Scientific Performance Assessments Through a Gender Lens: a Case Study on Evaluation and Selection Practices in Academia

Mathias Wullum Nielsen

Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University, Denmark/mwn@stanford.edu

Abstract

The focus on *excellence* and *quality assurance* in the academy has spawned a significant increase in the use of bibliometric measures in performance assessments of individual researchers. This article investigates the organizational consequences of this development through a *gender lens*. Based on a qualitative case study of evaluation and selection practices at a Danish university, a number of potential gender biases related to the use of bibliometric performance measures are identified. By taking as default the research preferences, approaches and career paths of a succesful group of predominantly male scholars, evaluators using bibliometrics risk disadvantaging candidates diverging from the norm with implications for gender stratification. Despite these potential biases, bibliometric measures come to function as technologies supporting a managerial narrative of the gender-blind organization. They adhere to the prevailing ethos of the academic meritocracy by standardizing the criteria for organizational advancement and ensuring transparency and accountability in the selection process. While bibliometric tools in this sense may lead to the recruitment of scientists with a strong CV and track record, they may at the same time prevent many talented researchers diverging from the norm from being recognized and succeed as academics.

Keywords: gender and science, bibliometric indicators, research performance, research management, qualitative methods

Introduction

As observed by Simmel (1950: 412; originally written in 1903), the economic rationalization of modern society "has filled the days of so many with weighing, calculating with numerical determinations, with a reduction of qualitative values to quantitative ones". This development has obliterated many important qualitative differences, as certain social forms have become taken for granted as useful means for weighing and valuing elements of the social world, while diluting relationships and things that do not assimilate to these particular forms (Espeland and Stevens, 1998). Simmel's renowned sociological insight can be usefully extended to the recent decades' national and international New Public Management-driven reforms of higher education and

research (henceforth HER). These reforms have paved the way for new procedures for determining accountability and quality assessment directed at evaluating and controlling organizational performance.

Governments have been introducing market structures into the science systems for more than two decades. Institutional research budgets increasingly depend on performance-based funding; consequently, the competition between research organizations, groups and individuals is expanding rapidly (Gläser and Laudel, 2007: 109). At the global level, the OECD has played an important role in promoting the policy paradigm of the 'knowledge-based economy', while the World Bank's 2009 report on the "challenge of establishing world-class universities" has placed even further rhetorical emphasis on issues of 'high performance' and 'market-type competition' in the governance of HER institutions (Jessop, 2008: 26; Salmi, 2009). The question of global competitiveness also constitutes a key element in the EU policy discourses on research and development. This is seen in the Lisbon Strategy's (old and new) objective of making the EU the most competitive knowledge-based economy in the world (European Commission, 2000; 2005). Likewise, the more recent *Europe 2020* vision statement calls for *smart growth* by strengthening "research performance, promoting innovation and knowledge transfer throughout the union" (European Commission, 2010: 9).

The institutionalization of research evaluation constitutes a crucial component in this overall development (Whitley, 2007: 5), which has also spawned a significant expansion in the use of bibliometric indicators and metrics in the assessment of the performance of individual researchers (Weingart, 2005; Van den Brink et al., 2013).

Bibliometric indicators are advanced analytical tools used to assess scientific productivity, visibility and impact. Apart from their scholarly purposes, these tools are frequently employed by managers and politicians to organize competition among research institutions and boost the performance of individuals, groups, departments and faculties (Weingart, 2005; Addis and Brouns, 2004). As observed by Espeland and Sauder (2007: 2–3), the introduction of such standardized measures "can initiate sweeping changes in status systems, work relations, and the reproduction of inequality". Similarly, Weingart (2005: 127) argues that "not only the behaviour of individuals but that of organizations may be affected by bibliometric measures in ways that are clearly unintended". Thus, given their potentially profound organizational consequences, bibliometric indicators deserve further consideration.

This article investigates the constitutive effects related to the use of bibliometric indicators and metrics in the evaluation of research performance – through a *gender lens*. The *gender lens* enriches the study by focusing attention on the differential impact of practices of individual performance assessments on women and men. More specifically, this approach questions the taken-for-granted assumptions underpinning the existing organizational structures and practices and explores how some styles and forms of work become privileged in shaping the distribution of opportunities and rewards while others do not (Bailyn, 2011).

I raise three main questions: (a) how is scientific performance assessed in the recruitment and promotion of academic researchers? (b) How do quantitative metrics and indicators influence this process? (c) What are the potential gender consequences related to the use quantitative metrics and indicators in this process? The article draws on findings from a qualitative case study on evaluation and selection processes in the recruitment of senior research staff at a Danish university. This case study was based on qualitative interviews with 24 department heads and a document analysis of 44 assessment reports from appointments for associate professorships.

Scholars have already raised concerns about the gender effects of the proliferation of individual performance measures in university settings (more on this below). However, no studies have elucidated the gendered implications of how such measures are put into practice in the day-to-day activities of managers and research evaluators. The key contribution thus lies in the attempt to make visible the practice level of scientific performance assessments and its potential implications for gender stratification. Since academic recruitment practices are often treated with a high degree of

confidentiality and sensitivity (Van Den Brink, 2010), the study also provides unique opportunity to gain new insights into an otherwise closed realm of evaluation and selection.

The article proceeds as First, I briefly describe the existing literature on gender consequences of the emerging evaluation regime and reflect on the relevance of the study. Second, I outline the selected theoretical perspectives and empirical insights, which may aid in the analysis. Third, I touch upon methodology and present the empirical scope. Fourth, I present the analysis, and fifth I discuss the main findings and conclude..

Gender and new regimes of evaluation

Numerous authors have already pointed out the potential downsides to the emerging performance management regime in HER, which have been argued to narrow the approaches that researchers employ and the span of areas in which they engage (Guena and Martin, 2003; Rafols et al., 2011). Whitley (2007: 10), for instance, observes that when "evaluations become more important for both researchers and employers, the costs of pursuing deviant strategies increase, and pressures to demonstrate how one's work contributes to dominant disciplinary goals will grow". According to him, this pressure is strongest for researchers in temporary positions "who need to show the merits for their research as assessed by current disciplinary priorities and standards in order to gain employment" (Whitley, 2007: 10). Morley (2003, IX) raises a more fundamental concern, asking: "if quality assurance is about standards and conformance, what place is there for difference and diversity?" In her view, the growing emphasis on quantitative assessments of scholarly output intensifies organizational demands for *prescriptive performance* within established regimes of logic and reinforces gendered power relations in academia (Morley, 2003: 48).

A specific strand of literature focuses on the gender consequences of this development. As observed by Thomas and Davies (2003), the integration of seemingly gender-neutral quantitative performance metrics into managerial practices can be viewed as instrumental in identifying and challenging the subtle processes of discrimination and nepotism in academia. Scholars have, for instance, suggested that NPM-driven managerial approaches strengthening the transparency and accountability of academic organizations may allow research active women to display their merits and claim their right to promotion (see, e.g. Morley, 2005; Luke, 1997). As noted by Thomas and Davies (2003), however, these approaches may simultaneously intensify employee workloads and promote a competitive and individualistic research culture, creating stronger tensions between the responsibilities of work and family life and leading to 'chilly climates' for researchers with a preference for more supportive and collegial working styles. Knights and Richards (2003) add another perspective to the discussion by arguing that the success criteria for the existing NPM-driven audit regime are slanted in favour of male career patterns. In line with the pioneer work of feminist organizational scholar Joan Acker (1990), the main argument underpinning these concerns is that transparent and standardized evaluation and appointment criteria, when taking as default the organizational behavior of dominant groups, will not necessarily counter gender inequality.

Despite a growing scholarly interest in investigating the stratifying outcomes of scientific performance assessments, there is little research on how such measures are put into practice in the day-to-day activities of research institutions. As pointed out by Van den Brink et al. (2013: 181), the existing research on indicators of scientific performance "rarely pays attention to the implementation process, power processes and context". The authors try to fill this gap by employing a 'practice perspective' aiming to describe the use of such measures in the managerial activities of promotion and selection. Very similar to this approach, the study at hand takes a step towards a more in-depth understanding of how the resources and ideas introduced by bibliometric indicators influence the evaluative practices of department heads and evaluators, thus entailing potential gender consequences. By 'gender consequences', I refer here to the unequal career outcomes for male and female academics.

As returned to below, the article's main argument is that whereas bibliometric measures cannot not *per se* be considered discriminative in their features, they are at risk of reinforcing an evaluative culture that disadvantages scholars diverging from the norm, i.e. the research behaviour and career paths of a successful group of predominantly male academics. One should of course not underestimate the heterogeneity of gender roles at play in academic organizations. However, despite the many different ways of being a man and woman in academia, such measures may indirectly be gendered in their stratifying outcomes. Indeed, other factors such as class, sexuality and ethnicity may also operate to influence academics' conformity to, or deviance from, the prevailing image of the ideal career path or track record. Gender, in other words, merely represents one of several intersecting social categories influencing the career outcomes of academics. Consequently, whereas this study limits its focus to gendered aspects of scientific performance measures, an underlying objective will be to employ the gender lens to raise broader questions about the potentially stratifying outcomes of how academic work is evaluated.

Analytical reflections

Following West and Zimmerman (1987), this study conceives gender as something organizational actors 'do' rather than something they 'have'. Gender roles and categories are expressions of socially acquired behaviours and attributes produced and reproduced over time.

Clearly, this analytical approach challenges the idea of gender as a unitary conception structured around the male/female dichotomy. Recognizing the performative dimension of gender does, however, not necessarily imply that one should dismiss 'women' and 'men' as collective analytical categories. As Gunnarsson (2011: 32) rightly observes, it is possible to acknowledge abstract concepts "such as 'women' and 'men' as qualitatively different from lived reality [and] use them effectively without any expectation that they will correspond to this lived reality in any clear-cut sense". Indeed, such categories have been highlighted as instrumental to analysing the material and institutional conditions and structures forming and perpetuating gender inequalities (Fraser, 1995; Gunnarsson, 2011).

Feminist science studies have played an important part in teasing out the different forms of gender-based oppression at play in the academy. Pioneer studies in this tradition remind us how gendered norms and stereotypes operate to influence scientific approaches and interpretations (see e.g. Haraway, 1989). Further, unspoken notions about the ideal scientist and the scientific enterprise in general have been shown to clash with expectations about women and their roles in society (Schiebinger 1999: 69). The main argument underpinning this branch of scholarship is that seemingly impartial and objective scientific practices are shaped by implicitly gendered cultural assumptions privileging certain scientific perspectives and certain ways of being or becoming a scientist. Haraway's (1997) figuration of the scientist as a 'modest witness' constitutes an illustrative example in this regard:

> …the modest witness is the legitimate and authorized ventriloquist of the object world adding nothing from his mere opinions, from his biasing embodiment. And so he is endowed with the remarkable power to establish facts. He bears witness: he is objective; he guarantees the clarity and purity of objects (…) His narratives have a magical power – they lose all trace of their history as stories, (…) as contestable representations, or as constructed documents in their potent capacity to define the facts. (Haraway, 1997: 24)

What Haraway is proposing here is that the prevailing narrative of academic science as a 'culture of no culture' (Traweek, 1988: 1) and the scientist as a disembodied (male) truth-seeker, facilitates a detachment of scientific judgment from its socially situated vantage point. Scientists ascribing to this narrative, in other words, risk becoming blind to the cultural apparatus of historically sedimented and sometimes implicitly gendered ideas and assumptions influencing their judgments.

As advocated in this article, one could contend that this risk not only pertains to the construction of scientific facts but also to academic managers' and evaluators' quantitatively driven judgments

of scientific merits. What I am hinting at here is that the employment of seemingly objective measures in a biased system may operate to reinforce existing biases (Feller, 2003). If university managers, for instance, when recruiting senior research staff, fail to take into account the implicit structural and cultural obstacles encountering many women, the use of seemingly gender-neutral performance measures end up privileging traditional 'male' career patterns, since crucial factors such as career breaks, domestic responsibilities, research time and non-traditional publication behaviour are left out of the bibliometric equation. Wendy Espeland's work (Espeland and Stevens, 1998; Espeland and Sauder, 2007) with the concept of commensuration provides important insights into the social dynamics related to this particular problem.

Commensuration

According to Espeland and Stevens (1998: 314), commensuration can be viewed as a fundamental process in social life, "which transforms different qualities into a common metric". The sociological investigation of this process is important, because it "changes the terms of what can be talked about, how we value, and how we treat what we value" (Espeland and Stevens, 1998: 315).

In a 2007 article, Espeland and Sauder draws attention to how university rankings rest on commensuration. A fundamental premise underlying this research is the assumption that processes of commensuration in certain situations and contexts can become so deeply institutionalized and taken for granted that they contribute to forming the things and relationships they are developed to measure (Espeland and Stevens, 1998: 329). This is because organizational processes of commensuration inevitably produce various forms of *reactivity*[1] and change how people make sense of everyday situations (Espeland and Sauder, 2007: 10–11). A university ranking (or bibliometric performance indicator), for instance, can make it easier for organizational representatives to ignore qualitative characteristics and nuances that are not expressed in a particular metric. In addition, they can construct new relationships between objects and entities by transforming distinctive qualities into a common comparable metric. This leads to new hierarchical relationships between ranked universities or, as in the case of research performance metrics, allegedly objective comparisons between researchers with very different research propensities, career paths and publication behaviour. Further investigations of how such metrics are employed by research managers and evaluators (e.g. assessment committees) requires a clear conceptual understanding of their methodological applications and limitations. Drawing on Latour's and Woolgar's (1986) concept of *modalities*¸ Gläser and Laudel (2007) provide exactly that.

Amateur bibliometrics and modalities

Gläser and Laudel (2007: 117) use the term *amateur bibliometrics* to describe the "practice of producing bibliometric analyses of an evaluative character by actors with little or no professional background in the field and with little or no knowledge or regard for the modalities involved". Here, modalities refer to the 'modifying statements' employed by interested parties to weaken or make more solid the applicability of scientific findings. In this study, the concept is used to account for the qualifying statements that limit the methodological applications of bibliometric indicators to "specified conditions and ways of use" (Gläser & Laudel, 2007: 117). These modalities constitute a crucial element in the evaluative practices related to the assessment of scientific performance, because bibliometric indicators, like any other scientific method, rely on a set of "assumptions about applicability and proper procedure" (Gläser and Laudel, 2007: 117). Since most department heads and committee evaluators have only modest or no scholarly training in bibliometrics, a particular analytical focus on their regard for the modalities involved seems highly relevant. My own use of the modalities concept in the analysis of assessment reports and interviews with department heads, however, goes beyond the ideas introduced by Gläser and Laudel (2007). I place particular emphasis on the subtle gender dynamics embedded in, or emerging from, the evaluative use of bibliometrics, hence adding a new dimension to their approach. In addressing these modalities, I draw heavily on contributions from the bibliometric literature. Yet, before I turn

to discuss this strand of scholarship, it is important to account for field-specific and disciplinary variations in how research activities are structured and organized across academic fields and disciplines, which points to the relevance of sociologist Richard Whitley's (1984) work on *The Intellectual and Social Organization of the Sciences* (henceforth ISOS).

Task uncertainty and mutual dependence

In ISOS, Whitley (1984) develops two useful analytical dimensions for understanding the social organization of the sciences. These dimensions aid to the present study's analysis of how evaluative practices vary across disciplines and fields. According to Whitley (1984: 120), scientific fields vary in their "need to adhere to particular standards of competence and criteria of significance in order to reward important reputations for contributions". He labels this the dimension of *mutual dependence*. More specifically, this notion refers to the relative dependence of a field on knowledge produced in other fields in order to make significant scientific contributions in its own field, but also the extent to which scientists are expected to explicate how their contributions connect to the work of other scientists (Fry, 2004). Moreover, disciplinary fields differ in terms of level of *task uncertainty,* which relates to their compliance with widely accepted work procedures, standardized methods, problem definitions, theoretical goals and their ability to produce visible and replicable research results (Felt and Stöckelová, 2009; Whitley, 1984).

As several scholars note, the social sciences and humanities (henceforth SSH) are characterized by wider variations in paradigms, epistemic cultures, scientific communication practices and perceptions of excellence and quality than the natural and health sciences (henceforth NHS) (Felt and Stöckelová, 2009; Lamont, 2009; Moed et al. 2002). It is therefore also reasonable to contend that most SSH disciplines represent a lower degree of mutual dependence and a higher level of task uncertainty than what is the case in the NHS disciplines. These characteristics have the following bibliometric implications: First, the comprehensive variations in communication media within SSH limits the relevance of employing bibliometric measures, such as citation counts, journal impact factors and h-indices in these fields, since the existing bibliometric databases (e.g. Thomas Reuters' *Web of Science* [WOS]), lack systematic coverage of anthology articles, conference proceedings and monographs. Second, the themes and topics in the SSH literature are sometimes more locally anchored than is the case with the NHS literature, and scholars within these fields therefore also publish more frequently in non-English journals. This feature also speaks against the use of bibliometric measures, because WOS and SSCI have great limitations with respect to language and geographical coverage (Archambault and Gagné, 2004).

Modalities and gendered outcomes of scientific performance metrics

Journal rankings: Journal rankings and impact factors[2] have a number of frequently overlooked modalities (see e.g. Fleck, 2013). Especially, the use of such measures as proxies of publication quality imposes strong biases. Seglen (1997) illustrates how merely 15 per cent of a typical journal's scholarly papers receive more than 50 per cent of its overall citations. Publishing in a highly ranked journal does, in other words, not guarantee scholarly impact, because most of the citations accrued by top journals normally adhere to a limited number of papers (see also Christenson and Sigelman, 1985). It should be mentioned, however, that this bias in the assessment of research merits can be overcome by accounting for publication-based citation rates (article impact) (Moed et al., 2002; Weingart, 2005). As documented in the empirical analysis, this appears to be common practice at most NHS departments, whereas the situation in the social sciences is different, as citation counts are less prevalent.

Scholars focusing on the social sciences have already documented clear differences in the average impact factor scores and journal ratings of the publication outlets in which male and female academics publish their work (Brooks et al., 2014; Davenport and Snyder, 1995; Hunter and Leahey, 2010; McElhinny et al., 2003). In comparison, studies investigating gender differences in the performance of NHS researchers find no consider-

able variation concerning average journal impact factors (see e.g. Bordons et al., 2003; Mauleon and Bordons, 2006). These field-specific variations should be interpreted in view of strong differences in task uncertainty and mutual dependence across the NHS and SSH. As mentioned earlier, the SSH fields, despite many within-group differences, are characterised by wider variations and struggles between paradigms, epistemic cultures, scientific communication practices, regional and international research traditions and perceptions of excellence and quality than the NHS. And if women are overrepresented among the scholars engaged in 'non-mainstream' approaches and topics in the SSH, a reliance on journal rankings and impact factors may entail indirect biases in recruitment and selection processes.

Current research focusing on the social sciences provides some evidence supporting this assumption. Several studies point to noteworthy gender variations in methodologies and epistemological frameworks, with women gravitating towards constructivist styles and qualitative approaches and men towards positivist styles and quantitative approaches (see e.g. Breuning et al. 2005; Mallard et al. Plowman and Smith 2011). If SSH evaluators rely on journal rankings and impact factors as proxies of scientific merit, such gender variations may produce unequal career outcomes, since qualitative methods and constructivist epistemological styles are less prevalent in the most highly regarded social science journals (Bennett et al., 2003; Donovan, 2007; Macdonald and Kam, 2007; Svensson, 2006; Willmott, 2011).

A similar concern could be raised with respect to topic selection. Dolado et al. (2005) map variations in sub-field preferences among researchers in 50 internationally top-ranked economics departments and find women's representation to be highest in areas pertaining to lower-status journals such as Health, Education, Welfare, Labour and Demographic Economics and Economic History (Ritzberger, 2008). Similarly, Light (2013) carves out ten specialization clusters in the sociological literature, of which women are overrepresented in areas such as gender-race-sexualities, family-demography-youth, and medical-mental-health-aging, with a lower likelihood of being published in prestigious journals.

Citation counts: Although performance assessments based on citations rates per paper involve a number of methodological shortcomings,[3] this measure – from a gender perspective – appears to be the least problematic. The existing research on gender differences in citation rates provides inconclusive and ambiguous results, which may be due to strong institutional variations in citation patterns dependent on scientific discipline, geographic location and even gender composition in the field. While a number of studies have found a citation bias in favour of men (e.g. Aksnes et al., 2011; Maliniak et al., 2013; Lariviere et al,. 2013), most of the existing research finds women to be equally or in some cases even more cited than men (e.g. Long, 1992) (for a literature review, see Nielsen, 2016a). A recent study compared the research impact of 3,923 female and male researchers at Aarhus University; and in line with most of the existing literature found only trivial differences in the field normalized citation rates and relative shares of men and women contributing to the top 10% most cited articles internationally (Nielsen, 2016a). Matters, however, look quite different if we turn to the *h*-index.

H-index: Basically, the *h*-index (Hirsch, 2005) is a proxy for research performance developed to capture both publication rates and citation impact in a common metric. This metric combines the number of articles published by a researcher with the number of citations received by these papers and provides an estimate of the highest number of papers that have each received the same number of citations. This means that a researcher with an *h*-number of 6 has published six papers, of which each paper has received at least six citations (García-Pérez, 2009).

A number of modalities must be taken into account when using the *h-index* to assess research merits. First, the *h-index* is slanted in favour of researchers who publish in sub-fields with high citation frequencies[4]. Second, the *h-index* is highly correlated with research output (number of publications), and in this sense heavily depends on scientific age (i.e. active years as a researcher) and gender (more on this below), since a researcher's pool of scholarly papers and the citations that each paper receives increase over time (Kelly and Jennions, 2006). Third, the *h*-index privileges

individuals publishing with multiple co-authors. Collaborative authors obviously have higher research outputs than single authors and have more colleagues to cite their collaborative publications (García-Pérez, 2009; Kelly and Jennions, 2006). According to Van Raan (1998: 427), collaboration also "implies a considerable broadening of the audiences around the authors, enhanced by more intensive networking".

Some of the modalities presented have been argued to disadvantage women (García-Pérez, 2009; Symonds et al., 2006). Several studies for instance detect a gender bias in research collaboration in favour of male researchers (e.g. Abramo et al., 2013; Bozeman and Corley, 2004; Kyvik and Teigen, 1996; Prpic, 2002)[5]. A recent study of researchers at Aarhus University also indicates that women on average publish more single-authored papers and have a slightly lower propensity for international research collaborations (Nielsen, 2016a). But the most crucial of these modalities from the perspective of gender equality concerns the *h*-index's high correlation with publication output. Historically, women have been found to publish fewer scholarly papers than men (Cole & Zuckerman, 1984); and while this gender difference have been shown to decrease over time (Xie and Shaumann, 2003), the most recent literature documents a continuous bias in favour of men (Mairesse and Pezzoni, 2015). The sociology of science offers a variety of explanations to this so-called productivity puzzle. Some scholars argue that childbearing lowers women's scholarly output during the early career stages (Kyvik and Teigen, 1996; Mairesse and Pezzoni, 2015). Others relate the gender gap to systemic causes such as variations in employment rank and access to funding (Xie and Shauman, 1998), degree of disciplinary specialization (Leahey, 2006), differences in collaborative patterns, and time dedicated to research and other tasks (Taylor et al., 2006).

Data and methods

This study unfolds within the framework of a larger research project focusing on the structural challenges to gender equality at Aarhus University. The case-study approach provides unique opportunity to relate the qualitative findings of this article to relevant quantitative patterns identified in complementary papers revolving around gender and scientific performance in the same organization (see Nielsen, 2015 2016a). Aarhus University is a public institution of HER with more than 40,000 students and approximately 11,000 employees. The university employs around 4,000 researchers (including approximately 1500 PhDs) and comprises a broad range of disciplinary domains and fields of research.

Assessment reports

Since bibliometric measures are used for a variety of goals and purposes in academic organizations, one of the main challenges has been to cover the diversity of the existing institutional evaluative procedures and practices in an adequate manner.

With considerable effort – and some luck – I have been able to gain access to 44 systematically selected assessment reports[6] from recruitments for associate professorships at Aarhus University in the period 2005–2012. The Danish associate professorship title – as is the case in the Anglo-American promotion model – is a tenured position, normally following three to five years of temporary or fixed-term employment as either postdoc or assistant professor. The American tenure track system is rarely employed in Denmark and positions at the postdoctoral level may be filled for a maximum period of four to five years at the same institution. The associate professorship appointment is therefore often considered to be the first real 'safe haven' for young Danish scholars attempting to establish a research career in the academy. When this study was carried out, women comprised 17 per cent of the full professors, 33 per cent of the associate professors, 41 per cent of the postdocs/assistant professors and 51 per cent of the PhDs at Aarhus University.

The assessment reports provide unique opportunity to investigate to what extent and how bibliometric measures and indicators are employed when appointment committees assess applicants' scientific merits. More specifically, I have focused on the weight ascribed to these measures in the judgment of the research candidates' existing research curriculum and aimed to clarify which kinds of research behaviour are rewarded when

Table 1: Selection Criteria – Assessment Reports:

- Only publicly announced vacancies for associate professorships or equivalent positions within the period 2005–2012 were considered relevant - Only vacancies with at least three applicants were considered relevant - Only vacancies with both male and female applicants were considered relevant - Twelve reports from each of Aarhus University's four faculties were requested (Arts, Business and Social Science, Health and Science & Technology)

using bibliometric tools. Criteria for the selection of assessment reports are specified in Table 1.[7]

Specifications on the research disciplines represented in this documentary material (i.e. the disciplines in which the relevant vacancies have been announced) and year of appointment are available in the Appendix, Table A. Moreover, analytical displays illustrating the different types of scientific performance measures employed in the evaluation of applicants in the documentary material are enclosed in the Appendix, Table B, C, D and E. I also draw on the official procedural documents guiding the evaluative work of assessment committees. These documents have been used for two purposes: a) to inform the qualitative interviews with the department heads; and b) to obtain information on the official status ascribed to scientific performance measures by assessment committees and department heads.

Before proceeding, it is relevant to note that assessment committees at Aarhus University do not prioritize among the candidates or select the final nominee for positions at associate and full professor level. Rather, their task is limited to identifying the qualified applicants for a given vacancy. The department heads typically make the final appointment decision in correspondence with the faculty deans. Indeed, existing research on recruitment and selection practices at Aarhus University, documents that department heads play an important part in identifying potentially relevant candidates for research vacancies at the university prior to the actual recruitment process (Nielsen, 2016b). A closer look at the evaluative practices of this group of managers thus seems particularly informative for the purposes of this study.

Interviews

I have conducted qualitative interviews with 24 of the 27 department heads at Aarhus University. I have used an open-ended interview approach, mixing conversation and structured questions to collect data. More specifically, I have asked questions in three broad areas of academic management: *evaluation criteria related to academic appointments* (how is research performance assessed in this process and what types of performance are rewarded?), *use of bibliometric measures and indicators in mid-level research management* (how are they used and for what purposes?), *the introduction of new models of performance assessment* (have they affected existing procedures of performance measurement and management and how?). Further, I have asked the interviewees to consider whether and how gender-related issues influence the selection practices in their departments and how they account for such issues in the recruitment process. Twenty of the 24 interviews have been conducted face-to-face, typically in the department head's office. The rest have been carried out over the phone. The interviews – lasting between 30 and 90 minutes – are analysed using Nvivo software. Analytical displays highlighting the different metrics employed by department heads in the evaluation of scholarly merits are enclosed in Table F, G, H and I in the Appendix.

Obviously, scientific performance merely represents one of several selection criteria in academic recruitment and selection. When asked to describe the central characteristics of the ideal research candidate, the interviewees emphasized other evaluative criteria such as preceding experience with research management, ability to obtain funding, and teaching qualifications. Yet research performance was highlighted as the core criteria in most interviews, and as noted by several

interviewees, a strong track record with respect to scientific outcomes is considered a premise for obtaining funding and managing research projects.

Analysis

The presentation of the analysis is divided into three main parts. First, I elaborate on the different bibliometric approaches employed by assessment committees in the identification of qualified applicants for associate professor positions. This part of the analysis draws exclusively on findings from the assessment reports. Second, I discuss the modalities and potential adverse gender consequences related to the use of different types of bibliometric measures. This part is mainly based on interviews with department heads and insights from the procedural documents guiding evaluative practices at Aarhus University. Moreover, this part includes selected examples from the assessment reports to illustrate how bibliometric measures are employed in the evaluation process.

Bibliometric measures at Aarhus University

At Aarhus University, despite many within-group variations, the NHS disciplines use bibliometric measures and indicators more frequently than the SSH.[8] As illustrated in Table D and E in the Appendix, publication counts, citation counts, h-indices, counts of first-author and senior-author publications, journal impact scores and measures of increasing or decreasing productivity trends over time are all frequently used bibliometric tools in assessment committee's evaluations of research merits in the NHS. The emphasis on bibliometric measures tends to be particularly advanced in disciplines such as biochemistry, biology, computer science and biomedicine.

In addition to quantitative measures, assessment committees in the NHS also make judgments concerning the quality and prestige related to publication channels. In computer science, for instance, where conference proceedings play an important role, evaluators make clear distinctions between contributions to 'high standing', 'medium standing' and 'low standing' conferences, but also between 'mediocre' and 'leading' journals. This type of distinction is present in most of the assessment reports, and several assessment committees also highlight authors' contributions to top journals, such as *Nature*, *Science* and *The Lancet*, as pivotal indicators of scientific merit.

In comparison, assessment reports in the SSH (see Table B and C in the Appendix) are considerably longer and characterized by more in-depth, qualitative evaluations of applicants' scholarly contributions. This is especially the case within the humanities and in the less quantitative parts of the social sciences (e.g. sociology and business communication). While the less quantitative SSH disciplines also account for publications in national and international peer-reviewed journals and articles in anthologies and monographs, bibliometric measures such as citation counts and *h*-indices receive no emphasis whatsoever. Instead, the evaluators usually provide comprehensive and in-depth assessments elaborating on the theoretical and analytical approaches employed by applicants and discussing how they contribute to the existing research literature in the field.

The more modest emphasis on quantitative measures of research output and past research achievements in the qualitative parts of the SSH may open space for a greater focus on the actual content of an applicants' work and his/her future research potential. This could serve as one of the explanations why more women succeed in obtaining permanent research positions in these disciplines than in the more quantitative parts of the social sciences and the NHS. This is discussed further below.

The predominantly quantitative areas of the social sciences (e.g. business administration and economics), in addition to publication counts, also make use of discipline-specific journal rankings and, in some cases, journal impact factors. This indicates a higher degree of mutual dependence in these disciplines than is the case in the humanities and the more qualitative part of the social sciences. As pointed out by Fry (2004), disciplines with a high level of mutual dependence must agree on what is considered a valid contribution to the research literature in their field, and thus have more tightly controlled research cultures and communication systems.

As discussed earlier, the potential gender bias associated with the evaluative use of journal impact factors and journal ratings is mainly an issue in the SSH .The following discussion therefore limits its focus to these fields.

Bibliometric indicators – modalities and potential gender consequences

The emphasis on journal rankings and impact factors in the SSH, according to the interview material, is strongest in disciplines weighted in favour of quantitative approaches (e.g. economics, business administration, political science). In these areas, the department heads frequently use terms such as 'internationally recognized journals', 'top journals', 'highly ranked field journals', 'general field journals', 'mediocre journals' and 'unknown journals' to make distinctions on the quality and prestige of applicants' scholarly contributions. Reflecting on the question of how research merits are evaluated, a department head explains:

> The easiest thing for the appointment committee to assess – because it's well-documented – that's the articles written by the applicants. And yes, this varies a lot, because people are different and their views on what counts as quality vary a lot. But they need to make an evaluation of the candidates' track records. What's the quality of their papers? And some of that can be assessed relatively easily, right? … In some areas it's relevant to look at the impact factor, in other areas it's relevant to look at what kind of journals this is, right? (Social Sciences)

As illustrated in the quote, the accreditation of knowledge through rankings and impact factors is attractive for SSH evaluators, because it offers seemingly objective tools for overcoming internal disagreements on what counts as quality. By transforming distinctive qualities into common comparable metrics, such tools establish easily applicable hierarchical relationships between journals or scholars, hereby making it easier for evaluators to ignore qualitative nuances and differences that are not expressed in the metrics (Espeland and Sauder, 2007).

As touched upon earlier, however, relying on journal impact factors and rankings is not an adequate strategy for evaluating a researcher's impact on the existing literature. Further women have been shown to more frequently engage in topics and approaches less prevalent in the most highly regarded social science journals, and on average score lower on journal ratings and impact factors than men. When asked about whether any gender differences exist in the research interests, topics and approaches of his staff, a department head from the social sciences comments:

> No, I don't think so. And this is because of these women, who are highly aware of the risk of researchers ending up in the trenches. Well, I don't know whether it's a 'trench', but… There seems to be a tendency among Danish women that some research areas are more popular than others. Children and education, for instance – that kind of research seems to attract more women … And what has been very explicit among the older women in this department is the importance of career supervision to make sure that young women don't all end up doing research on children. There's nothing wrong with that, but there seems to be a tendency. (Social Sciences)

As observed by Whitley (2007, 10), the standardization of research objectives and quality criteria "means that the diversity of intellectual goals and approaches … decline over time". This is illustrated quite well in this example, where some research topics are considered less likely to reach top journals and hereby become taken-for-granted as "trenches" in the field. Journal ratings and impact factors, in this sense, not only endow evaluators with analytical tools to establish hierarchical relationships between scholarly publications; they at the same time contribute to form the content of knowledge production by changing how people make sense of their day-to-day activities, thus producing career obstacles for (women) researchers diverging from the prevailing approaches and research topics. While such processes of commensuration risk advantaging an already successful group of predominantly men scholars, more research is needed to fully estimate their stratifying gender effects.

Output, impact and the temporal morphology of the academic career

When asking the department heads to describe the central characteristics of the 'employable' candidate for tenured positions, most respond by emphasizing the weight and volume of applicants' existing research as the most central criterion. This is also evident in the assessment reports, where research merit in most cases is evaluated before anything else. In the following, attention is focused specifically on three different approaches to evaluating research output: citation counts, h-indices and measurement of performance over time.

Citation counts: The evaluative practice of counting and comparing the citations obtained by candidates in their papers (i.e. research impact) is widely employed in the natural sciences and parts of the health sciences, and most of the department heads describe this measurement as an important supplement to publication counts (See Table H and I, Appendix). Further, citation counts are stated to be an important element in the faculty guidelines in the natural sciences defining the basic criteria for assessing scientific merits. More specifically, the document states that the qualified candidate should have many papers [professor positions] or some papers [associate professor positions] with good citation numbers (dependent on scientific field and number of years after PhD) (Aarhus University, 2013). As discussed earlier, paper-based citation counts — in a gender perspective — can be viewed as the least problematic of the typical scientific performance metrics employed in individual research assessments, since women and men tend to be cited at similar rates at Aarhus University.

Yet adverse gender consequences may be at play when citation counts are employed to capture a researcher's cumulative scholarly impact over time. As illustrated in the bar-charts and figure text in Figure 1[9] (taken from an assessment report in the natural sciences), this way of counting citations privileges candidates with high publication rates and many co-authored publications (more on this below). The multiplicity of contextual factors circumscribing researchers' cumulative citation rates are here rendered invisible in favour of seemingly objective comparisons of past performance. By taking as default the track records and career patterns of a successful group of predominantly male scholars, such evaluative practices, in other words, risk disadvantaging candidates diverging from the norm with potential implications for gender stratification.

H-index: The *h*-index, as mentioned earlier, is a proxy developed to capture both publication rates and citation impact in a common metric. This metric has been heavily promoted by journals such as *Science* and *Nature* (Symonds et

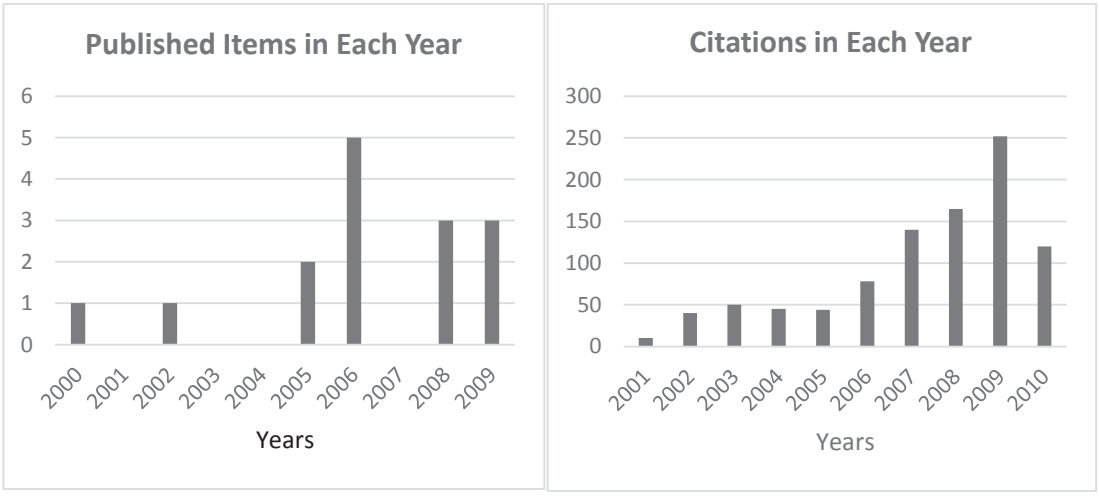

Figure 1: Measuring performance rates over time (assessment report, NHS).
Over the past 5 years: 13 papers published with a significant impact (>6)
Over the past 10 years: 9 papers published with a significant impact (>6)
Last year he was cited 256 times and 118 this year to-date (across articles)

al., 2006) and is widely employed by assessment committees and department heads in the health and natural sciences at Aarhus University (See Table H and I, Appendix). The aforementioned guidelines used for assessing scientific merit in the natural sciences also emphasize the *h*-index as a central performance criterion. More specifically, this document states that candidates for full professorships are expected to have an *excellent h*-index, while applicants for associate professorships should have a *good h*-index (dependent on scientific field and number of years after PhD) (Aarhus University, 2013). As discussed earlier, a number of modalities must be taken into account when using the *h*-index in research assessments; and if left unnoticed, some of these modalities have been argued to put women at a disadvantage. To briefly restate, the *h*-index privileges researchers publishing in sub-fields with high citation frequencies, and researchers who have many co-authorships and many scholarly publications.

When reflecting on their own evaluative practices of recruitment, many of the department heads also emphasize some of these modalities. Referring to the methodological shortcomings of *h*-indices, a department head from the natural sciences comments:

> You really need to be careful. You can't compare an experimental scientist with a theoretical scientist in this field. There is an inherent difference in the h-index and in between. (Natural Sciences)

This quote touches upon the first of the modalities discussed above, which is an issue raised by several of the department heads. The second modality, relating to the strong correlation between research output and the h-index, is also a crucial methodological concern among many interviewees. Three interviewees note:

> Obviously, citations depend on the age of the candidate right? There is also the *h*-index, which measures impact and so on, but it is very dependent on age, and of course you look at that right? If the *h*-index is very different from that of other candidates, then you start to wonder; because this means that this person doesn't get cited very often (Natural Sciences)

> In my opinion, the emphasis on the *h*-index is far too strong. My experience tells me that it takes many years for a researcher to reach high citations. This is why I think we should avoid emphasizing this issue too much. But of course, they're expected to have reached a certain number of citations and publications, but … I don't expect the *h*-index to be a double-digit number, but it has to be over seven or eight, depending on age. (Natural Sciences)

> We take this with a pinch of salt, and I clearly understand the connection between things here. We don't say "okay this man with an *h*-index of 35 is better, he's better than the one with an index of 17". This isn't necessarily the case, but he's older, and he's been publishing more articles, and that leads to more citations for a normal employee. And then you can ask… do you want a young employee or an old, experienced one? That's another question. Sometimes you might need one with 'more hair on the chest'. But in another vacancy, you might need a young, dynamic researcher with 'ants in the pants'. (Natural Sciences)

As the quotes indicate, department heads call for circumspection when interpreting the *h*-indices of applicants. However, while issues of scientific age tend to be very important to the interviewees, gender is not mentioned by any of the NHS department heads when reflecting on the limitations of the *h*-index. The quotes also illustrate that although *h*-indices are considered less useful in the evaluation of younger researchers, applicants for tenured positions are still expected to have reached a certain *h* when applying for positions at the associate professor-level. The third quote stands out as particularly interesting. It tends to be structured around a temporal norm of the typical (male) career trajectory, distinguishing between two idealized images of the successful scientist – the highly experienced older candidate and the hyper-productive younger combatant. As returned to below, one might question how this temporal norm complies with the (un)usual career paths of many female researchers.

As noted earlier, examining the applicants' performance trends over time is a frequently employed evaluative practice within the NHS and parts of the social sciences. Comments similar to those below are also present in several of the 48 reports.

> The applicant has a scientific production, which documents experience with different techniques within cardiovascular research. However, the production seems stagnant at a modest level, and an increase is definitely needed. (Health Sciences)

> Given that the applicant completed a PhD in 1997, the number of peer-reviewed journal articles may be a little disappointing, but both the quality and quantity still seem to be merely adequate for a tenured position. (Social Sciences)

Similarly, the department heads clearly emphasize the crucial importance of considering performance trends when evaluating scientific merits. Two interviewees from the more quantitative fields in the social sciences explain:

> These things influence our evaluations, and they also influence me when I recruit staff. I understand that a researcher needs time to settle, but an increase in productivity is important, because it indicates success. (Social Sciences).

> Well, the quality is something that you… it's relatively easy, right? You can look at what kind of journal it is, and there will be several rankings of journals. And then we look at trends. Does it look like the person's productivity is rising or falling? (Social Sciences)

The strong evaluative emphasis on research output and performance trends over time may entail indirect gender consequences. As noted by Espeland and Stevens (2008) "numbers often help constitute the things they measure by directing attention, persuading, and creating new categories for apprehending the world"; and although many of the department heads clearly emphasize that bibliometric indices are only one of many strategies for evaluating scientific merit, it is reasonable to contend that these metrics promote an evaluative culture privileging past achievements over future potential, thus rewarding traditional career paths and publication patterns.

Many years ago, March and Simon (1958: 165) introduced the term 'uncertainty absorption' to describe social processes in which "inferences are drawn from a body of evidence, and the inferences instead of the evidence itself, are then communicated" (cf. Espeland and Stevens, 2008). The analysis of the assessment reports reveals a similar pattern. Issues of quality and content tend to receive less emphasis in the evaluative practices of the NHS and the more quantitative parts of the social sciences than in the qualitative SSH disciplines. Evaluators may therefore fail to account for the nuances of existing contributions and future potentials in more than just quantitative ways. Reflecting on how the proliferation of bibliometric indicators has affected the evaluative practices, an interviewee with many years of leadership experience in the natural sciences comments:

> This is something which has emerged within the last 5–10 years. Before, you couldn't measure … well before Google scholar, it wasn't possible for us to measure citations within this discipline. So, what we did 10–15 years ago if we wanted to hire a person was to look at their CVs and then classify each paper and say "these five are good conferences, these are mediocre and the rest is shit". That was how you measured the researcher – by looking at conferences and journals. And if you go 10 years back, you looked more at each of the papers. Well, there were fewer papers and fewer applicants. It was less demanding at that point. And if you go 20–30 years back – at that point the assessment committees read your papers. They don't do that anymore! (Natural Sciences)

Although most research evaluators and department heads would probably hesitate to support the last part of this quote, the development described by the interviewee constitutes an illustrative example of how the accessibility of advanced bibliometric tools has spawned a certain type of 'uncertainty absorption'. Evaluative inferences tend to be drawn from the output of quantitative performance metrics rather than the actual research that these metrics are developed to measure.

While this development may harm both women and men, the existing literature indicate that a disproportionate number of younger female researchers facing "the dilemma of synchronizing the often-conflicting demands of three clocks: the biological clock, the career clock (as in timetables for tenure), and a spouse's career clock" (Sonnert and Holton, 1996: 70) may experience crucial challenges conforming to this new evaluative regime.

To be sure, not all female researchers become mothers or have male spouses, and numerous other factors also contribute to the gender gap in publication rates. Nonetheless, family commitments, in the large perspective, tend to play a crucial role at this career level.

A 2008 study focusing on Danish academics, for instance, reveal that women in top research positions have fewer children than their male colleagues (Verner, 2008). These findings echo the work of Mason and Goulden (2004) showing that American male researchers who father children early in their career are 38 per cent more likely to achieve tenure than women in the same situation.

Scholars have pointed to conflicts between family responsibilities and the gendered structure of the academic career path, when explaining these patterns. Bailyn (2004), for instance, asserts that the first stages of the academic trajectory, which often coincide with the point in life when many academics start families, are characterized by enormous pressure for quick success. According to her, this pressure may put a disproportionate share of female academics at a slight disadvantage due to gender variations in family-career tensions.

The existing literature on the question of gender, family characteristics and publication rates is, however, inconclusive. Kyvik (1990) and Kyvik and Teigen (1996) provide evidence of a negative impact of motherhood on scientific publication rates, while Fox and Faver (1985) and Fox (2005) find women with young children to publish at higher rates than women with no children or school-age children. Similarly, Cole and Zuckerman (1991) fail to identify negative effects of motherhood on scientific publication rates.

In this case, it is relevant to note that the studies presented above limit their focus to researchers who have already obtained tenure track or tenured positions. This means that they do not account for the initial processes of selection and exclusion in the research system. In other words, they are merely comparing the impact of motherhood on the pool of researchers who have already 'survived' one or several steps of academic promotion. In this sense, Fox and Faver (1985), Fox (2005) and Cole and Zuckerman (1991) are not providing sufficient evidence for rejecting the hypothesis that early-career researchers experience cumulative disadvantages due to motherhood and domestic obligations.

Moreover, the studies by Fox and Faver (1985), Fox (2005) and Cole and Zuckerman (1991) all draw on American data, while Kyvik (1991) and Kyvik and Teigen (1996) provide insights into the Norwegian situation. This means that structural and socio-cultural differences between the countries may contribute to explaining some of the disparities in the outlined results. Seierstad and Healy (2012) highlight the Scandinavian countries' family-friendly policies and their inherent affirmation of women as the main carers of the family as a distinctive structural feature limiting the advancement of female researchers in this particular socio-cultural context. The Scandinavian countries, for instance, all provide significantly longer periods of paid maternity and parental leave than the US, meaning that Scandinavian women (and some men) will have longer periods of research inactivity early in their careers than their American colleagues (Kyvik and Teigen, 1996).

Another crucial concern in this regard relates to the question of whether gender differences in the weekly allocation of time for research activities are taken into account when employing bibliometric measures to evaluate scientific achievements. In a 2012 survey aiming to assess the psychological work climate at Aarhus University, all researchers were asked to estimate how they, on average, distributed their work time across different types of tasks and activities. As illustrated in Table 2, women's self-estimated weekly allocation of time for research was lower than that of their men colleagues across all scientific ranks with a women to man ratio of 0.91:1 (25.0/27.4) for PhDs, 0.78:1 (16.9/21.7) for postdocs, 0.84:1 for associate professors (10.5/12.5) and 0.86:1 (11.6/13.5) for full professors. The difference was particularly noteworthy for faculty in postdoc level positions, which as illustrated in the existing literature is a career stage characterized by high demands for quick success in terms of scientific achievements (see e.g. Müller, 2014).

One way of interpreting these data could be that on average male researchers are better at administrating their time in terms of direct

Table 2: Weekly time spending (Psychological Work Place Assessment)

	PHD		POSTDOC		ASSOC. PROF.		FULL PROF.	
Tasks	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**
Keep yourself updated	6.0	5.2	6.7	5.8	7.8	6.7	9.7	7.7
Research	25.0	27.4	16.9	21.7	10.5	12.5	11.6	13.5
Research administration	1.8	1.7	4.0	3.1	5.1	4.3	5.7	5.2
Other administration	0.7	0.9	1.7	1.5	4.0	3.6	4.9	3.9
Teaching/ preparation	4.4	4.3	7.3	6.6	7.9	7.3	8.4	9.0
Supervision	0.8	0.9	2.4	2.9	4.0	3.6	5.0	4.8
Dissemination	0.6	0.5	1.0	0.7	1.2	1.6	1.9	2.0
Other (e.g. consultancy)	3.7	3.2	4.9	2.1	6.8	6.4	6.1	4.4
Total	**43.0**	**44.1**	**43.9**	**44.4**	**47.2**	**46.0**	**53.4**	**50.5**

N = Grade D: F(382), M(350); Grade C: F(182), M(262); Grade B: F(273), M(591); Grade A: F(47), M(212). Source: Human Resources, AU.

research outcomes. Another interpretation, however, might be that women take on broader 'invisible' organizational responsibilities (Fletcher, 2001), thus contributing to the functioning of the university in ways which are left unnoticed in bibliometrically based assessments of research trends over time. In other words, it is crucial that evaluators and managers take such issues into account when evaluating and comparing academics' scientific achievements.

In the last part of my interviews, I specifically asked the interviewees to consider whether and how gender-related issues influenced the recruitment and selection practices in their departments. As expected, several of them responded in words quite similar to the statements below:

> We DO NOT look at whether the applicant is a man or a woman when we recruit. The selection has nothing to do with that. We look at the qualifications … We don't think, "we want a man" or "we want a woman". (Health Sciences)

> Well honestly, I have to admit that I don't look at it [gender]. What we want is the best – the person that fits the picture the best. Sometimes it's a girl, sometimes it is a boy … or women and men. Well, it's not an issue, and I know that some people think it should be … Actually, our gender balance is OK, and this is also because we have many talented girls, but you're right… many of them leave before they go very far …When I look at recruitment, this isn't something… We want talented girls, and we have that, but it's not something we… We primarily look at qualifications. (Social Sciences)

As illustrated, the department heads consider the existing recruitment and selection processes to be gender-neutral and clearly emphasize that only the very best candidates will get through. I coin this "the narrative of the gender-blind organization". Interestingly, the performance measures discussed in this article come to function as managerial technologies supporting this narrative. They adhere to the prevailing ethos of the academic meritocracy by standardizing the criteria for organizational advancement and ensuring transparency and accountability in the selection processes, thus reducing the space for the practice of direct discrimination and nepotism. In

view of the persistent gender inequalities in the academic system, however, this narrative can be viewed as problematic. It disregards the potential differential impact of scientific performance measures on women and men, and reinforces prevailing disparities in resources and opportunities. By adhering to the gender-blind narrative, the department heads may overcome accusations of any type of direct discrimination or nepotism, but they may at the same time unintendedly prevent many talented researchers with 'unusual' research interests and career trajectories from succeeding as academics.

During the interview with the first of the two department heads quoted above, I chose to follow up on the interviewee's reflections on the *gender-neutral* nature of the existing recruitment practices. I clarified that the main purpose of my interviews had not been to reveal issues of direct nepotism and discrimination against women. Rather, I aimed to obtain a better understanding of the unintended differential impact of existing recruitment and selection practices on women and men. Interestingly, this made the interviewee open up and recognize one of the central biases related to these practices:

> Well, I agree. Clearly, things will be distorted, because we look at the h-index and things like that. Because it depends on your publication productivity and women simply haven't had the time to write the necessary number of publications. In view of that, I agree. Already at that point, we distort things. But this isn't intentional, right? (Health Sciences)

The interviewee's response constitutes an illustrative example of how social processes of commensuration related to the use of bibliometric measures decontextualize knowledge and render some aspects of organizational life invisible by shaping and constraining the cognition and behaviour of research evaluators (Espeland and Stevens, 2008).

In this regard, it is relevant to note that most interviewees, when asked directly about whether and how they compensate for parental and maternity leave periods in their selection practices, clearly emphasize that such breaks in a career are always taken into account and that having children would never disadvantage an applicant. Interestingly, however, the issue of a potential relation between CV gaps and parental leave periods is not raised once in any of the 44 assessment reports. Likewise, very few of the interviewees account for the subsequent periods of increased domestic responsibilities related to starting a family, which may limit the productivity of many (women) researchers with small children. Instead, they adhere to the idea that researchers, when 'back in business', should be measured against the same objective criteria as anyone else. This idea is epitomized most clearly in the quotes below.

> It's evident in the CV when children are in the picture. What's interesting is whether they're capable of getting back on track … One of the persons we hired had two maternity leave periods, and there were also a couple of years without any scientific production – and that made good sense. If people have shown their worth and shown that they're capable of getting back on track… In that case, I would have no worries about hiring. (Natural Sciences)

> One or 2 years of absence due to parental leave isn't that important. Well, for us it's all about getting the best candidate, and if that's a woman with children… well, that's fantastic and impressive, but that's not what we look at. We look at their qualifications regarding research, teaching and so on. (Health Sciences)

As illustrated, the prevailing understandings of research potential, capabilities and scientific worth tend to be strongly intermingled with issues of past performance and research output, which may contribute to explaining why more female than male researchers continue to face challenges in obtaining permanent recruitment at Aarhus University.

Conclusion

This study has attempted to make visible aspects of gender biases in how quantitative metrics are put in to practice in scientific performance assessments. Drawing on assessment reports and interviews with department heads, the study illustrates how bibliometrics, when used at the individual

level, can serve to perpetuate existing gender inequalities in academia by providing indisputable and easily measurable proxies for merit that decontextualizes scientific achievements and transforms different qualities into common metrics.

The study adopts a 'practice perspective' teasing out the modalities employed by research evaluators when using such metrics; modalities which often tend to disregard variations in scientific styles, career trajectories and particular life-circumstances, hereby disadvantaging candidates diverging from the (male) norm.

Consider, for instance, the accreditation of knowledge through journal ratings and impact factors in the social sciences. As illustrated in the interview material, this form of assessment is attractive among evaluators, because it offers seemingly objective tools for overcoming internal disagreements on what counts as quality.

In a gender perspective, however, the use of such proxies may entail adverse consequences. A disproportionate share of female researchers have been proven to engage in topics, styles and methodologies with a lower likelihood of being published in prestigious journals, and women, on average, score lower on impact factor scores and journal ratings than their men colleagues.

At the same time, the use of such proxies risk narrowing the diversity of intellectual goals and approaches at play in departments, hereby producing career obstacles for (women) researchers diverging from prevailing approaches and topics.

Further, publication counts, *h*-indices and assessments of cumulative citation and publication rates tend to be employed with little regard for the non-traditional circumstances and career obstacles facing many female researchers. Indeed, the use of such measures often come at the expense of longer, more in-depth and content-focused evaluations of scientific merits. They privilege past achievements over future potential, hereby putting scholars that do not comply with the temporal norms of the typical (male) career trajectory at a slight advantage.

Following Haraway (1989), one could contend that these measures facilitate a detachment of evaluative judgment from it socially situated vantage point. Despite their potential gender effects, they come to function as technologies supporting a managerial narrative of the gender-blind organization. They standardize the criteria for organizational advancement and ensure transparency and accountability in the selection process. By relying on such measures, managers may overcome accusations of any type of direct discrimination or nepotism. However, while this approach may lead to the recruitment of scientists with a strong CV and track record, it may at the same time prevent many talented researchers with interest and career trajectories diverging from the norm from being recognized for their contributions and succeed as academics.

One of the strategies that evaluators can adopt to overcome this potential bias is to move beyond the prevailing narrative of the gender blind organization. As clearly illustrated in this study, many department heads are eager to employ bibliometric measures in fair and objective ways, which implies being sensitive and responsive to gendered variations in research interests, approaches, and career developments. At first hand, this may seem counterintuitive, since gender categorization has been proven to implicitly influence academic assessments and evaluations (Valian, 1999). However, the academic appoint process is not double-blinded by nature (the first names of candidates are always given to evaluators), meaning that unconscious gender bias will operate irrespective of whether we explicitly account for gendered variations in the selection process or not. In line with the recently announced *Leiden manifesto* offering principles to guide the use of research metrics in evaluations of performance (Hicks et al., 2015), my suggestion therefore is to always supplement the use of quantitative proxies for merit with in-depth and systematic qualitative considerations about variations in expertise, experience, activities and career progression along gendered lines; even when comparing large numbers of researchers. Such an approach could help render visible some of the potential gender biases related to the use of quantitative performance metrics, hereby making academic recruitment and selection processes less gendered in their stratifying outcomes.

The potentially gendered aspects of quantitative performance assessments highlighted in this study leave ample room for further investigations. One question concerns the potential adverse gender effects related to the use of journal rankings and impact factor scores as proxies for quality in the SSH disciplines. Obtaining an in-depth understanding of this question, implies a more systematic mapping of gender variations in research interests, topics, approaches and methods. In this regard it is also crucial to account for variations across different stages of the academic career, since the publication behaviour of women (and men) leaving the academy at an early scientific age (i.e. years after PhD) may vary considerably from the publication behaviour of those who remain.

Likewise, a comparative study investigating the influence of family formation on the publication productivity of early career academics across varying socio-cultural contexts could provide much needed information adding further nuance to the scholarly debate over the existence of a 'motherhood penalty'.

References

Arhus University (2013) *Criteria for evaluating candidates for positions as full professor, associate professor, senior scientist, or senior advisor*. Available at: http://scitech.medarbejdere.au.dk/fileadmin/site_files/scitech.medarbejdere.au.dk/files/Evaluation_Criteria.pdf (accessed 2 March 2015).

Abramo G, D'Angelo CA and Caprasecca A (2009) The Contribution of Star Scientists to Overall Sex Differences in Research Productivity. *Scientometrics* 84 (3): 821–33.

Acker J (1990) Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations. *Gender & Society* 4 (2): 139–58.

Addis E and Brouns M (2004) *Synthesis Report on the Workshop in Gender and Excellence in the making.* Brussels: European Commission.

Aksnes, DW, Rorstad K, Piro F and Sivertsen G (2011) Are female researchers less cited? A large-scale study of Norwegian scientists. *Journal of the American Society for Information Science and Technology* 62(4): 628-636.

Archambault É and Gagné É V (2004) The use of bibliometrics in the social sciences and humanities. *Science-Metrix Final Report*. Available at: http://www.science-metrix.com/pdf/SM_2004_008_SSHRC_Bibliometrics_Social_Science.pdf (accessed 2 March 2017).

Bailyn L (2004) Time in careers: Careers in time. *Human Relations* 57(12): 1507-1521.

Bailyn L (2011) Redesigning work for gender equity and work-personal life integration. *Community, Work & family* 14(1): 97-112.

Bennett A, Barth A and Rutherford KR (2003) Do we preach what we practice? A survey of methods in political science journals and curricula. *Political Science and Politics* 36(03): 373-378.

Bordons MF, Morillo M, Fernandez T and Gomez I (2003) One step further in the production of bibliometric indicators at the micro level: Differences by gender and professional category of scientists. *Scientometrics* 57(2): 159-173.

Breuning M, Bredehoft J and Walton E (2005) Promise and Performance: An Evaluation of Journals in International Relations. *International Studies Perspectives*, 6(4): 447-461.

Brooks C, Fenton EM and Walker TJ (2014) Gender and the evaluation of research. *Research Policy* 43(6): 990-1001.

Bozeman B and Corley E (2004) Scientists' Collaboration Strategies: Implications for Scientific and Technical Human Capital. *Research Policy* 33: 599–616.

Christenson JA and Sigelman L (1985) Accrediting knowledge: Journal stature and citation impact in social science. *Social Science Quarterly 66*(4): 964-975.

Cole JR and Zuckerman H (1991) Marriage, motherhood, and research performance in science. In: Zuckerman H, Cole JR and Bruer JT (eds) *The Outer Circle: Women in the Scientific Community*. New York: W.W. Norton, pp. 157-170.

Davenport E and Snyder H (1995) Who cites women? Whom do women cite? An exploration of gender and scholarly citation in sociology. *Journal of Documentation* 51: 404–410.

Dolado JJ, Felgueroso F and Almunia M (2005) *Do men and women-economists choose the same research fields? Evidence from top-50 departments*. IZA Discussion Paper 1859. Available at: http://ssrn.com/abstract=866886 (accessed 2 March 2015).

Donovan C (2007) The hidden perils of citation counting for Australasian political science. *Australian Journal of Political Science* 42(4): 665-678.

Espeland W and Sauder M (2007) Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology* 113(1): 1-40.

Espeland W and Stevens ML (1998) Commensuration as a social process. *Annual Review of Sociology* 24: 313-343.

Espeland W and Stevens ML (2008) The sociology of quantification. *European Journal of Sociology* 49(3): 401-436.

European Commission (2000) *Presidency conclusion: Lisbon European Council, March 23-24*. Brussels: European Commission.

European Commission (2005) *Communication to the Spring European Council. Working together for growth and jobs: A new start for the Lisbon Strategy.* Brussels: European Commission.

European Commission (2010) *Europe 2020: A European strategy for smart, sustainable and inclusive growth. Communication from the commission*. Brussels: European Commission.

Evans HK and Bucy EP (2010) The representation of women in publication: An analysis of political communication and the international journal of press/politics. *Political Science and Politics* 43: 295-301.

Feller I (2003) Measurement of scientific performance and gender bias. In: Addis E and Brouns M (eds) *Synthesis Report on the Workshop in Gender and Excellence in the making.* Brussels: European Commission, pp. 35-41.

Felt U and Stöckelová T (2009) Modes of ordering and boundaries that matter in academic knowledge production. In: Felt U (ed) *Knowing and Living in Academic Research.* Prague: Institute of Sociology of the Academy of Sciences of the Czech Republic, pp. 41-124.

Fleck C (2013) The impact factor fetishism. *European Journal of Sociology* 54(2): 327-356.

Fox MF and Faver CA (1985) Men, women, and publication productivity: Patterns among social work academics. *Sociological Quarterly* 26: 537-549.

Fox MF (2005) Gender, family characteristics, and publication productivity among scientists. *Social Studies of Science* 35(1): 131-150.

Fletcher JK (2001) *Disappearing acts: Gender, power, and relational practice at work*. MA, Boston: MIT Press.

Fraser N (1995) False antithesis. In: Benhabib S, Butler J, Cornell D and Fraser N (eds) *Feminist contentions*. New York, NY: Routledge, pp. 59-74.

Fry J (2004) The cultural shaping of ICTs within academic fields: Corpus-based linguistics as a case study. *Literary and Linguistic Computing* 19(3): 303-319.

Garcia-Pérez MA (2009) A multidimensional extension to Hirsch's H-index. *Scientometrics* 81(3): 779-785.

Guena A and Martin BR (2003) University research evaluation and funding: An international comparison. *Minerva* 41(4): 277-304.

Gunnarsson L (2011) A defence of the category "women"*. Feminist theory* 12(1): 23-37.

Gläser J and Laudel G (2007) The social construction of bibliometric evaluations. In: Whitley R and Gläser J (eds) *The Changing Governance of The Sciences: The Advent of Research Evaluation Systems*. Doodrecht: Springer, pp. 101-123.

Haraway DJ (1989) *Primate visions: Gender, race, and nature in the world of modern science*. Hove: Psychology Press.

Haraway DJ (1997) *Modest– Witness@ Second– Millennium. FemaleMan– Meets– OncoMouse: Feminism and Technoscience*. Hove: Psychology Press.

Hicks D, Wouters P, Waltman L, de Rijcke S. and Rafols I (2015) Bibliometrics: The Leiden Manifesto for research metrics. *Nature* 520: 429-31.

Hirsch JE (2005) An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences* 102: 16569-16572.

Hunter L and Leahey E (2010) Parenting and research productivity: New evidence and methods. *Social Studies of Science* 40(3): 433–451.

Jessop B (2008) A cultural political economy of competitiveness and its implications for higher education. In: Jessop B, Fairclough N, Wodak R (eds) *Education and the Knowledge-Based Econonomy in Europe*. Rotterdam: Sense Publishers, pp. 11-39.

Kelly CD and Jennions MD (2006) The h index and career assessment by numbers. *TRENDS in Ecology and Evolution* 21: 167–170.

Knights D and Richards W (2003) Sex discrimination in UK academia. *Gender, Work & Organization* 10(2): 213-238.

Kyvik S (1990) Motherhood and scientific productivity. *Social Studies of Science* 20(1): 149-160.

Kyvik S and Teigen M (1996) Child care, research collaboration, and gender differences in scientific productivity. *Science, Technology, & Human Values* 21(1): 54-71.

Lamont M (2009) *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.

Lariviere V, NI C, Gingras Y, Cronin B and Sugimoto CR (2013) Global gender disparities in science. *Nature* 504(7479): 211-213.

Latour B and Woolgar S (1986) *Laboratory Life: The Construction of Scientific Facts.* Princeton, NJ: Princeton University Press.

Leahey E (2006) Gender differences in productivity: Research specialization as a missing link. *Gender & Society* 20(6): 754-780.

Light R (2013) Gender inequality and the structure of occupational Identity: The case of elite sociological publication. *Networks, Work, and Inequality* 24: 239-268.

Long JS (1992) Measures of sex-differences in scientific productivity. *Social Forces* 71(1): 159–178.

Luke C (1997) Quality assurance and women in higher education. *Higher education* 33(4), 433-451.

Macdonald S & Kam J (2007) Ring a Ring o'Roses: Quality Journals and Gamesmanship in Management Studies. *Journal of Management Studies* 44(4): 640-655.

Mairesse J and Pezzoni M (2015). Does gender affect scientific productivity? A critical review of the empirical evidence and a panel data econometric analysis for French physicists. *Revue economique* 66(1):65-113.

Mallard G, Lamont M and Guetzkow J (2009) Fairness as appropriateness: Managing epistemological differences in peer review. *Science, Technology and Human Values* 34(5): 573-606.

Maliniak D, Powers RM and Barbara F (2013) *The gender citation gap* in *international relations*. *International Organization* 67(4):889-922.

Mason MA and Goulden M (2004) Do babies matter (Part II)? Closing the baby gap. *Academe* 90 (6): 11-16.

March JG and Simon HA (1958) *Organizations.* New York: Wiley.

Mauleón E and Bordons M (2006) Productivity, impact and publication habits by gender in the area of materials science. *Scientometrics* 66(1): 199–218.

McElhinny B, Hols M, Holtzkener J, Unger JS and Hicks C (2003) Gender, publication and citation in sociolinguistics and linguistic anthropology: the construction of a scholarly canon. *Language in Society* 32: 299–328.

Moed HF, Luwel M and Nederhof AJ (2002) Towards research performance in the humanities. *Library Trends* 50(3): 498–520.

Morley L (2003) *Power and quality in higher education*. Berkshire: Open University Press.

Morley L (2005) The micropolitics of quality. *Critical Quarterly* 47(1-2): 83-95.

Müller R (2014) Postdoctoral life scientists and supervision work in the contemporary university: A case study of changes in the cultural norms of science. *Minerva* 52(3), 329-349.

Nielsen MW (2015) Gender consequences of a national performance-based funding model: new pieces in an old puzzle. *Studies in Higher Education*. Epub ahead of print 12 September 2015 DOI: 10.1080/03075079.2015.1075197.

Nielsen MW (2016a) Gender inequality and research performance: moving beyond individual-meritocratic explanations of academic advancement. *Studies in Higher Education* 41(11): 2044-2060.

Nielsen MW (2016b) Limits to meritocracy? Gender in academic recruitment and selection processes. *Science & Public Policy* 43 (3): 386-399.

Plowman DA and Smith AD (2011) The gendering of organizational research methods: Evidence of gender patterns in qualitative research. *Qualitative Research in Organizations and Management: An International Journal* 6(1): 64-82.

Prpic K (2002) Gender and productivity differentials in science. *Scientometrics* 55(1): 27–58.

Rafols I, Leydesdorff L, O'Hare A, Nightingale P and Stirling A (2012) How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research Policy* 41(7): 1262-1282.

Ritzberger K (2008) A ranking of journals in economics and related fields. *German Economic Review* 9(4), 402-430.

Salmi J (2009) *The Challenge of Establishing World Class Universities*. Washington DC: The World Bank.

Schiebinger L (1999) *Has Feminism Changed Science?* Cambridge, MA: Harvard University Press.

Seglen P (1997) Why the impact factor of journals should not be used for evaluating research. *British Medical Journey* 314: 497-502.

Seierstad C and Healy G (2012) Women's equality in the Scandinavian academy: A distant dream? *Work Employment & Society* 26(2): 296-313.

Simmel G (1903) The metropolis and modern life (transl. Wolff K). In: Weinstein D (ed) (1950). *The Sociology of Georg Simmel*. New York: Free Press, pp. 409-424.

Symonds MRE, Gemell NJ, Braisher TL, Gorringe KL, Elgar MA (2006) Gender differences in publication output: Towards an unbiased metric of research performance. *PLoS ONE* 1 (1): doi:10.1371/journal.pone.0000127

Sonnert G and Holton G (1996) Career patterns of women and men in the sciences. *American Scientist* 84(1): 63-71.

Svensson G (2006) The paradoxnoia of top journal (s) in marketing. *European Journal of Marketing'* 40(11/12): 1153-1168.

Thomas R and Davies A (2003) Gender and New Public Management: Reconstituting academic subjectivities. *Gender, Work and Organization* 9(4): 372-397.

Traweek S (2009) *Beamtimes and lifetimes: The world of high energy physicists*. Boston, MA: Harvard University Press.

Valian V (1999) The cognitive bases of gender bias. *Brooklyn Law Review* 65:1037-1061.

Van den Brink M (2010) *Behind the scenes of science: Gender practices in the recruitment and selection of professors in the Netherlands*. Amsterdam: Amsterdam University Press.

Van Den Brink M, Fruytier B and Thunnissen M (2013) Talent management in academia: Performance systems and HRM policies. *Human Resource Management Journal* 23(2): 180-195.

Van Raan AFJ (1998) The influence of international collaboration on the impact of research results: some simple mathematical considerations concerning the role of self-citations. *Scientometrics* 42(3): 423-428.

Verner M (2008) *Gender differences in rank within the academic profession: The case of Denmark. Paper presented at the IZA* Workshop on Research in Economics: Rewards, Evaluation and Funding. Bonn, Germany. Available at: http://conference.iza.org/conference_files/ReinEco2008/verner_m907.pdf (accessed 12 May 2016).

Weingart P (2005) Impact of bibliometrics upon the science system: Inadvertent consequences. *Scientometrics* 62(1): 117-131.

Whitley R (1984) *The Intellectual and Social Organization of the Sciences.* Oxford: Oxford University Press.

Whitley R (2007) Changing governance of the public sciences: The consequences of establishing research evaluation systems for knowledge production in different countries and scientific fields. In: Whitley R, Gläser J (eds) *The Changing Governance of the Sciences.* Doodrecht: Springer, pp. 3-27.

West C and Zimmerman DH (1987) Doing gender. *Gender & society* 1(2): 125-151.

Willmott H (2011) Journal list fetishism and the perversion of scholarship: reactivity and the ABS list. *Organization* 18(4): 429-442.

Xie Y & Shauman KA (2003) *Women in science: Career processes and outcomes.* Cambridge, MA: Harvard University Press.

NOTES

1 Espeland and Sauder (2007: 6) define *reactivity* as a process in which "individuals alter their behavior in reaction to being evaluated, observed or measured".

2 Basically, the journal impact factor is an indicator developed to measure the importance or influence of a specific journal within a scientific field for a given time period by providing an estimate of the average frequency with which scholarly papers are cited in the journal during the preceding 2 or 5 years.

3 For thorough discussion of these shortcomings, see Gläser and Laudel (2007) and Weingart (2005).

4 Whereas field normalized citation scores can be used to overcome this bias in paper-based citation counts, the h-index does not normalize citations across fields.

5 Whereas fractionalization can be used to overcome this bias in publication counts, the *h*-index is not based on a fractionalized count of publications.

6 The human resources department has been very helpful in this regard by opening up their recruitment records. Due to the time-demanding process of anonymizing the reports, the administration limited my access to 48 systematically selected assessment reports. The selection of the final reports has taken place on the basis of a dataset provided by the human resources department, including statistical information regarding all recruitments for research positions from 2005–2013.

7 Originally, the assessment reports were obtained for a broader case study on evaluative practices in academic recruitment and promotion, which is also reflected in the selection criteria. 12 assessment reports were requested from each of the university's four scientific areas. However, due to complications in identifying the reports at the human resource department I ended up with altogether 44 documents. 12 from Science & Technology, 13 from Health, 11 from Business and Social Science, and 8 from ARTS (humanities).

8 NHS refers here to disciplines pertaining to departments in AU's faculties of Science & Technology and Health. SSH refers to disciplines pertaining to departments in AU's faculties of Business and Social Science and ARTS (humanities).

9 Figures and text are reconstructions of photo-copied figures appearing in one of the assessment reports.

Appendix:

Table A: Overview of the appointment reports

Science/Technology	Health Sciences	Business/Social sciences	Humanities
Molecular biology (2009)	Public health (2008)	Business administration (2012)	Theology (2009)
Computer science (2012)	Biomedicine (2012)	Political Science (2010)	Literature (2011)
Geoscience (2012)	Clinical medicine (2012)	Business Communication (2012)	European studies (2011)
Mathematics (2011)	Biomedicine (2011)	Business administration (2011)	French (2006)
Chemistry (2011)	Public health (2011)	Law (2013)	Aesthetics and communication (2011)
Cross disciplinary (2010)	Cross disciplinary (2006)	Business administration (2012)	Media science (2007)
Biology (2010)	Public health (2006)	Management (2007)	History (2011)
Molecular biology (2005)	Biomedicine (2007)	Business communication (2006)	History (2011)
Mathematics (2007)	Clinical medicine (2011)	Economics and business (2007)	
Chemistry (2006)	Public health (2011)	Law (2010)	
Computer Science (2008)	Clinical medicine (2012)	Economics and business (2010)	
Environment. Science (2010)	Biomedicine (2007)		
	Public Health (2005)		

Table B: Humanities (appointment reports)

	Pub. count	***Citation count***	***h index***	***impact factor* (prestige)**	***Journal rankings* (over time)**	***Productivity trends***	***Co-authored/ solo-authored* (relative to productivity)**	***Scientific Age***
Theology	X							
Literature	X							
European Studies	X							
French	X							
Aesth. & Comm.	X						X	
Media Science								
History	X						X	X

Table C: Business/Social Sciences (appointment reports)

	Pub. count	***Citation count***	***h index***	***impact factor*** **(prestige)**	***Journal rankings*** **(over time)**	***Productivity trends***	***Co-authored/ solo-authored***	***Scientific Age*** **(relative to productivity)**
Business adm.	X			X	X		X	X
Political Science (Sociology)	X							
Business Com.	X				X			
Law								
Management	X				X			
Economics	X				X			X

Table D: Health (appointment reports)

	Pub. count	***Citation count***	***h index***	***impact factor*** **(prestige)**	***Journal rankings*** **(over time)**	***Productivity trends***	***Author sequence*** **(first/middle/ last)**	***Scientific Age*** **(relative to productivity)**
Public Health	X	X	X	X	X	X	X	
Biomedicine	X	X		X	X	X	X	
Clinical medicine	X			X	X	X	X	
Cross-disciplinary	X			X			X	

Table E: Science and Technology (appointment reports)

	Pub. count	***Citation count***	***h index***	***impact factor***	***Journal rankings* (prestige)**	***Productivity trends* (over time)**	***Author sequence* (first/middle/ last)**	***Scientific Age* (relative to productivity)**
Molecular biology	X	X	X	X	X	X	X	X
Computer science	X	X	X		X			X
Geoscience	X						X	X
Mathematics	X							
Chemistry	X					X		
Cross-disciplinary (Physics)	X							
Biology	X	X	X	X	X		X	
Environment. Science	X						X	

Table F: Humanities (interviews)

	Pub. count	***Citation count***	***h index***	***impact factor***	***Journal rankings* (prestige)**	***Productivity trends* (over time)**	***Author sequence* (first/middle/ last)**	***Scientific Age* (relative to productivity)**
Culture and Society	X				X			
Education	X							

Table G: Business/Social Sciences (interviews)

	Pub. count	***Citation count***	***h index***	***impact factor***	***Journal rankings* (prestige)**	***Productivity trends* (over time)**	***Author sequence* (first/middle/ last)**	***Scientific Age* (relative to productivity)**
Business administration	X			X	X			
Business communication	X				X			
Economics	X			X	X	X		
Political Science	X				X			
Interdisciplinary (Business/ Engineering)	X	X		X	X	X	X	X

Table H: Health (interviews)

	Pub. count	***Citation count***	***h index***	***impact factor***	***Journal rankings* (prestige)**	***Productivity trends* (over time)**	***Author sequence* (first/middle/ last)**	***Scientific Age* (relative to productivity)**
Public Health	X		X		X	X	X	
Biomedicine	X	X	X	X	X	X	X	
Clinical medicine	X	X	X	X		X	X	
Legal medicine	X							
Odontology	X		X					

Table I: Science and Technology (interviews)

	Pub. count	***Citation count***	***h index***	***impact factor***	***Journal rankings* (prestige) *(conferences)***	***Productivity trends* (over time*)***	***Author sequence* (first/middle/ last)**	***Scientific Age* (relative to productivity)**
Molecular biology	X	X	X	X		X		
Computer science	X	X	X		X			
Geoscience	X	X	X		X	X		X
Mathematics	X	X			X			
Chemistry	X	X	X					X
Biology	X	X	X	X	X			
Physics	X	X	X		X	X		
Engineering	X	X	X		X			
Agroecology	X		X	X				
Food science	X	X	X		X			
Animal science	X		X	X	X			X
Environment. Science	X		X	X	X			