Minerva
https://doi.org/10.1007/s11024-020-09419-0

Bibliometrics in Academic Recruitment: A Screening Tool Rather than a Game Changer

Ingvild Reymert[1]

© The Author(s) 2020

Abstract This paper investigates the use of metrics to recruit professors for academic positions. We analyzed confidential reports with candidate evaluations in economics, sociology, physics, and informatics at the University of Oslo between 2000 and 2017. These unique data enabled us to explore how metrics were applied in these evaluations in relation to other assessment criteria. Despite being important evaluation criteria, metrics were seldom the most salient criteria in candidate evaluations. Moreover, metrics were applied chiefly as a screening tool to decrease the number of eligible candidates and not as a replacement for peer review. Contrary to the literature suggesting an escalation of metrics, we foremost detected stable assessment practices with only a modestly increased reliance on metrics. In addition, the use of metrics proved strongly dependent on disciplines where the disciplines applied metrics corresponding to their evaluation cultures. These robust evaluation practices provide an empirical example of how core university processes are chiefly characterized by path-dependency mechanisms, and only moderately by isomorphism. Additionally, the disciplinary-dependent spread of metrics offers a theoretical illustration of how travelling standards such as metrics are not only diffused but rather translated to fit the local context, resulting in heterogeneity and context-dependent spread.

Keywords Bibliometrics · Research Quality · Academic Profession · Recruitment Processes · Candidate Evaluation · Peer Review

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11024-020-09419-0) contains supplementary material, which is available to authorized users.

✉ Ingvild Reymert
Ingvild.reymert@nifu.no

[1] NIFU - Nordic Institute for Studies in Innovation, Research and Education, PB 2815, Tøyen, 0653 Oslo, Norway

Published online: 10 September 2020

Introduction

Across the world, universities and organizations are converging with a growing set of universal standards (Brunsson and Jacobsson 2000; Drori et al. 2006). Faced with rationalization, universities have adopted the idea of the "world class" university and have developed into organizational actors or goal-oriented entities expected to be accountable to stakeholders (Krücken and Meier 2006). Standards and rankings represent the operationalizations of the world class university and universal quality standards (Ramirez and Meyer 2013) and are also applied to satisfy the increased accountability and auditing demands as rationalization is more easily achieved with fewer objectives (Brunsson and Sahlin-Andersson 2000). In addition, the increased use of standards has resulted in global convergence (Baert and Shipman 2005; Power 1999), for instance, how rankings and instruments of comparison in the higher education and science sector have spurred universities to remodel their organizational structures (Sauder and Espeland 2009; Paradeise and Thoenig 2015).

Bibliometrics (metrics), referring to the quantitative measures of scientific and scholarly publications, are examples of such global standards and have become increasingly important in the higher education and research field (Sivertsen 2017; Wilsdon et al. 2015). Metrics are widely perceived as indicators of research quality (Aksnes et al. 2019), and where research quality is a complex concept and hard to define, the numerical basis of metrics implies that they are countable, rankable and easily understood (Krücken and Meier 2006).

Metrics are not always welcomed by researchers. There are examples of metrics misuse and gaming, and scholars have criticized metrics for being more driven by data than by judgments (De Rijcke et al. 2016; Wilsdon et al. 2015). Both the Leiden Manifesto and the 2012 San Francisco Declaration on Research Assessment (DORA), the latter signed by thousands of organizations and individuals, have expressed concerns about the use of metrics on individuals. However, despite critiques and recommendations, studies have suggested that metrics are used to assess individual researchers and have gained importance in recruitment processes, which were traditionally based on tenured scholars' qualitative evaluations of the candidates' works (Hammarfelt and Rushforth 2017; Hicks et al. 2015; Stephan et al. 2017).

Increased reliance on metrics and change in the use of assessment criteria in recruitment processes is noteworthy as recruitments represent core university processes acquiring key university resources, scholars, and are one of the fundamental peer review processes in science (Langfeldt and Kyvik 2011). Peer review processes are also found to be highly stable, resisting reforms aiming to alter them (Musselin 2013). Accordingly, recruitment processes are processes where change is least expected. Stability in organizational processes is further supported by historic institutionalists describing organizations and their processes as stable and maintained by policy feedback effects (Pierson 1993; Thelen 1999). Despite arguing that organizations are changing, neo-institutionalists also specify that organizations often buffer core organizational processes or their day-to-day activities from change (Meyer and Rowan 1977). An increased reliance on metrics in these robust processes would thus

not have occurred without resistance and would imply that the propagation of metrics has even reached core university processes.

However, the literature on the use of metrics in recruitments is scarce and provides little evidence for the importance of metrics in relation to other quality criteria, or whether metrics have replaced or only supplemented traditional peer review. In addition, the literature covers only few disciplines, thus making generalizability to the whole science community problematic. This lack of research could partly be due to the confidential nature of recruitment processes. Decision-making in recruitment processes are mostly classified information, also for researchers aiming to analyze these evaluation processes. However, we were able to get access to confidential reports with candidate evaluations from professor recruitments in economics, sociology, physics and informatics at the University of Oslo between 2000 and 2017. These highly unique data have enabled us to go beyond the literature's more superficial description of metrics use and disentangle the dynamics inside recruitment processes accounting for how, and with which importance metrics are used in relation to other assessment criteria and more traditional peer review in recruitments. We thus address the following research questions:

> How have assessment criteria, more specifically the use of bibliometric indicators in academic recruitment, changed over time in different academic disciplines? And how can organizational theory explain the similarities and differences in the use of bibliometrics?

To answer these questions, we understand universities as a collection of disciplinaries in sub-units with their own concepts of research quality and their own evaluation cultures (Lamont 2009; Whitley 1984). In the following section we review prior studies on academic recruitment processes. We then develop three separate expectations derived from three different strands in organizational theory: *Stable evaluation cultures over time* based on a path-dependency perspective (Thelen 1999), *increased reliance on metrics in candidate evaluation over time* aligned with a neo-institutional perspective (DiMaggio and Powell 1983; Meyer and Rowan 1977), and *discipline-dependent change* building on a translation perspective (Czarniawska and Sevón 1996). Afterwards we present the research context, data and method before turning to the analysis. Finally, we discuss our data in relation to our expectations before summarizing the results in the conclusion.

Relevant Studies

Recruitment processes represent one of the fundamental peer review processes in academia where candidates are evaluated on multiple assessment criteria, such as their research output, teaching experience, language skills, administrative and leadership experience and social skills (Van den Brink and Benschop 2011; Herschberg et al. 2018; Levander et al. 2019). Of these, a candidate's research output has been identified as the most salient criterion (Van den Brink and Benschop 2011). However, research quality is not a straightforward concept, but rather a fluid, negotiated,

socially constructed and multifaceted concept covering both originality, plausibility, solidity, and academic and societal relevance (Polanyi 2000; Lamont 2009; Langfeldt et al. 2020). Montgomery and Hemlin (1991) moreover observed how different aspects of research quality were emphasized in different phases of a recruitment process, where candidates were acknowledged for their stringency and productivity in the early stages of the recruitment process, whereas the interview committee regarded originality and breadth of their work as more important.

Traditionally, recruitment processes were conducted by tenured academics qualitatively assessing the candidates' works (Fürst 1988; Musselin 2010). However, recent studies have suggested that metrics are increasingly applied in these candidate evaluations. Recruiters request the candidates' h-index, candidates boost their CVs with metrics, and recruitment committees favor lengthy publication records (Hicks et al. 2015; Stephan et al. 2017; Van den Brink and Benschop 2011), and in Norway metrics are used in recruitment processes despite governmental recommendations (Aagaard 2015). However, these studies only identify use of metrics and not how and with which importance metrics are applied, nor whether metrics have outperformed traditional peer review.

Two Swedish studies have, however, contributed with comprehensive accounts of metrics use in recruitment processes. Hylmö (2018) examined expert reports from academic recruitment in economics from 1989 to 2014 at four leading Swedish universities. The author found a shift from more traditional qualitative peer review evaluations towards a stronger reliance on publication numbers in top international journals. In addition, he observed how the length of evaluation reports had decreased, implying a stronger reliance on metrics rather than the more page consuming qualitative candidate evaluation. Similarly, Hammarfelt and Rushforth (2017) explored the use of metrics in recruitments in biomedicine and economics from 2005 to 2014 in Sweden with content analysis of committee reports. The authors found that evaluators possessed high knowledge of the metrics they used and how they cautiously guarded disciplinary norms when employing metrics. In addition, they found discipline-dependent use of metrics: whereas evaluators in biomedicine relied on journal impact factor and h-index, evaluators in economics placed emphasis on the candidates' number of publications in top journals (Hammarfelt and Rushforth 2017). However, whereas Hammarfelt and Rushforth (2017) did not consider the relative importance of metrics in relation to other assessment criteria, Hylmös' (2018) study only covered economics, making generalizations difficult considering the particular use of metrics in economics. Thus, the literature on the use of metrics in recruitment processes does not fully explain *how*, *with which importance* and *whether* their application supplements or replaces traditional peer review.

Analytical Framework

Universities are often regarded as loosely coupled organizations and as a collection of different academic disciplines. These disciplines are independent systems for knowledge production and validation with their own epistemological traditions, notions of research quality and peer review (Clark 1978; Lamont 2009; Whitley

1984). The disciplines educate and employ their own scholars, and at the University of Oslo as site for this study, recruitments are conducted within the disciplines and not at university level. Studies on recruitment processes have also addressed how disciplines evaluate candidates differently, for instance, how candidates' international experience and teaching experience were more valuable qualifications in the natural sciences than in social sciences (Herschberg et al. 2018; Levander et al. 2019). The use of metrics in research evaluations also strongly depends on the disciplines (Hammarfelt and Rushforth 2017; Wilsdon et al. 2015). Hence, we conceptualize the university as an organization with highly independent disciplines (or subunits) with recruitment processes possessing diverse evaluation cultures.

The academic disciplines may also be labeled *academic fields*, but to avoid confusion with organizational fields as an analytic concept developed by neo-institutional scholars (Meyer and Rowan 1977; DiMaggio and Powell 1983), we address sociology, economics, informatics and physics as disciplines. The disciplines operate in two organizational fields: a) *disciplinary* organizational fields consisting of scholars, departments, conferences, journals, norms and cultures, and b) in a broader *university or academic* organizational field including all disciplines in science.

In the analyses, we aim to detect change and stability in the university disciplines' evaluation practices. Universities are often perceived as stable organizations resistant towards 'new ways of doing things' (Colyvas and Powell 2006), and if they change, they change mostly incrementally or through organizational layering (Clark 1983). In addition, peer review has proven especially resistant towards change (Musselin 2013). This robustness corresponds with the historical institutionalist perspective which sees organizations as stable products of their context, originating from critical junctures (Thelen 1999). Historical development and context matter for organizational decision-making. Consequently, institutional structures render path dependency, feedback mechanisms (Pierson 1993, 2004) and lock-in effects (Sydow et al. 2009). Recruitment processes as core organizational activities may even be more stable as organizations often buffer them from change in the organizational field (Meyer and Rowan 1977). We thus expect recruitment processes to be sites where change is least expected, and where peer review evaluations are stable with persisting differences according to discipline, and not suddenly adopting metrics as new parameters of research quality.

However, universities and their organizational processes are not unchangeable, and across the world, universities are converging with a growing set of universal standards (Brunsson and Jacobsson 2000; Drori et al. 2006). This understanding stems from a neo-institutional view on organizations, which highlights organizational change and understands organization as open systems that in the wake of uncertainty adapt to rules and myths that are taken for granted in the surrounding environment in order to gain legitimacy (DiMaggio and Powell 1983; Meyer and Rowan 1977). Recruitment processes are associated with high levels of uncertainty where quality is hard to define, and candidates difficult to rank. Disciplines are moreover associated with different status, where economics, for instance, is recognized as a highly prestigious discipline and has a strong use of metrics (Fourcade et al. 2015). These uncertainties and status differences may trigger evaluators to mimic other colleagues' use of metrics. Appointments moreover need to be perceived as legitimate

(Scott and Davis 2007), and metrics are suited tools to consolidate candidate evaluations and offer a sense of objectivity as they are considered as more objective than qualitative research evaluations (Ramirez and Meyer 2013). In addition, whereas a committee's peer review could be disputed, a candidate's superior metrics could not. Hence, even though recruitment processes are sites where change is least expected, we would expect accountability and desire for legitimacy to stimulate increased use of metrics in recruitment processes.

Whereas the path dependency perspective suggests stability, and isomorphism suggests change, the translation perspective proposes that eventual change would be context dependent (Czarniawska and Sevón 1996; Wedlin and Sahlin 2017). Translation scholars have contested that ideas and standards are passively diffused, resulting in homogenization, as suggested by the isomorphic perspective (Czarniawska and Sevón 1996). Instead, they argue that organizations actively translate and contextualize travelling ideas to fit their local context through editing and translation processes (Sahlin-Andersson 1996). These editing processes may change both the form of the idea, its focus, content, and meaning (Wedlin and Sahlin 2017). This perspective brings the path dependency and isomorphism perspectives together and proposes that the different disciplines would adopt metrics differently in order to gain legitimacy and at the same time preserve their evaluation cultures. We thus expect sustained differences, but still change which reproduces the differences. The translation perspective is highly relevant for studying the use of metrics in recruitment processes, as both the evaluation cultures of the academic disciplines are robust, and the diffusion of metrics is strong. The translation perspective also illuminates the necessity to study how metrics are used in different disciplines, as we expect the disciplines to adopt metrics differently.

We thus develop three contesting expectations; Firstly, aligned with the path-dependency perspective (Thelen 1999) we expected stable and discipline-dependent evaluation cultures that are relatively unaffected by the general spread of metrics in science. Secondly, relying on the isomorphism perspective (DiMaggio and Powell 1983) we anticipate increasingly reliance on metrics in the candidate evaluations in all disciplines. Finally, building on translation theory (Czarniawska and Sevón 1996) we expect sustained disciplinary-differences as change will reproduce the differences.

Data and Method

The Research Context

This paper follows a case study design (Bennett and Checkel 2015) and draws on empirical data from academic recruitment in four disciplines at the University of Oslo (UiO) in Norway between 2000 and 2017. We selected Norway as the site for the study due to unique data availability. The exclusive material from professor recruitments allowed us to investigate when and how metrics were applied in these processes over time. The Norwegian material is further relevant in a globalized

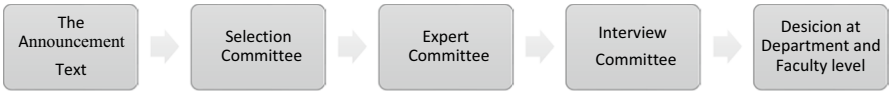

Fig. 1 The Stages in the Recruitment Processes

world where Norwegian universities are converging toward a globalized, rationalized university model (Hansen et al. 2019; Ramirez and Christensen 2013) and subjected to the global increase of metrics in science (Aagaard 2015; Maassen et al. 2011; Sivertsen 2016). In addition, the expert committees in Norwegian recruitment processes consist of both national and international professors, contributing with an international context to the nationally anchored processes. However, to substantiate case relevancy, it should be noted that the Norwegian higher education system chiefly consists of public institutions regulated by the same law and funding model. Public funding from the national budget represents 80% of university income (Hægeland 2015). At the same time, Norwegian universities have a high level of autonomy for internal organization and management. UiO is the oldest and, up until recent university mergers in the sector, the largest, and was a preferred case since the university is less troubled by applicant shortage than other smaller and more peripheric universities may be.

Recruitment processes at UiO are executed at department level and governed by national and university regulations. These processes are long action chains, where decisions are taken at different stages (displayed in Figure 1). First the vacancies are announced publicly according to national laws. Then universities appoint an expert committee consisting of internal and external professors as expected by national regulations, where the internal professors often serve as secretaries. The administration is not present. University regulations further instruct the expert committee to evaluate the candidates according to their academic, pedagogic, personal, management and administration qualifications, and their publications record, where academic qualifications are the most important. However, these rules only define academic qualifications in broad terms and neglect to provide guidelines for the use of metrics, which leaves the committee members in control of defining research quality and whether to apply metrics in their evaluations. The expert committees are expected to reach consensus on the candidate ranking, but if consensus fails, disagreements are listed with several conclusions in the report. The highest ranked candidates are then called for an interview consisting of an internal professor(s), a student, department leaders, and administration staff. Finally, the department and faculty councils decide on eventual job offer(s). The UiO regulations also state that, before appointing an expert committee, a selection committee may be appointed to identify the most prominent candidates based on a more superficial evaluation of their CVs. The selection committee consists mostly of internal professors and, often, the department head.[1] Serving as the internal representatives

[1] A few selection committees studied in this paper also included an external professor.

in recruitment committees alternated among the staff, and the appointed internal representatives often served in two committees in the same processes, but never in all three of them (selection, expert, and interview committees).

To study development over time, we have chosen to investigate recruitment reports from the selection, expert and interview committees which were written at the time of the events. These reports are written in a formal language and include a short summary of the individual candidate's qualifications and research contribution, ending with an overall evaluation and a candidate ranking. The committee reports and the names of committee members are semi-official documents and are available for all applicants, evaluators and department leaders to read. Applicants are also able to complain about the expert committees' evaluation and candidate ranking. This transparency incentivizes evaluators to write consistent accounts and suggest that the reports may be rationalized without less legitimate evaluation criteria such as preferring a candidate based on their gender, ethnicity, or personal acquaintanceship, which has been documented to be influential by prior studies (Van den Brink et al. 2006; Tavares et al. 2019). This paper will neither confirm nor deny such effects. However, this transparency further signals the reports' strategic position and reflects what is perceived as legitimate reasoning around research quality in the disciplines and what tenured professors define as research quality in a semi-transparent process as well as their evaluation of metrics. Hence, these reports are highly valuable as proofs of the legitimate status of metrics within the disciplines.

Data Collection

Universities are collections of relatively independent academic disciplines with different evaluation cultures and use of metrics (Lamont 2009; Wilsdon et al. 2015; Whitley 1984). A major epistemic distinction exists between the social sciences and natural sciences, but there are also internal variations within these traditions (Becher and Trowler 1989). To obtain a representative sample we thus selected two disciplines with epistemic differences from each tradition, aiming to also cover variations between and within these two research traditions. In the social sciences, we therefore chose *economics*, with documented strong use of metrics (Hylmö 2018; Fourcade et al. 2015), combined with *sociology*, representing a more typical discipline within the epistemic tradition of the social sciences (Christensen and Klemsdal 2019; Becher and Trowler 1989). In the natural sciences, some disciplines are more typical of what Becher and Trowler (1989) labeled pure sciences while others resemble technologies. To capture this variation, we selected *physics* as an example of the former and *informatics* as the latter (Becher and Trowler 1989). To ensure similarities between the recruitment processes, we excluded temporary positions, affiliated positions, and targeted positions.

Table 1 Number of recruitment processes 2000–2017

Research Tradition	Discipline	Year			
		2000–2005	2006–2012	2013–2017	*Total*
Social sciences	Economics	3	4	4	*11*
	Sociology	2	4	5	*11*
Natural sciences	Physics	4	7	5	*16*
	Informatics	8	3	10	*21*
Total		*17*	*18*	*24*	*59*

Method

Content analysis refers to a research technique used to make replicable and valid inferences from texts (or meaningful matter) to the contexts of text use (Krippendorff 2013). This analysis included 59 recruitment processes with 1,172 applicants, 57 announcement texts, 11 selection committee reports, 59 expert committee reports, and 29 interviews committee reports written in English and Norwegian (See appendix A1 in the online supplementary material for a document overview). We divided the recruitment cases into three periods (2000–2005; 2006–2012; 2013–2017) to obtain roughly equal numbers of recruitment processes and to trace development over time[2]. Table 1 shows the number of included cases by field and time period. To secure anonymity, names and numbers have been altered in citations.

To analyze the *relative* importance of metrics we first identified the frequency of use and then evaluated the importance of the different assessment criteria applied in the candidate evaluations. We first analyzed the documents with the NVivo computer software program where we coded text containing references to the candidates' qualifications with predefined categories of assessment criteria. Mostly, parts of the text were only coded with one predefined category, but in situations where a sentence contained references to multiple qualifications, the sentences were attached with multiple codes. The predefined codes were based on prior studies on academic recruitment processes (Herschberg et al. 2018; Van den Brink and Benschop 2011; Van den Brink et al. 2006) and the UiO's instructions for candidate evaluations (see Table 2). We subdivided *Research Quality* into the qualitative assessment of the candidates' research and *Metrics* referring to the quantitative measures of scientific and scholarly publications (Wilsdon et al. 2015). Following this definition, we coded both arguments more broadly referring to the quantitative measures of scientific output such as "*X has too few publications*" and arguments referring more specifically

[2] Some of these categories contain few cases. To ensure that eventual observed development was not due to one or two deviant cases, we also applied other categorizations of the cases: We divided the cases into two groups instead of three, we investigated the cases divided by each year, we divided the cases into before and after the implementation of selection committees where relevant, and merged fields in time periods to increase the number of cases. These different categorizations did not alter the results, and since dividing the cases into three periods displayed roughly equal numbers of cases and enabled us to trace development over time, we settled on this categorization.

Table 2 General categories of assessment criteria in academic recruitment processes

Assessment criteria category	Definition
Research quality	
Qualitative Research Evaluations	Qualitative research evaluations of a candidate's research
Matching Research Profile	Assessment of the match between the candidate's research profile and the specific call
Metrics	The quantitative analysis of scientific and scholarly publications
Future Potential	The candidate's future potential for research output
Teaching	Teaching experience such as supervising, pedagogic courses, formal pedagogic competence and pedagogic publication
Third mission activities	The diffusion and use of scientific and scholarly knowledge in society and the economy (broadly defined, such as industry relations and dissemination activities)
International experience	International experience such as international work experience, international networks, language skills and relations to a specific country
Administration skills	Administrative and leadership experience
Personality and personal background	Social skills, personality, gender, and ethnicity
Grants	Records and experience in attracting funding and grants

to metrics such as "*X most cited paper has received 78 citations and his h-index is 14*". In addition, we categorized the quantitative measures of scientific output in more detail into different types of metrics as we soon will describe. Patents were not defined as metrics. Creswell (2013) outlined three ways to construct categories: predefined categories, categories defined by data, or a mix. Following Creswell's (2013) latter strategy, we made two additional sub-categories out of the arguments from the evaluation committees of *Research Quality*: (1) to which degree the candidate's research profile matched the specific vacancy (Matching Research Profile); and (2) the candidate's future potential for research output (Future Potential).[3]

After analyzing the frequency of use, we evaluated the importance of the different criteria for the outcome. In these evaluations, criteria applied in the final ranking were regarded as more important than criteria mentioned in the general candidate description, and we classified the most important criteria which constituted the basis of the candidate ranking. We often defined one to four criteria as the most important since there was seldom only one criterion that clearly outperformed the others. Information from department and faculty reports was used to identify which of the committee's rankings eventual job offers were based on and, subsequently, the committee's relative importance.

After mapping the importance of metrics in relation to other criteria we further investigated which types of bibliometric indicators were used. In the analysis, we coded metrics after categories based on the data (see Table 3), as suggested by

[3] Table A2 in the appendix (see online supplementary material) provides examples of evaluations coded in these categories.

Table 3 Metrics applied in recruitment processes

Type of metrics	Definition
Metrics on journal publications	
Volume of journal publications	Number of journal publications
International journal publications	Number of publications in international journals, often in contrast to Norwegian journals. We also distinguished between international and internationally recognized journals
Top international journals	Number of publications in top international journals, often in contrast to international journals
Publications in thematic journals	Number of publications in relevant journals for the specific call
Metrics on impact and citations	
Citations and h-index	Citation numbers and h-index
Other types of publications	
Conference proceedings	Number of conference proceedings and workshop presentations
Volume of books	Number of books and book chapters

Creswell (2013). Thus, these categories do not reflect the most common metrics in science but those most frequently applied in the documents.[4]

Results

In this section we first present the results from our analysis of the reports from the expert committees which constituted the backbone of the recruitment processes and were present in all disciplines. We then show how the various committee types (selection, expert and interview committees) differed in their use of metrics, reflecting the stagewise nature of the recruitment processes.

Different Importance of Metrics in Discipline-dependent Evaluation Cultures

Our material revealed discipline-dependent evaluation practices where the disciplines assessed the candidates with their own criteria. The social sciences mostly made open calls, asking for general competence in the discipline and evaluated candidates on different aspects of research performance. In sociology, the expert committees produced lengthy qualitative evaluations of the candidates' research, whereas the expert committees in economics wrote shorter evaluations with a stronger emphasis on metrics. In contrast, the natural sciences advertised more defined positions, often linked to a specific research group, requiring specific research profiles with particular technical skills. Subsequently, the expert committees mainly assessed whether the candidates' research profile matched the requirements. The natural sciences expert committees often excluded candidates

[4] Appendix Table A3 (see online supplementary material) shows examples of text coded in the subcategories.

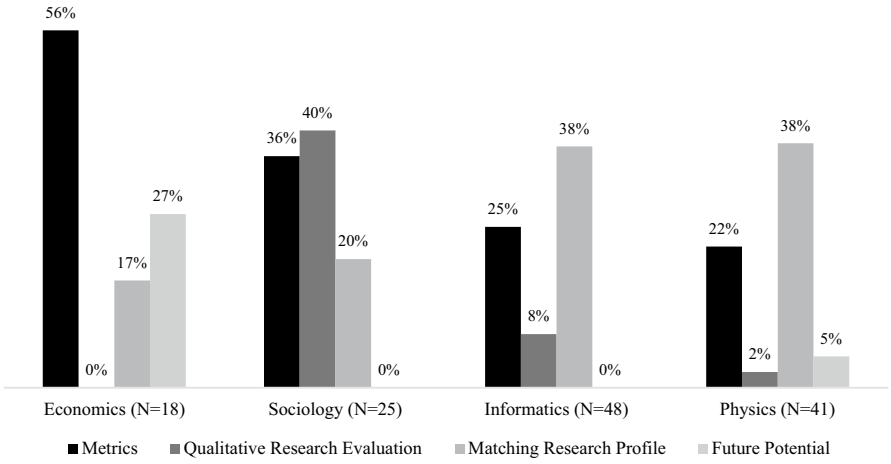

Fig. 2 Most important assessment criteria in the expert committees by academic discipline (percentage). N refers to the number of important assessment criterion/a detected in the expert committee

with irrelevant research profiles, and seldom penalized candidates for having weak publications records if they possessed the preferred competence. Thus, both research quality and metrics played different roles in the diverse evaluation cultures. These characteristics remained relatively stable throughout the study, and the use and importance of metrics must be understood within the context of these discipline-specific evaluation cultures.

In these evaluation cultures metrics were important assessment criteria but seldom the most salient and importance varied across the disciplines. Figure 2 shows how often *the four most important assessment criteria* were applied by the expert committee as the most decisive criteria for the final outcomes. The figure displays how metrics were the most important criterion for the expert committees in economics, but only the second most important criterion in the other disciplines. In sociology, qualitative evaluations were more important than metrics, whereas having the desired research profile was more important than metrics in the natural sciences. Appendix Table A4 (see online supplementary material) shows further details of all the assessment criteria applied by the expert committees.

Metrics were, thus, often used as the second most important criteria but seldom as a replacement for traditional peer review, aside from in economics where metrics gradually replaced the expert committees' quantitative evaluations. Metrics were also often used to rank equally qualified candidates where other criteria proved incapable. For instance, a committee in sociology, almost exclusively basing their entire evaluation on quantitative evaluations, used metrics to differentiate the last two finalists: "*Tara has, however, less publications in refereed-journals*" (recruitment number 1102). Metrics were also used as a benchmark excluding candidates with weak publication records despite other strong qualifications: "*Theodore's activity is well in line with the vacancy announcement, but*

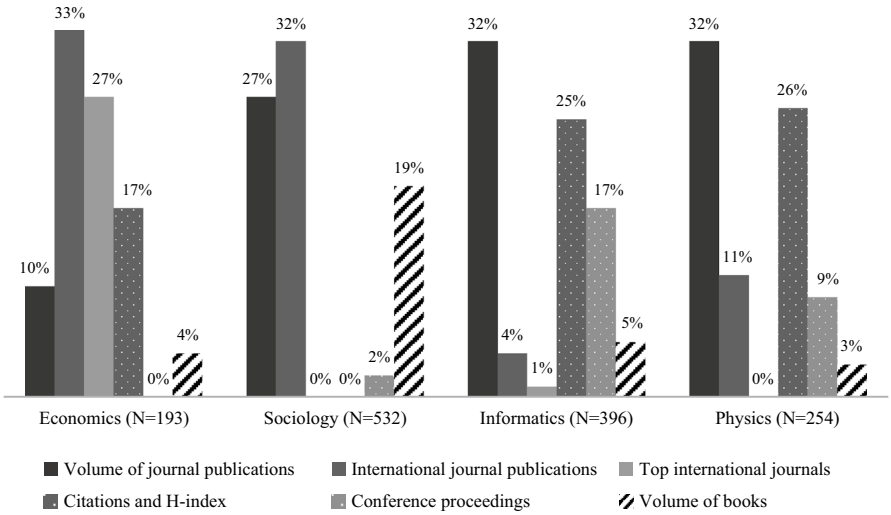

Fig. 3 Types of metrics applied by expert committees in each discipline (percentage). N refers to the number of arguments coded as metrics in the expert committee

the overall scientific impact is limited. And with only 3 published papers after 2005, the candidate does not meet the description 'outstanding physicist with an excellent research profile'" (recruitment number 1209).

Discipline-dependent Use of Metrics

The disciplines further relied on different types of metrics. Social sciences chiefly preferred metrics referring to publication volumes and journal quality, while the natural sciences relied more on various metrics such as the number of publications, impact metrics and the number of conference proceedings. Figure 3 illustrates these findings and displays *how often the different types of metrics* were applied by the expert committees. The figure demonstrates how the number of publications in international journals and top international journals were most commonly applied in economics, while citations and h-index were used more frequently in the natural sciences. Although the use of different types of metrics expanded, as we soon will describe, this discipline-specific use of metrics persisted throughout the study period. Table A5 in the appendix (see online supplementary material) provides further details of the expert committees' use of metrics.

A Modest Increase of Metrics in the Expert Committees

Even though we foremost observed highly stable evaluation cultures in the expert committee, we also discovered a modest but discipline-dependent increased use of metrics. We found the strongest metrics increase in economics —the discipline also valuing metrics the highest (see Figure 2). The announcement texts

in economics stated that candidate evaluations would particularly emphasize scientific output and international publications and, here, the expert committees employed metrics as the most important assessment criterion throughout the period. In addition, we saw an increased reliance on metrics in the expert committees' reports evolving from lengthier evaluations of the candidate's research to shorter summaries of their CVs and metrics. The metrics, thus, not only served as additional information, but to some extent as a replacement of more quantitative evaluations. However, the increasing reliance on metrics did not imply the disappearance of more qualitative candidate evaluations, and the strong reliance on publications in top journals was questioned by an expert committee as late as 2013.

We also found an increased reliance on metrics in informatics though less prominent than in economics. In early announcement texts, informatics only called for vague "research" qualifications, and not until 2013 did these texts request publication lists from the applicants. From 2015 onwards, these texts also mentioned that candidates with "*a strong record of publications in relevant fields*" were preferred. Throughout this study period, the expert committees showed increased reliance on metrics, but always inferior to possessing the desired research profile.

In contrast to economics and informatics, we detected a more moderate increase of metrics in sociology. The announcement texts in sociology emphasized that the evaluations would strongly weigh scientific output and number of international publications, but this attention to metrics was gradually reduced from 2012 when the announcements also stated that "*In the assessment of publications, originality, quality and scope will be emphasized."* Nevertheless, metrics appeared more frequently in the expert committee reports from 2006 onwards, but whether this use also signified an increased significance is questionable as the qualitative evaluations remained the backbone of the candidate evaluation throughout the period. For instance, some expert committees ranked candidates with shorter publication records higher if the quality of their work was regarded as better. Other expert committees argued that metrics were not the same as research quality, while one committee relied exclusively on qualitative evaluations and neglected the use of metrics.

Whether the use of metrics increased in physics is also more doubtful. The announcement texts mostly referred to research quality in general terms and, throughout the period, the expert committees treated metrics as being less important than possessing the desired competence for the position.

Expanding Use of Different Types of Metrics

Over time, the expert committees expanded their use of different types of metrics. Whereas the earliest expert committees only applied a few simple metrics, the committees gradually expanded their use to multiple types of metrics. This expansion was most evident in informatics, where the expert committees in recent years described the candidates with a range of different metrics.

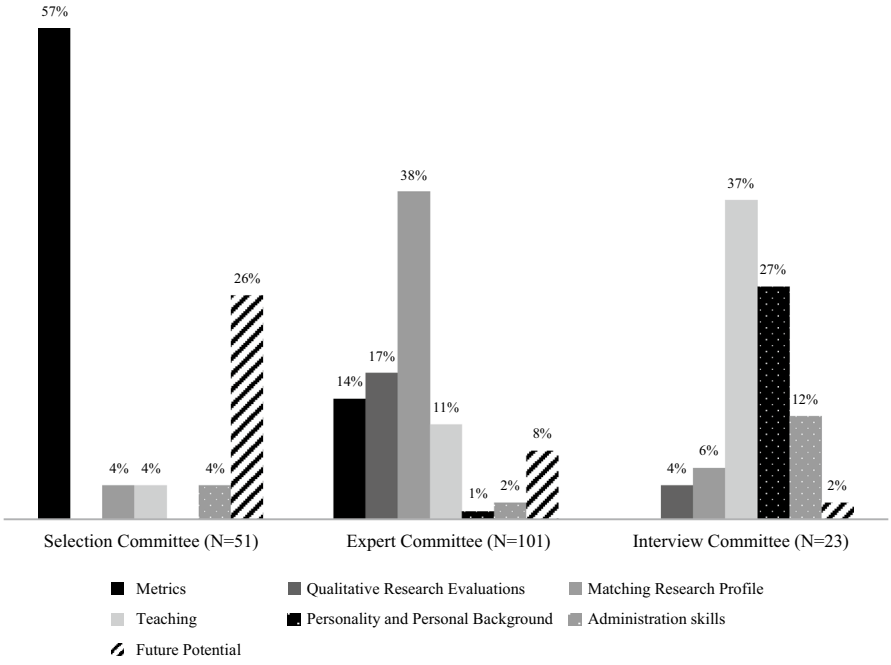

Fig. 4 Most important criterion in the three different committee types (percentage). N refers to the number of the most important assessment criterion/a in the different committees

> Jon has an impressive publication list of his age with over 210 co-authored publications, including 74 peer-reviewed journal articles, 134 peer-reviewed conference publications, 9 book chapters and 2 books. Jon has given 16 invited talks and in addition he holds 1 patent. Currently Jon has 5 journal papers in press or in review process. Jon has an h-index of 14, according to Scopus, which reflects a high number of citations to his publications (From recruitment process number 1306).

The nature of the expanded use of metric types varied between the disciplines. In economics, we observed a shift from relying on international publications to a stronger reliance on publications in the top five journals, while in sociology the expert committees shifted from focusing on the number of refereed publications to the number of publications in recognized international journals. In the natural sciences, we also detected a steady rise in citations and h-index use.

Metrics as a Screening Tool

A major observation in the analysis was how metrics were used chiefly as a screening tool by the selection committees which almost exclusively screened the candidates on metrics. In contrast, the expert committees applied numerous evaluation criteria while the interview committees foremost assessed their candidates on teaching experience and personal skills. Figure 4 displays the different assessment criteria

the committees use and shows the frequency of the most important criterion in the three different committees. This figure shows that metrics were the most important criteria in over half of the selection committees, but only in one of the expert committees and in none of the interview committees. Note that the figure includes all four disciplines which unequally contribute to the totals. There are more expert and interview committees in the natural sciences and more selection committees in the social sciences.

The selection committees were relatively new committees and used mostly in the social sciences where they were introduced in 2012. These committees were instructed to select the most eligible candidates, based on their CVs, for further assessment by the expert committees. In economics, these committees ranked the candidates by their number of publications in top international journals, expecting more from senior candidates. In sociology, these committees ranked the candidates by their publication records, rewarding publications in recognized international journals in sociology the most. Before the introduction of selection committees, some expert committees also conducted more superficial selections of the candidates based on their CVs, research, and publication records before evaluating the most qualified researchers more thoroughly. However, these earlier candidate selections did not exclusively emphasize metrics but also other evaluation criteria such as teaching experience and administrative skills. The description of the candidates' research was also lengthier. Thus, the implementation of selection committees boosted the overall importance of metrics in the recruitment process.

The introduction of selection committees to the social sciences correlated with a sharp increase of applicants in these disciplines from an average of 12 applicants per recruitment before 2012 to 71 applicants per recruitment after 2012. Selection committees were less common in the natural sciences with eight candidates per recruitment, on average. The few selection committees in the natural sciences did not apply a one-dimensional use of metrics either; they also evaluated the candidates on other criteria such as their research profiles and experience with grant proposals. Apart from the increased number of applications, there were no other changes or development in the rest of the recruitment procedures that co-occurred simultaneously with the introduction of selection committees. Another important structural finding was the different importance of the various committee types in the four disciplines. The interview committees in the natural sciences were, for example, more influential than in the social sciences; whereas 9% of the interview committees in the social sciences changed the expert committees' candidate rankings, 18% of the interview committees in physics and 28% in informatics reorganized the rankings of the expert committees. Hence, the committees' relative importance reflected the relative prestige of the associated assessment criteria. Teaching experience and personality-related aspects are more important in the natural sciences as these criteria are more closely related to the interview committees, which were more influential in the natural sciences.

Table 4 summarizes the results from our analyses over time and across disciplines.

Table 4 Summary of key developments in recruitment processes

	2000–2005	2006–2012	2013–2017
Economics	• Announcement texts emphasized the candidate's publications in international journals • The expert committees used recognized and top international journals as the most important criteria	• Announcement texts also included *future potential* as an assessment criterion • Selection committees were introduced • The expert committees expanded their use of different types of metrics	
Sociology	• Announcement texts emphasized the candidate's publications in international journals and future potential • The expert committees' qualitative evaluations of the candidate's work were most important, but the number of publications in international journals were the second-most important	• The expert committees expanded their use of different types of metrics	• The announcement texts tuned down the emphasis on the volume of publications • Selection committees were introduced • The expert committees more frequently used the volume of publications in *recognized* international journals instead of the volume of publications in international journals
Physics	• The announcement texts did not mention metrics as a criterion • The expert committees applied matching research profiles as the most important criteria, but the volume of publications was the second-most important	• The expert committees expanded their use of different types of metrics	• The announcement texts mentioned metrics
Informatics	• The announcement texts did not mention metrics as a criterion • The expert committees applied matching research profiles as the most important criteria, but the number of publications and conference proceedings were the second-most important criterion	• The expert committees expanded their use of different types of metrics	• The announcement texts mentioned strong publication records in relevant fields

Discussion

We initiated this paper by asking how the use of metrics in academic recruitment had changed over time in different academic disciplines, and how organizational theory could explain the similarities and differences in the use of metrics. We developed three different expectations; Firstly, following the path-dependency perspective (Thelen 1999), we expected stable evaluation cultures relatively unaffected by the general spread of metrics. Secondly, relying on the isomorphism perspective (DiMaggio and Powell 1983), we conversely anticipated increased reliance on metrics. Finally, resting on translation theory (Czarniawska and Sevón 1996), we expected discipline-dependent change. In this section we will discuss our results in relation to our analytical framework and expectations.

Stable Evaluation Processes

To understand the use of metrics in recruitment processes we must understand the context in which metrics are used. In this analysis we found stable and discipline-dependent evaluation cultures where the disciplines evaluated the candidates with their own criteria and notion of research quality, also addressed by other studies (Herschberg et al. 2018; Levander et al. 2019). These evaluation cultures reflected the disciplines' epistemic traditions (Lamont 2009), as, for instance, how economics as a methodologically quantitative oriented discipline placed emphasis more strongly on metrics (Hylmö 2018; Fourcade et al. 2015), while sociology, as a more heterogeneous discipline comprising both qualitative and quantitative research traditions (Christensen and Klemsdal 2019), relied more strongly on quantitative research evaluations. The different evaluation cultures also mirrored differences in academic work (Välimaa 1998), as, for instance, how the more specific and narrow announcement texts in the natural sciences assigned to certain research groups probably were due to research groups being more common in these research traditions (Kyvik and Reymert 2017). Subsequently, an evaluation of whether the candidates possessed the right competence and matched the research groups' needs was more important.

The disciplines' evaluation cultures proved highly stable and the way the discipline evaluated the candidates did not change substantially over time. The robustness of peer review has also been addressed by prior studies (Musselin 2013), and could be explained by their fundamental connection to the disciplines' epistemological traditions and academic work (Lamont 2009; Välimaa 1998). This observed stability supports the path-dependency perspective, proposing organizational stability instead of organizational change (Thelen 1999). However, universities are robust organizations (Colyvas and Powell 2006), and recruitment processes as core organizational processes are even more stable (Meyer and Rowan 1977). Thus, stability in this study was expected and must be understood in relation to the fundamental position of the recruitment processes at universities as robust organizations.

A Moderately Increased Reliance on Metrics

Contrary to the literature suggesting an escalation of metrics (Stephan et al. 2017), we only detected a moderately increased reliance on metrics in recruitment processes as suggested by Hicks et al. (2015). Moreover, metrics were already in use in the earliest recruitment processes, and the importance of metrics in the candidate evaluations was more of a modest and steady increase or continuum than an escalation. The increased use of metrics was further chiefly a result of the new selection committees' use of metrics as screening tools, while peer reviews conducted by the expert committees were relatively unchanged. These selection committees were introduced in the social sciences alongside a rapid growth in the number of candidates, probably due to the internationalization of the academic job market (Chou and Gornitzka 2014), and metrics were applied in decreasing the number of candidates which reduced the complexity in the evaluations. This finding alludes to the pivotal observation in organization studies that actors are incapable of handling the complexity of reality in decision-making processes, and ought to reduce complexity (March and Simon 1958). Similarly, evaluators are nearly incapable of reading and evaluating the entire works of 50 or more applicants and metrics are suited to reduce this complexity. Metrics used as judgment tool has also been discovered by prior studies (Hammarfelt and Rushforth 2017) and is in accordance with bibliometricians' recommendations as supplements rather than replacements of traditional expert judgment (Hicks et al. 2015). However, even though metrics were foremost applied as screening tools, this was nevertheless a new and prominent role of metrics in the recruitment processes, which may send strong signals to future applicants that strong records of metrics are needed to be considered for a position.

Despite the observed stability in evaluation practices and the primary role of metrics as screening tools, we also detected a moderately increased reliance on metrics in the expert committees. This was most evident in economics but also present in sociology and informatics. Whereas metrics were used primarily as a very important criterion by the expert committees in sociology, informatics, and physics, the expert committees in economics also tended to use them as a replacement of traditional peer review. The disciplines' use of metrics also became more complex, applying a range of different metrics. Considering the relatively short time period of this study, this moderate increase should not be underestimated, but rather implies that metrics have reached core organizational processes where changes are least expected (Meyer and Rowan 1977).

The increased reliance on metrics could not only be understood as an attempt to reduce complexity in recruitment processes but is hard to understand detached from the general global spread of standards in organizations (Brunsson and Jacobsson 2000; Power 1999), and increased accountability demands (Krücken and Meier 2006; Ramirez 2006). In the analysis we observed that expert committees often expressed difficulty when ranking the candidates and these difficult decisions were to be taken in a semi-official process with a larger research community surveillance and candidates being able to complain. This context creates strong need for the results to be perceived as legitimate, which may be satisfied with metrics as they may be perceived as more objective (Ramirez and Meyer 2013). Moreover,

whereas a committee's peer review could be disputed, a candidate's superior metrics could not and could thus prevent complaints. The use of metrics in these processes could therefore be understood as an example of how standards satisfy legitimacy and accountability demands (Brunsson and Sahlin-Andersson 2000), and the semi-official context with high legitimacy demands helps us understand why metrics have reached core university processes as uncovered in this study.

Moreover, the observed uncertainty and difficulty in ranking candidates may also have triggered mimetic behavior where use of metrics in one discipline was mimicked by other disciplines (DiMaggio and Powell 1983). The use of external professors may have further paved way for normative isomorphism (DiMaggio and Powell 1983), as the external and international professors may have brought with them metrics use from their home institutions. However, this study is unable to account for how metrics have spread, but the theory of isomorphism exemplifies many potential paths for this spread.

The Disciplines' Different Use of Metrics

As the use of metrics was discipline-dependent, so was the intensity of the spread. In addition, the disciplines applied different types of metrics reflecting the established peer review cultures. For example, as a more homogenous discipline, economics (Hylmö 2018; Fourcade et al. 2015) relied on the number of publications in the top journals only, while sociology being a more heterogeneous discipline (Christensen and Klemsdal 2019) relied on the number of publications in a much broader set of recognized journals. Similarly, as conference proceedings are valued as research output in the natural sciences (Wilsdon et al. 2015), these disciplines logically quantified the candidates' number of conference proceedings, which was absent in the social sciences. This diverse use of metrics suggests that the disciplines have translated and selected those metrics that suit their notions of research quality best, aligned with expectations derived from the translation of idea perspective arguing that organizations adapt and translate universal templates and travelling ideas to fit their own local context (Czarniawska and Sevón 1996).

Finishing our discussion, we conclude that the chiefly stable evaluation cultures in the recruitment processes provide an empirical example of how core university processes are mainly characterized by path dependency (Thelen 1999; Pierson 1993) and only moderately by isomorphism (DiMaggio and Powell 1983). This was further in line with our expectations as recruitment processes were sites where change was less expected. However, we did detect modest change, but this change was highly discipline-dependent, thus offering stronger support to the translation perspective (Czarniawska and Sevón 1996) than isomorphism. The disciplines' diverse use of types of metrics is hence an empirical example of how metrics not only are passively and evenly diffused as may be derived from isomorphism (DiMaggio and Powell 1983), but they are rather actively translated to fit the local context as suggested by translation scholars (Czarniawska and Sevón 1996).

Conclusion

Metrics have proliferated in science (Hicks et al. 2015), and studies have even suggested an increased reliance on metrics in academic recruitment processes (Stephan et al. 2017). However, the literature on the use of metrics in recruitment processes offer little evidence for how and with which importance metrics are applied in these evaluation processes (Stephan et al. 2017; Aagaard 2015; Hicks et al. 2015). The lack of empirical studies could partly be due to the confidentiality of recruitment processes and thus unavailability of data. However, in this study we were able to get access to unique data containing confidential reports from professor recruitments in economics, sociology, physics and informatics at the University of Oslo between 2000 and 2017, which enabled us to explore the internal dynamics inside the recruitment processes, and account for how metrics are used in candidate evaluations.

In the study, we display stable and discipline-dependent evaluation cultures where metrics foremost were applied as a screening tool to limit the number of eligible candidates rather than replacements of traditional peer review. However, the exception is economics, where metrics also replaced more qualitative research evaluations to some extent. Hence, using metrics in recruitment processes does not necessarily imply fundamental change or elimination of peer reviews but could rather suggest that metrics are used as a supplement. However, even though the disciplines' evaluation cultures were foremost unchanged, we also detected moderately increased reliance on metrics in the peer review process. This moderate increase indicates that the spread of metrics has also reached core university processes where change is least expected (Musselin 2013). Hence, this observation of robust evaluation practices provides an empirical example of how core university processes are chiefly characterized by path-dependency mechanisms (Pierson 1993; Thelen 1999) and only moderately by isomorphism (DiMaggio and Powell 1983). Additionally, the disciplinary-dependent spread of metrics offers a theoretical understanding of how travelling standards such as metrics are not diffused but rather translated to fit the local context resulting in context-dependent spread as suggested by translation scholars (Czarniawska and Sevón 1996). The disciplinary differences in recruitment processes have been addressed by prior studies (Herschberg et al. 2018; Levander et al. 2019; Van den Brink and Benschop 2011; Van den Brink et al. 2006; Hammarfelt and Rushforth 2017), but not well covered and this study has contributed with important descriptions of the evaluation cultures in sociology, economics, physics and informatics.

This paper has analyzed semi-official recruitment reports written in a rationalized language. These documents contribute valuable information of the reported use of evaluation criteria and what is perceived as legitimate reasoning around research quality and the valuation of metrics in candidate evaluations by tenured scholars. Moreover, they are relatively free from recollection bias as they are written at the time of the evaluations and, thus, used in studies of academic recruitments (Hammarfelt and Rushforth 2017; Hylmö 2018). However, these documents do not openly account for the recruiters' strategic behavior (Musselin 2010), inbreeding effects

(Altbach et al. 2015; Tavares et al. 2019), or gender bias (Van den Brink and Benschop 2011). Nor do they show unreported quantitative or qualitative research evaluations. These documents are, thus, only indicators of and not equivalent to the verified use of metrics. How the field-specific evaluation cultures detected in this study align with prior studies of these fields' characterized notion of quality strengthens these documents as indicators of the verified research evaluation. Nevertheless, as these unreported evaluations leave more subtle traces in the documents, these effects are harder to detect and must be studied with other research methods. For instance, comparing the description of female and male candidates (Fürst 1988) or calculating their chances to proceed in recruitments, controlled for their qualifications (Lutter and Schröder 2016), may uncover gender biases. Interviews, experiments, and ethnographic methods may also be helpful to uncover these effects.

The literature on evaluation practices in academic recruitment processes is scarce, and this study has contributed with a comprehensive account of the use of metrics at the University of Oslo over the last two decades across academic disciplines. These observations should moreover be common in the international field of science as University of Oslo along with other universities are subject to global trends (Ramirez and Christensen 2013) and increased use of metrics (Aagaard 2015). Our selection of disciplines further covered epistemic differences between and within the major research tradition of social and natural sciences. However, this study also poses new questions and illustrates the need for additional research. Firstly, what are the effects of the observed use of metrics (De Rijcke et al. 2016)? For example, to what degree did the metrics-oriented selection committee single out qualified candidates? This paper has not been able to answer this question, but independently of the answer, only the awareness of the selection of candidates by metrics may stop researchers from following more risky ideas, afraid of losing the needed publications records to proceed in future recruitments (Laudel 2017). The effects of the moderate use of metrics should, thus, be more closely studied since a moderate increase does not necessarily imply moderate effects. Secondly, to control for different types of vacancies, we only studied permanent professor positions, which are the most important positions at universities with the most experienced candidates. However, follow-up studies may investigate whether there is a different use of metrics when hiring for other types of positions, such as postdocs. We also urge scholars to explore more profoundly the mechanisms that contribute to change and stability in evaluation cultures in recruitment processes.

Lastly, in our analysis we observed how the expert committees controlled the definition of research quality and decided on use of metrics, which implies that the modestly increased reliance on metrics in peer review has been a result of researchers' own choices despite scholarly protests as seen in the Doha-declaration. Scholars' control over peer review processes has also been address by prior studies (Musselin 2013). However, the researchers were not the only factor influencing the use of metrics in recruitment processes. Which type of evaluation committees were appointed and which missions and importance they were assigned with also affected the use of metrics reflecting how organizational structure affects organizational outcome (March and Simon 1958; Egeberg 2012; Gulick 1937). For instance, the introduction of selection committees in the recruitment processes boosted the

importance of metrics in the recruitment processes. Thus, even though the scholars controlled the definition of research quality in peer review, they were not in control of the organizational structure of the processes which also affected the importance of metrics. We thus urge scholars not only to investigate the dynamics within peer review, but also how organizational structure impacts the importance of peer review in recruitment processes.

Acknowledgments A special thanks to research professor Liv Langfeldt and the Centre for Research Quality and Policy Impact Studies, professor Francisco Ramirez and the SCANCOR community at Stanford, lecturer Robyn Lockwood, professor emeritus Tom Christensen, two anonymous reviewers, and the author's two supervisors, professor Tobias Bach and professor Taran Thune, for their support and expertise.

Funding Funding was provided by Norges Forskningsråd (Grant No. 256223). The author is submitted as a PhD student at the Department of Political Science at the University of Oslo.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aagaard, Kaare. 2015. How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy* 42: 725–737. https://doi.org/10.1093/scipol/scu087.

Aksnes, Dag W., Liv Langfeldt, and Paul Wouters. 2019. Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *SAGE Open*. https://doi.org/10.1177/2158244019829575.

Altbach, Philip G., Maria Yudkevich, and Laura E. Rumbley. 2015. Academic inbreeding: local challenge, global problem. *Asia Pacific Education Review* 16: 317–330. https://doi.org/10.1007/s12564-015-9391-8.

Baert, Patrick, and Alan Shipman. 2005. University under siege? *European Societies* 7(1): 157–185. https://doi.org/10.1177/2158244019829575.

Becher, Tony, and Paul Trowler. 1989. *Academic Tribes and Territories. Intellectual Enquiry and the Culture of Disciplines*. New York: The Society for Research into Higher Education and Open University Press.

Bennett, Andrew, and Jeffery T. Checkel. 2015. *Process tracing: from metaphor to analytic tool*. Cambridge: Cambridge University Press.

Brunsson, Nils, and Kerstin Sahlin-Andersson. 2000. Constructing Organizations: The Example of Public Sector Reform. *Organization Studies* 21(4): 721–746. https://doi.org/10.1177/0170840600214003.

Brunsson, Nils, and Bengt Jacobsson. 2000. The Contemporary Expansion of Standardization. In *A World of Standards*, eds. Nils Brunsson and Bengt Jacobsson, 1–16. New York: Oxford University Press.

Chou, Meng-Hsuan, and Åse Gornitzka. 2014. *Building the Knowledge Economy in Europe*. Massachusetts: Edward Elgar Publishing Limited.

Christensen, Tom, and Lars Klemsdal. 2019. Social Science Disciplines in Complex Development Contexts - the Professional Dimension of Reputation Management. In *Universities as Agencies*, eds. Tom Christensen, Åse Gornitzka, and Francisco Ramirez, 117–154. Palgrave Macmillan, London: Reputation and Professionalization.

Clark, Burton. 1978. Academic Differentiation in National Systems of Higher Education. *Comparative Education Review* 22(2): 242–258.

Clark, Burton. 1983. *The higher education system: academic organization in cross-national perspective*. Berkeley, California: University of California Press.

Colyvas, Jeannette A., and Walter W. Powell. 2006. Roads to institutionalization: The remaking of boundaries between public and private science. *Research in Organisational Behavior* 27: 305–353. https://doi.org/10.1016/S0191-3085(06)27008-4.

Creswell, John W. 2013. Data Analysis and Representation. In *Qualitative Inquiry & Research Design. Choosing Among Five Approaches*, eds. John W. Creswell and Cheryl N. Poth, 179–212. London: Sage.

Czarniawska, Barbara, and Guje Sevón. 1996. *Translating Organizational Change*. New York: Walter de Gruyter.

De Rijcke, Sarah, Paul Wouters, Alex D. Rushforth, Thomas P. Franssen, and Björn Hammarfelt. 2016. Evaluation practices and effects of indicator use - A literature review. *Research Evaluation* 25: 161–169. https://doi.org/10.1093/reseval/rvv038.

DiMaggio, Paul J., and Walter W. Powell. 1983. The iron cage revisited: Institutional isomorphism and collective rationality in organization fields. *American Sociological Review* 48: 147–160.

Drori, Gili S., John W. Meyer, and Hokyu Hwang. 2006. *Globalization and Organization, World Society and Change*. New York: Oxford University Press.

Egeberg, Morten. 2012. How Bureaucratic Structure Matters: An Organizational Perspective. In *The Sage Handbook of Public Administration*, eds. B. Jon and Cheryl N. Pierre, 116–127. London: Sage.

Fourcade, Marion, Etienne Ollion, and Yann Algan. 2015. The Superiority of Economists. *Journal of Economic Perspectives* 29(1): 89–114.

Fürst, Elisabeth L'orange. 1988. *Kvinner i Akademia - inntrengere i en mannskultur? [Women in Academe – intruders in a male culture]*. Oslo: NAVFs sekretariat for kvinneforskning.

Gulick, Luther. 1937. Notes on the Theory of Organization. In *Papers on the Science of Administration*, eds. Luther Gulick and Lyndall Urwick, 3–45. New York: Institute of Public Administration.

Hammarfelt, Björn, and Alex D. Rushforth. 2017. Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation* 26(3): 169–180. https://doi.org/10.1093/reseval/rvx018.

Hansen, Hanne Foss, Lars Geschwind, Jussi Kivistö, Elias Pekkola, Rómulo Pinheiro, and Kirsi J. Pulkkinen. 2019. Balancing accountability and trust: university reforms in the Nordic countries. *Higher Education* 78(3): 557–573. https://doi.org/10.1007/s10734-019-0358-2.

Herschberg, Channah, Yvonne Benschop, and Marieke Van den Brink. 2018. Selecting early-career researchers: the influence of discourse of internationalisation and excellence on formal and applied selection criteria in academia. *Higher Education* 76: 807–825. https://doi.org/10.1007/s10734-018-0237-2.

Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. 2015. The Leiden Manifesto for research metrics. *Nature* 22(April). http://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351.

Hylmö, Anders. 2018. *Disciplined reasoning: Styles of reasoning and the mainstream hetrodoxy divide in Swedish economics*. Lund, Sweden: Lund University.

Hægeland, Torbjørn. 2015. Finansiering for kvalitet, mangfold og samspill - Nytt finansieringssystem for universiteter og høyskoler [Funding for quality, diversity and interaction - New funding system for universities and university colleges]. In *Report to the Ministry of Education and Research*, ed. Ministry of Education and Research. Oslo.

Krippendorff, Klaus. 2013. *Content Analysis And Introduction to Its Methodology*, 3rd ed. The United States of America: SAGE Publications Inc.

Krücken, Georg, and Frank Meier. 2006. Turning the University into an Organizational Actor. In *Globalization and Organization. World Society and Organizational Change*, eds. Gili Drori, John W. Meyer, and Hokyu Hwang, 241–257. New York: Oxford University Press.

Kyvik, Svein, and Ingvild Reymert. 2017. Research collaboration in groups and networks: differences across academic fields. *Scientometrics* 113: 951–967. https://doi.org/10.1007/s11192-017-2497-5.

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

Langfeldt, Liv, and Svein Kyvik. 2011. Researchers as evaluators: tasks, tensions and politics. *Higher Education* 62: 199–212. https://doi.org/10.1007/s10734-010-9382-y.

Langfeldt, Liv, Maria Nedeva, Sverker Sörlin, and Duncan A. Thomas. 2020. Co-existing Notions of Research Quality: A Framework to Study Context-specific Understandings of Good Research. *Minerva* 58(1): 115–137. https://doi.org/10.1007/s11024-019-09385-2.

Laudel, Grit. 2017. How do National Career Systems Promote or Hinder the Emergence of New Research Lines? *Minerva* 55(3): 341–369. https://doi.org/10.1007/s11024-017-9314-4.

Levander, Sara, Eva Forsberg, and Maja Elmgren. 2019. The meaning-making of educational proficiency in academic hiring: a bland spot in the black box. *Teaching in Higher Education*. https://doi.org/10.1080/13562517.2019.157660.

Lutter, Mark, and Martin Schröder. 2016. Who becomes a tenured professor, and why? Panel data evidence from German sociology, 1980-2013. *Research Policy* 45(5): 999–1013. https://doi.org/10.1016/j.respol.2016.01.019.

Maassen, Peter, Eli Moen, and Bjørn Stensaker. 2011. Reforming higher education in the Netherlands and Norway: the role of the state and national modes of governance. *Policy Studies* 32(5): 479–495. https://doi.org/10.1080/01442872.2011.566721.

March, James G., and Herbert A. Simon. 1958. *Organizations*. Cambridge, Massachusetts: Blackwell Publishers.

Meyer, John W., and Brian Rowan. 1977. Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology* 83(2): 340–363. https://doi.org/10.1086/226550.

Montgomery, Henry, and Sven Hemlin. 1991. Judging Scientific Quality. A Cross-disciplinary Investigation of Professorial Evaluation Documents. *Göteborg Psychological Reports* 21(4).

Musselin, Christine. 2013. How peer review empowers the academic profession and university managers: Changes in relationships between the state, universities and the professoriate. *Research Policy* 42(5): 1165–1173. https://doi.org/10.1016/j.respol.2013.02.002.

Musselin, Christine. 2010. *The Market for Academics [Les marché des universitaires]*. New York, USA: Routledge.

Paradeise, Catherine, and Jean-Claude Thoenig. 2015. *In Search of Academic Quality*. UK: Palgrave Macmillan.

Pierson, Paul. 2004. *Politics in Time: History, Institutions, and Social Analysis*. Princeton: Princeton University Press.

Pierson, Paul. 1993. When Effect Becomes Cause: Policy Feedback and Political Change. *World Politics* 45(4): 595–628.

Polanyi, Michael. 2000. The Republic of Science: Its Political and Economic Theory. *Minerva* 1: 54–73.

Power, Michael. 1999. *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Ramirez, Francisco. 2006. The Rationalization of Universities. In *Transnational Governance: Institutional Dynamics of Regulation*, eds. Marie-Laure Djelic and Kerstin Sahlin-Andersson, 225–244. Cambridge: Cambridge University Press.

Ramirez, Francisco, and Tom Christensen. 2013. The formalization of the university: rules, roots and routes. *Higher Education* 65: 695–708. https://doi.org/10.1007/s10734-012-9571-y.

Ramirez, Francisco, and John W. Meyer. 2013. Universalizing the University in a World Society. In *Institutionalization of World-Class University in Global Competition*, eds. Jung Cheol Shin and Barbara M. Kehm, 257–275. New York: Springer.

Sahlin-Andersson, Kerstin. 1996. Imitating by Editing Success: The Construction of Organizational Fields. In *Translating Organizational Change*, eds. Barbara Czarniawska-Joerges and Guje Sevón, 69–92. Berlin: Walter de Gruyter.

Sauder, Michael, and Wendy Nelson Espeland. 2009. The Discipline of Rankings: Tight Coupling and Organizational Change. *American Sociological Review* 74(1): 63–82. https://doi.org/10.1177/000312240907400104.

Scott, William Richard, and Gerald F. Davis. 2007. *Organizations and organizing: Rational, Natural, and Open System Perspectives*. Upper Saddle River, N.J: Pearson Prentice Hall.

Sivertsen, Gunnar. 2017. Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications*. https://doi.org/10.1057/palcomms.2017.78.

Sivertsen, Gunnar. 2016. Publication-Based Funding: The Norwegian Model. In *Research Assessment in the Humanities. Towards Criteria and Procedures*, eds. Michael Ochsner, Sven E. Hug, and Hans-Dieter Daniel, 79–90. Zürich: Springer.

Stephan, Paula, Reinhilde Veugelers, and Jian Wang. 2017. Blinkered by bibliometrics. *Nature*. https://doi.org/10.1038/544411a.

Sydow, Jörg, Georg Schreyögg, and Jochen Koch. 2009. Organizational Path Dependence: Opening the Black Box. *Academy of Management Review* 34(4): 689–709. https://doi.org/10.5465/amr.34.4.zok689.

Tavares, Orlanda, Cristina Sin, and Vasco Lança. 2019. Inbreeding and Research Productivity Among Sociology PhD Holders in Portugal. *Minerva* 57(3): 373–390. https://doi.org/10.1007/s11024-019-09378-1.

Thelen, Kathleen. 1999. Historical institutionalism in comparative politics. *Annual Review of Political Science* 2: 369–404. https://doi.org/10.1146/annurev.polisci.2.1.369.

Van den Brink, Marieke, and Yvonne Benschop. 2011. Gender practices in the construction of academic excellence: Sheep with five legs. *Organization* 19(4): 507–524. https://doi.org/10.1177/1350508411414293.

Van den Brink, Marieke, Margo Brouns, and Sietske Waslander. 2006. Does excellence have a gender? *Employee Relations* 28(6): 523–539. https://doi.org/10.1108/01425450610704470.

Välimaa, Jussi. 1998. Culture and identity in higher education research. *Higher Education* 39(2): 119–138. https://doi.org/10.1023/A:1003248918874.

Wedlin, Linda, and Kerstin Sahlin. 2017. The Imitation and Translation of Management Ideas. In *The SAGE Handbook of Organizational Institutionalism*, eds. Royston Greenwood, Christine Oliver, Thomas B. Lawrence, and Renate E. Meyer, 102–127. London: SAGE Publications Ltd.

Whitley, Richard. 1984. *The Intellectual and Social Organization of the Sciences*. Oxford: Clarendon Press.

Wilsdon, James, Liz Allen, Eleonora Belfiore, Philip Campbell, Stephen Curry, Steven Hill, Richard Jones, Roger Kain, Simon Richard Kerridge, Mike Thelwall, Jane Tinker, Ian Viney, Paul Wouters, Jude Hill, and Ben Johnson. 2015. The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. ed. HEFCE. https://doi.org/10.13140/RG.2.1.4929.1363.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.