. Wouters, Alex D. Rushforth, Thomas P. Franssen, and Björn Hammarfelt. 2016. Evaluation practices and effects of indicator use—a literature review. *Research Evaluation* 25: 161–169. <https://doi.org/10.1093/reseval/rvv038>.
- Robinson-Garcia, Nicolas, Rodrigo Costas, Cassidy R Sugimoto, Vincent Larivière, and Gabriela F Nane. 2020. Task specialization across research careers. Edited by Peter Rodgers and Allison Morgan. *eLife* 9. eLife Sciences Publications, Ltd: e60586. <https://doi.org/10.7554/eLife.60586>.
- Robinson-Garcia, Nicolas, and Ismael Ràfols. 2020. The Differing Meanings of Indicators Under Different Policy Contexts. The Case of Internationalisation. In *Evaluative Informetrics: The Art of Metrics-Based Research Assessment : Festschrift in Honour of Henk F. Moed*, ed. Cinzia Daraio and Wolfgang Glänzel, 213–232. Cham: Springer International Publishing.
- Rumbley, Laura E., P. G. Altbach, and Liz Reisberg. 2012. Internationalization within the higher education context. *The Sage handbook of international higher education*: 3–26. <https://doi.org/10.4135/9781452218397>.
- Rushforth, Alexander, Thomas Franssen, and Sarah de Rijcke. 2019. Portfolios of Worth: Capitalizing on Basic and Clinical Problems in Biomedical Research Groups. *Science, Technology, & Human Values* 44. SAGE Publications Inc: 209–236. <https://doi.org/10/gdsm6b>.

- Sánchez-Barrioluengo, Mabel. 2014. Articulating the 'three-missions' in Spanish universities. *Research Policy* 43: 1760–1773. <https://doi.org/10.1016/j.respol.2014.06.001>.
- Sarewitz, Daniel. 2016a. Saving Science. *The New Atlantis* 49: 4–40.
- Sarewitz, Daniel. 2016b. The pressure to publish pushes down quality. *Nature News* 533: 147. <https://doi.org/10.1038/533147a>.
- Sarewitz, Daniel, and Roger A. Pielke Jr. 2007. The neglected heart of science policy: reconciling supply of and demand for science. *environmental science & policy* 10: 5–16. <https://doi.org/10.1016/j.envsci.2006.10.001>.
- Schimanski, Lesley A., and Juan Pablo Alperin. 2018. The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future. *F1000Research* 7: 1605. <https://doi.org/10.12688/f1000research.16493.1>.
- Spaapen, Jack, and Leonie van Drooge. 2011. Introducing 'productive interactions' in social impact assessment. *Research Evaluation* 20: 211–218. <https://doi.org/10.3152/095820211X12941371876742>.
- Stephan, Paula E., and Sharon G. Levin. 2001. Exceptional contributions to US science by the foreign-born and foreign-educated. *Population Research and Policy Review* 20: 59–79. <https://doi.org/10.1023/A:1010682017950>.
- Stirling, Andy. 2007. A general framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface* 4: 707–719. <https://doi.org/10/fh85wx>.
- Stürmer, Stefan, Aileen Oeberst, Roman Trötschel, and Oliver Decker. 2017. Early-Career Researchers' Perceptions of the Prevalence of Questionable Research Practices, Potential Causes, and Open Science. *Social Psychology* 48. Hogrefe Publishing: 365–371. <https://doi.org/10.1027/1864-9335/a000324>.
- Trencher, Gregory, Masaru Yarime, Kes B. McCormick, Christopher N. H. Doll, and Steven B. Kraines. 2014. Beyond the third mission: Exploring the emerging university function of co-creation for sustainability. *Science and Public Policy* 41: 151–179. <https://doi.org/10.1093/scipol/sct044>.
- Tversky, Amos, and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185. American association for the advancement of science: 1124–1131.
- Vandenbroucke, Jan P. 2002. The history of confounding. *Sozial- und Präventivmedizin* 47: 216–224. <https://doi.org/10.1007/BF01326402>.
- Viglione, Giuliana. 2020. Are women publishing less during the pandemic? Here's what the data say. *Nature* 581. Nature Publishing Group: 365–366. <https://doi.org/10/dwdw>.
- VSNU and ZonMw. 2019. Room for Everyone's Talent: Towards a New Balance in the Recognition and Rewards of Academics.
- Waijter, Cathelijan J. F., Christine Teelken, Paul F. Wouters, and Inge C. M. van der Weijden. 2018. Competition in Science: Links Between Publication Pressure, Grant Pressure and the Academic Job Market. *Higher Education Policy* 31: 225–243. <https://doi.org/10.1057/s41307-017-0051-y>.

- Wagner, Caroline S., Travis A. Whetsell, and Satyam Mukherjee. 2019. International research collaboration: Novelty, conventionality, and atypicality in knowledge recombination. *Research Policy* 48: 1260–1270. <https://doi.org/10.1016/j.respol.2019.01.002>.
- Walsh, John P., You-Na Lee, and Li Tang. 2019. Pathogenic organization in science: Division of labor and retractions. *Research Policy* 48. Academic Misconduct, Misrepresentation, and Gaming: 444–461. <https://doi.org/10.1016/j.respol.2018.09.004>.
- Waltman, Ludo. 2017. Special section on performance-based research funding systems. *Journal of Informetrics* 11: 904. <https://doi.org/10.1016/j.joi.2017.05.015>.
- van der Weijden, Inge, Christine Teelken, Moniek de Boer, and Mariske Drost. 2016. Career satisfaction of postdoctoral researchers in relation to their expectations for the future. *Higher Education* 72: 25–40. <https://doi.org/10.1007/s10734-015-9936-0>.
- Whitley, Richard. 2000. *The intellectual and social organization of the sciences*. 2nd ed. Oxford: Oxford University Press.
- Wildgaard, Lorna, Jesper W. Schneider, and Birger Larsen. 2014. A review of the characteristics of 108 author-level bibliometric indicators. *Scientometrics* 101: 125–158. <https://doi.org/10.1007/s11192-014-1423-3>.
- Woolston, Chris. 2020. Postdocs under pressure: ‘Can I even do this any more?’ *Nature* 587. Nature Publishing Group: 689–692. <https://doi.org/10.1038/d41586-020-03235-y>.
- Zhang, Lin, Gunnar Sivertsen, Huiying Du, Ying Huang, and Wolfgang Glänzel. 2021. Gender differences in the aims and impacts of research. SocArXiv. <https://doi.org/10.31235/osf.io/9n347>.
- Zhou, Qiuju, Ronald Rousseau, Liying Yang, Ting Yue, and Guoliang Yang. 2012. A general framework for describing diversity within systems and similarity between systems with applications in informetrics. *Scientometrics* 93: 787–812. <https://doi.org/10.1007/s11192-012-0767-9>.
- Zitt, Michel. 2005. Facing diversity of science: A challenge for bibliometric indicators. *Measurement: Interdisciplinary Research and Perspectives* 3: 38–49.