

Controlling the Future of Academe

Academic and Managerial Logics in Professorial
Recruitment

Ingvild Reymert



PhD Thesis

Department of Political Science
Faculty of Social Sciences
University of Oslo

© **Ingvild Reymert, 2021**

*Series of dissertations submitted to the
Faculty of Social Sciences, University of Oslo
No. 875*

ISSN 1564-3991

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard.
Print production: Reprintsentralen, University of Oslo.

For my mother.

Preface

This thesis has been submitted to the Department of Political Sciences and is supported by the Research Council of Norway through The Centre for Research Quality and Policy Impact Studies (R-QUEST), Grant Number 256223, with Nordic Institute for Studies in Innovation, Research and Education (NIFU) as host. I have been supervised by Professor Tobias Bach and professor Taran Thune. The thesis is a collection of four papers; two were written in collaboration with Jens Jungblut (University of Oslo), Siri B. Borlaug (NIFU), Liv Langfeldt (NIFU), and Dag W. Aksnes (NIFU), and two by me as sole author.

Acknowledgements

This thesis has been written with substantial valuable support from colleagues, friends, and family. I would first like to thank my two supervisors, Tobias and Taran, for valuable, constructive, and excellent advice, knowledge, and expertise. This thesis could not have been completed without you. A special thanks for being flexible and creative during these COVID times and for finding new ways of supervising, such as walking around Østensjøvannet, when health regulations did not allow anything else. And for finding time to comment on my work when you have been placed in quarantine with lots of home-schooling responsibilities for your kids.

I also want to thank Professor Emeritus Tom Christiansen for reading through the first draft of the thesis and contributing with valuable comments and stimulating discussions since we became acquainted at SCANCOR in 2019. I also owe great thanks to Professor Emeritus Bern Aardal for continuing to give me valuable advice even though research policy is somewhat outside his field of expertise.

This thesis arose from the work of the R-Quest Centre overseen by Research Professor Liv Langfeldt. I am deeply grateful to be part of this dynamic international research group with colleagues from five European countries. I would like to give a special thanks to Liv, my co-authors (Jens, Siri and Dag), and my colleagues at the Centre for Science and Technology Studies (CWTS) for letting me stay there in May 2018, where especially Paul Wouters and Inge van der Weijden provided me with important feedback on my work. A special thanks to Gunnar, Thed, Ingeborg, Maria, Sarah, Thomas, Espen, Maria, Duncan, Agnete, Erlend, Kristoffer, Håvard, Fredrik, Anders, and Kody. I would also like to thank all my NIFU colleagues for theoretical discussions, coffee breaks, workout sessions, and social activities. A special thanks to Nicoline, Mari, Elizabeth, Vera, Astrid, Per Olaf, Svein, Lars, Bjørn, Silje, Silje, Magnus,

Asgeir, Hebe, Jannecke, Thea, and Lone. Finally, thanks to Marte Mangset (OsloMet) and Lars Klemsdal (UiO) for inspiring conversations and feedback on my work, and to the eight anonymous reviewers of the first three papers.

The thesis has been much inspired by my stay at SCANCOR at Stanford in 2019. I am deeply grateful to professor Francisco Ramirez for including me in this unique community, allowing me to participate in debates with leading scholars in organizational theory and a special thanks to professor John W. Meyer for discussing my work and professor Justin Grimmer for letting me audit his lectures on quantitative text analysis. I am very grateful for all the special friendships I made there and a special thanks to Maria, Johan, Barbara, Marouane, Alison, Ulf, Molli, Michael, Julia, Claus, Tobias, Judith, Kalliopi, and Lill. I would also want to thank Lecturer Robin Lockwood, who helped me discover the joy of writing academic English and Marisa and her family for welcoming us to California and opening doors for us in the Bay Area. Finally, thanks to all the children, adults, and families at Madera Grove who made Ida's and Klara's stay in Palo Alto such a great experience.

Writing a doctoral thesis is, for some reason, easier when one is with people in the same situation. Maybe it is the shared tragedy or the inspiration we give one another. I opt for the latter and give a special thanks to all the fellow doctoral students I met along the way, whether at NIFU, UiO, Stanford, conferences, seminars, or courses. A special thanks to Hans Rasmus, Sebastian, Stine, Karin, Leif Erlend, Torbjørg, Per Magnus, Karl, Tea and Øyvind (Postdoc). And to Professor Michal Kotnarowski for giving a great lecture at the ECPR school in Bamberg 2020.

A major contribution of this thesis is the unique material I was able to access. Collecting such material is not simply a question of obtaining permission; it also involves a substantial amount of work for university administrative staff to find documents dating back to 2000, some of which were in remote storage. I am extremely grateful for all the efforts these people made to find those documents, and I sincerely hope that they will find this work valuable. Thanks to all those I interviewed and those who answered the survey.

I would also like to thank my friends, especially "Gjengen" and my parents and parents-in-law for looking after my kids as the writing hours escalated. And to Ingrid, who was the greatest supporter of my academic work, I miss you so much. A special thanks to my mom and dad for all their feedback and support.

Finally, Thomas deserves the most thanks. I am not thanking you for taking half of the responsibility at home, as anything else would be unacceptable, but I thank you for all your love, patience, and moral support. And most of all, thank you for your intellectual contributions:

helping me with R, listening to my long arguments/monologues, discussing and disagreeing, and spicing up my arguments whenever I got too bland. To my dearest daughters Klara and Ida: after two minutes with you, all my academic worries are gone. I love you.

Ingvild Reymert

June 2021

Summary

Many public organizations like universities and hospitals have traditionally been steered by the professions. These organizations have, however, been challenged by New Public Management (NPM) reforms that have introduced a managerial logic as a new organizing principle. Consequently, these organizations have evolved into hybrid organizations influenced by both professional and managerial institutional logics. However, it remains unclear how we can understand the interplay between these institutional logics at the shop floor level of these organizations, and how the organizations handle these, sometimes, conflicting logics.

This thesis explores the dynamics between professional and managerial logics in one of the universities' most important processes: professorial recruitment. These hiring processes are crucial for universities as they control organizational membership, but also shape their cultural foundation. This thesis thus asks: *How do different logics and their interplay influence academic recruitment processes? How and why does their importance vary across time, countries, and disciplines?*

These questions are studied through historical institutional and neo-institutional perspectives and with a mixed methods approach that includes analysis of confidential recruitment reports, interviews, and cross-national survey data. This unique data set consists of 109 recruitment reports with selection committee reports, expert committee reports, and interview reports from 66 recruitment processes with over 1,000 applicants in four disciplines at three Norwegian universities from 2000 to 2017. This material opens the black box of what happens inside these confidential processes, which are generally not open to public scrutiny. Additionally, the thesis relies on 26 interviews with university leaders, academics, and human resources staff at three Norwegian universities and a cross-national survey from five northern European countries to account for national differences.

The thesis reveals how academic recruitment processes are primarily anchored in disciplinary evaluative cultures and are highly robust processes dominated by academic logic. At the same time, recruitment is increasingly influenced by managerial logic. However, the growing reliance on managerial logic was moderate and occurred only incrementally through layering, displacement, and translation processes. This moderate increase shows that managerialism has reached the strongholds of professionalism which we least likely expected to adopt managerialism. At the same time this finding uncovered the dominance and resilient of the academic logic and thus also say something of the limits of the proliferation of managerialism when faced with strong professions. There was also only limited tension

between managerial and academic logics; they were reconciled with sequential decision-making that involved attending to one logic and problem at a time.

Table of Contents

- PART I: INTRODUCTION 1**
- 1 INTRODUCTION..... 2**
 - 1.1 Background 2
 - 1.2 Research Questions 6
 - 1.3 Contributions..... 7
- 2 THEORETICAL FRAMEWORK 9**
 - 2.1 Institutional Logics 9
 - 2.1.1 Organizational Response to Multiple Institutional Logics..... 11
 - 2.2 Change and Stability in the Balance between Institutional Logics..... 12
 - 2.2.1 Historical Institutionalism 13
 - 2.2.2 Neo-Institutionalism..... 14
 - 2.3 Disciplinary Evaluative Cultures 16
- 3 RESEARCH DESIGN..... 18**
 - 3.1 Mixed Methods 20
 - 3.1.1 Document Analysis 20
 - 3.1.2 Interviews 21
 - 3.1.3 Cross-National Survey 22
 - 3.2 Validity and Reliability 22
 - 3.3 Ethical Considerations 26
- 4 PRESENTATION OF PAPERS 27**
- 5 MAIN FINDINGS 29**
 - 5.1 Institutional Logics in Academic Recruitment 29
 - 5.1.1 The Sequential Nature of Recruitment Reconciles Institutional Logics 31
 - 5.2 Discipline-Dependent Recruitment and Moderate Country Differences..... 34
 - 5.4 Bibliometrics as an Indication of Managerial Logic..... 37
 - 5.5 The Increased Influence of Managerial Logic 39
- 6 IMPLICATIONS 41**
- 7 SUGGESTIONS FOR FUTURE RESEARCH 44**
- 8 REFERENCES..... 45**
- PART II: PAPERS 56**
 - Paper 1: Handling Hybridity Through Sequential Decisions in Recruitment of Professionals 57
 - Paper 2: Bibliometrics in academic recruitment: A screening tool rather than a game changer 86
 - Paper 3: The role of metrics in peer assessments..... 113
 - Paper 4: Are evaluative cultures national or global? A cross-national study on evaluative cultures in academic recruitment processes in Europe 129
- APPENDIX 151**

Part I: Introduction

1 Introduction

The thesis is divided into two parts. The first part begins with this introduction as the point of departure for the research questions. I then present the theoretical framework in which I operationalize professionalism and managerialism as institutional logics before explaining change and stability in the dynamics between them from both historical institutional and neo-institutional perspectives. After introducing the theoretical framework, I elaborate on the research design and discuss how the thesis relates to fundamental debates in the philosophy of science and evaluate its ethical considerations. Finally, I discuss the main findings from the four papers, their implications, and propose questions for future research. Part two of the thesis contains the four papers that together answer the overall research questions.

1.1 Background

In most countries public organizations represent a large share of public spending, employ vast numbers of people, and deliver public services to the entire population. However, what are the organizing principles that enable these organizations to make decisions, act, and produce these services?

Traditionally many public organizations like universities, hospitals, and schools have been characterized as professional bureaucracies, which are steered by the professions and their expertise (Ferlie and Geraghty, 2005, Mintzberg, 1983, Abbott, 1988). In recent decades, professionalism as an organizing principle has, however, been challenged by New Public Management (NPM) reforms aimed at redesigning the public sector to be more businesslike. These reforms have introduced stronger internal hierarchies with a new managerial class, and implemented a managerial logic with increased efficiency demands, performance measurements, and a results-driven orientation (Christensen and Læg Reid, 2010, Hood, 1991, Pollitt and Bouckaert, 2017). The reforms have been motivated by reducing costs and making services more efficient (Freidson, 2001); for proponents, professional control is old-fashioned, bureaucratic, and sometimes even corrupt (Deem, 1998).

The NPM reforms have undermined professionalism as an organizing principle in public organizations (Aschhoff and Vogel, 2019, Sehested, 2002). Consequently, many organizations have emerged as stronger organizational actors more able to exercise power over the professions (Drori, Meyer and Hwang, 2006, Brunsson and Sahlin-Andersson, 2000).

Managerialism has, however, not fully replaced professionalism. Instead, both principles are found to be influential in public organizations (Skelcher and Smith, 2015, Denis,

Ferlie and Van Gestel, 2015). Professionalism and managerialism can be conceptualized as different institutional logics defined as “socially constructed, historical patterns of material practices, assumptions, values, beliefs, and rules by which individuals produce and reproduce their material substance, organize time and space and provide meaning of their social reality” (Thornton, Ocasio and Lounsbury, 2012: 2).

That many organizations, especially universities, hospitals, and other organizations with multiple task and professional groups, are faced with multiple logics is already well established in the literature (Greenwood et al., 2011). However, it remains unclear how we can understand the interplay between these logics at the shop floor of the organizations, and how organizations handle multiple and, in some instances, conflicting logics. How is this complexity handled in practice?

Universities are prototypes of professional bureaucracies steered by scholars and dominated by an academic logic which have experienced managerial reforms aimed at transforming them into corporate-managerial organizations (Deem, 1998, Gumpert, 2000). Worldwide universities have adapted global managerial standards (Ramirez, 2006, Musselin, 2013a) and have emerged as stronger organizational actors with a greater ability to muster their resources, as well as their staff (Krücken and Meier, 2006, Musselin, 2013b).

However, universities are also robust and resistant to change (Clark, 1978), and have remained powerfully influenced by an academic logic (Bleiklie, Enders and Lepori, 2017). Universities are thus best regarded as hybrid organizations in which multiple institutional logics interact (Bleiklie et al., 2017). Managerialism has not necessarily decreased the overall power of the academic profession; rather, it has caused an internal reconfiguration of power among scholars. For instance, researchers argue that an academic elite who act as gatekeepers of journals and grant proposal panels and have gained increasing influence at the expense of less privileged scholars who work on short-term contracts (Musselin, 2013a, Barrier and Musselin, 2009).

Stating that both managerialism and professionalism are important institutional logics in universities, however, does not account for the interplay and power balance between them or how universities respond to these competing logics. Different parts of universities are not equally influenced by managerial logic, and core organizational processes have been argued to be more robust in the face of external pressures (Pache and Santos, 2010, Meyer and Rowan, 1977).

Academic recruitment processes are examples of core university processes that experience multiple logics. Recruitment have traditionally been dominated by the academic

profession and an academic logic (Musselin, 2010), but have also become increasingly influenced by a managerial logic, with expanding human resources (HR) departments and university leaders increasingly surveilling these processes (Levander, 2017, Van den Brink, Fruytier and Thunnissen, 2013, Paisey and Paisey, 2018, Pietilä and Pinheiro, 2020). Yet, the literature on academic recruitments remains scarce, and it is not evident to what extent managerial logic has gained a foothold in a process.

Studying academic recruitment is also important since it represents one of universities' most crucial processes: deciding on their most pivotal resource, the talented scholars who enable them to conduct their two primary missions of research and teaching. As most of those appointed to permanent positions will keep their positions for many years, the academics hired today will represent their universities for decades into the future. Thus, recruitment involves highly strategic decisions (Scott and Davis, 2007, Selznick, 1957) that universities for several reasons often conduct very thoroughly, sometimes even tediously.

Additionally, academic recruitment is a key peer review process in which quality, institutional norms, culture, and values are negotiated (Musselin, 2010, Lamont, 2009). For individual academics, recruitment is a critical juncture that shapes their future careers. In an increasingly competitive labor market with fewer long-term positions (Barrier and Musselin, 2009), young scholars in particular may even adapt their research ideas to improve their chances in future recruitments (Laudel, 2017).

One observed indication of increased managerialism in academic recruitment processes is the proliferation of bibliometrics, which refers to the quantitative measurement of scientific and scholarly publications (Wilsdon et al., 2015), in candidate evaluations (Paisey and Paisey, 2018, Van den Brink et al., 2013). Despite recommendations to avoid using these measures when evaluating individual researchers (e.g., the San Francisco Declaration on Research Evaluation, or DORA), researchers have argued that bibliometrics have gained importance when evaluating academic candidates (Hammarfelt and Rushforth, 2017, Hicks et al., 2015, Stephan, Veugelers and Wang, 2017). The literature on the use of bibliometrics in recruitment is, however, scarce and provides little evidence for whether bibliometrics have replaced or only supplemented traditional peer reviews.

The relation of the use of bibliometrics to managerial and academic logics is also not straightforward. The use of bibliometrics has been seen as indicator of a managerial logic (Van den Brink et al., 2013) since bibliometric measurements are performance-based indicators (Talbot, 2005) and are related to increased demands for accountability (Bovens, 2005) and the "audit explosion" (Power, 1999), in which organizations are expected to demonstrate objective

output measures (Brunsson and Sahlin-Andersson, 2000). These are also central elements in NPM reforms (Hood, 1991, Pollitt and Bouckaert, 2017).

Bibliometrics, however, also represent measures of research quality (Aksnes, Langfeldt and Wouters, 2019). The volume of scholarly research output, publication channels, and degree of research relevance (citations) resonate with academic logic and more traditional notions of research quality. Bibliometrics are moreover based on scholars' own peer reviews (Musselin, 2013a). Especially in economics, counting publications in prestigious journals is perceived as a good indication of research quality (Hylmö, 2018). The use of bibliometrics relation to the academic and managerial logic should thus be more closely investigated.

Different theoretical perspectives contribute different expectations for the interplay between academic and managerial logics. While a neo-institutional perspective suggests that universities will adapt to managerialism as an emerging institutional logic in their surroundings (Meyer and Rowan, 1977, DiMaggio and Powell, 1983), a historic-institutional perspective instead argues that universities are stable and maintained by feedback mechanisms (Pierson, 1993) and suggests that academic logic will remain the dominant force in professorial recruitment due to path dependency.

Core processes in organizations (Pache and Santos, 2010) are also especially robust to change, and academic recruitment represents a stronghold of professionalism (Musselin, 2010). Professorial recruitment can therefore be seen as a "least likely" case (George and Bennett, 2005) to adopt managerialism in the public sector. Thus, even though this thesis focuses on academic recruitment, the findings of this thesis also have more general implications and can tell us something about the scope of managerialism in public organizations in general.

Universities are often described as systems with different disciplines representing independent systems of knowledge production with their own notions of research quality and evaluative cultures (Lamont, 2009, Whitley, 1984, Becher and Trowler, 1989). Each discipline also evaluates researchers and candidates for vacancies according to its own set of evaluative criteria (Musselin, 2010). This raises the questions of how discipline shapes academic recruitment and whether some disciplinary cultures are more closely related to managerial logic than others. Universities are moreover embedded in highly diverse national traditions and governance arrangements (Clark, 1978, Whitley, 2003), which can also lead to national differences in evaluative cultures.

Disciplinary and national differences in academic recruitment are underexplored. Some studies on academic recruitment only cover one field (Hamann, 2019, Hylmö, 2018), while others only distinguish between the social and natural sciences without disentangling the

specific differences between disciplines (e.g., Herschberg, Benschop and Van den Brink, 2018, Levander, Forsberg and Elmgren, 2019). Most studies are also single country studies (Herschberg et al., 2018, Van den Brink and Benschop, 2011, Levander, 2017, Hylmö, 2018, Pietilä and Pinheiro, 2020, Sanz-Menéndez and Cruz-Castro, 2019, Hamann, 2019), while comparative studies are lacking. This thesis thus adopts a comparative design by investigating evaluation practices in recruitment in five different disciplines in five different northern European countries to close this gap.

1.2 Research Questions

This thesis investigates academic recruitment as a core organizational process and how different institutional logics play out and are handled in recruitment processes. I thus propose a main research question with three sub-questions:

How do different logics and their interplay influence academic recruitment processes? How and why does their importance vary across time, countries and disciplines? (All papers)

1. Explaining institutional logics in academic recruitment: *What are the dominant institutional logics in academic recruitment, and have they changed over time? Why do universities handle these institutional logics in the way they do? (Papers 1 and 2)*
2. Explaining disciplinary and country differences: *Why do the institutional logics unfold differently in different disciplines (Papers 1, 2, 3, and 4) and countries (Papers 3 and 4)?*
3. Explaining the use of bibliometrics: *How are bibliometrics applied in academic recruitment, and how can we explain this use? Which institutional logics do these measures represent? (Papers 2 and 3)*

These questions are addressed in four papers; the first two explore recruitment documents and interviews from a case study in Norway, while the last two are based on a cross-national survey that contributes a comparative perspective to the thesis. All papers contribute findings to answering the main research question while also addressing one or more of the sub-questions.

The first paper identifies and explains the dominant institutional logics in academic recruitment and shows how these logics are handled. For this purpose, I analyzed confidential recruitment documents and conducted interviews with actors participating in these processes in sociology, economics, physics and informatics at three Norwegian universities to investigate the rationales and logics behind their actions. The second paper further explores the confidential recruitment reports for the 2000–2017 period at one of the universities in greater depth and continues to investigate the dominant logics in these processes, but with a temporal perspective. The paper also explains how bibliometrics are applied in recruitment processes, which also responds to the third research sub-question.

The last two papers constitute the comparative part of the thesis and draw on a cross national survey in five northern European countries in physics, economics and cardiology. In the third paper, I continue to investigate how academics apply bibliometrics in their peer evaluations but explore how that use differs between countries. National and disciplinary differences in evaluative cultures are further examined and more thoroughly theoretically explained in connection in the fourth paper.

1.3 Contributions

Empirically, this thesis uncovers how academic recruitment is a highly robust process that is still dominated by an academic logic, anchored in stable disciplinary evaluative cultures and maintained by feedback mechanisms including both resources and interpretive effects (Pierson, 1993). The recruitment are strongly discipline-dependent, but are only moderately affected by national context. The use of bibliometrics in recruitment is moreover moderate and primarily applied as a supplement to the existing peer review process and as a judgment tool that is often used to decrease the number of candidates. The disciplines also adhere to different types of bibliometrics that correspond with their established evaluative cultures.

I also show that professorial recruitment is changing as universities strive to satisfy new demands in their surroundings, as suggested by a neo-institutional perspective (DiMaggio and Powell, 1983). These changes have led to an incrementally increase in the influence of managerial logic even though academic logic remains the more important logic in these recruitment processes.

The robustness of the recruitment can explain why the observed change was moderate and unfolded incrementally, through layering, displacement, or translation processes (Streeck and Thelen, 2005, Czarniawska and Sevón, 1996). Thus, *theoretically*, this thesis contributes

with detailed coverage of how robust core organizational processes are incrementally changing in response to pressure in their environment while being sustained by feedback mechanisms that preserve their established practices.

The thesis also contributes with a new theoretical understanding of how multiple institutional logics are reconciled with sequential decision-making processes (Cyert and March, 1963), which involve separating the logics into different phases of the processes as a type of compartmentalization strategy (Kraatz and Block, 2008). This insight represents a novel finding in the literature on how organizations respond to multiple and partly competing logics.

For *the higher education literature*, this thesis illustrates how academic recruitment is still powerfully dominated by academic logic. At the same time, it is also increasingly influenced by managerial logic, which reflects the emergence of the managerial university (Deem, 1998, Gumpert, 2000). The dynamics between the two institutional logics further illustrates how universities are hybrid organizations serving different societal needs (Clark, 1998, Bleiklie et al., 2017). Universities also proved flexible in handling diverging internal (academic) and external (managerial) demands.

The thesis also contributes with important findings for *the public administration literature*, providing another example of declining professionalism in the public sector (Meyer et al., 2014, Bleiklie, Enders and Lepori, 2013, Scott et al., 2000). Professorial recruitments are strongholds of professionalism where we least likely would expect increased managerialism. Increased influence of managerialism here thus suggests an extensive proliferation of this development in the broader public sector. However, the academic logic proved very powerful in recruitment, which demonstrates that professionalism remains not only a strong but also the prevailing logic within public organizations. This finding also indicates something about the limits of NPM reforms when they are faced with strong professions in public organizations.

Finally, the thesis contributes with novel *methodological* approaches to the study of academic recruitment. Analyzing confidential recruitment reports to go beyond the more superficial accounts of these processes and using cross-national surveys to examine country-level differences are scarcely used in the study of academic recruitment. This novel approach enabled me to fill research gaps in the literature, such as describing the role of bibliometrics in academic recruitment and covering national differences, both of which have enjoyed only sparse coverage before.

2 Theoretical Framework

To explain the dominant institutional logics in academic recruitment and the interplay between them, I create an overarching theoretical framework combining complementary theoretical contributions, while the individual papers go into detail.

In the first part of this framework, I apply an institutional logic approach to identify the organizing principles that guide action in academic recruitment; I then define managerialism and professionalism as institutional logics. The framework's second part explains the dynamics between the logics across time, disciplines, and countries. I supplement the institutional logic approach with other perspectives from the neo-institutional framework such as isomorphism and translation to explain changes in the dynamics between the logics and a historic institutional perspective to explain their stable relations.

2.1 Institutional Logics

Institutional logics operates as institutional guidelines that relate to the central institutions of society, such as family, community, religion, the market, professions, corporations, and the state (Thornton et al., 2012, Friedland and Alford, 1991). These logics structure the world by constraining actors' choices and sense-making and providing them with words, power, and meaning. These logics function as scripts that actors can mobilize to justify their arguments and actions (Thornton et al., 2012, Friedland and Alford, 1991).

Most organizations experience multiple institutional logics that co-exist more or less peacefully (Greenwood et al., 2011). Organizations with persistent patterns of multiple logics are often referred to as hybrid organizations (Skelcher and Smith, 2015). The balance between multiple and conflicting institutional logics is referred to as organizational settlement (Rao and Kenney, 2008, Schildt and Perkmann, 2017) and can change over time when one logic gains more influence and others decline (Scott et al., 2000).

The institutional logics that organizations experience depend on which logics operate in their environment (Pache and Santos, 2010, Greenwood et al., 2011). The interplay between these logics within organizations further depends on internal power and group dynamics (Pache and Santos, 2010), as logics connect to specific organizational groups (Smets, Morris and Greenwood, 2012). Logics firmly linked to influential groups are stronger than logics more loosely tied to less powerful groups (Pache and Santos, 2010, Greenwood et al., 2011). Institutional logics have also been suggested to have different levels of influence over different parts of an organization (Pietilä and Pinheiro, 2020).

Universities are often described as typical *professional* bureaucracies (Mintzberg, 1983) collegially steered by the professions and their expertise (Krücken and Meier, 2006) and dominated by an academic logic in which the university's main purpose is the unfettered quest for knowledge (Musselin, 2006, Sauermann and Stephan, 2013, Bleiklie et al., 2017). According to this logic, the purpose of academic recruitment is to select the most academically qualified scholars, as evaluated by their peers and based on their research contributions (Merton, 1973, Musselin, 2010).

In recent decades, universities have however been increasingly influenced by a managerial logic that understands universities as corporate-managerial organizations that should be hierarchically steered and provide economic advantage through new knowledge and technology (Bleiklie et al., 2017, Deem, 1998, Krücken and Meier, 2006, Ramirez, 2006, Gumport, 2000, Slaughter and Leslie, 1997, Clark, 1998). The organizing principle of managerial logic is not the profession but the *organization*. The main concern of recruitment is thus to secure competitive advantages and meet organizational needs (Paisey and Paisey, 2018). The preferred candidates are scholars with the best bibliometric records who are able to attract external funding and meet organizational needs (Paisey and Paisey, 2018, Van den Brink et al., 2013). Under this logic, university leaders also play an important role in selecting new scholars (Van den Brink et al., 2013).

Scholars have also detected increased attention to guidelines and rules in academic recruitment (Van den Brink et al., 2013, Paisey and Paisey, 2018, Musselin, 2010). This focus to some extent reflects a legalistic-bureaucratic logic based on Weber's concerns about correctness, impartiality, and equal treatment (Meyer et al., 2014). However, the formalization of organizational procedures is also an integral component of becoming stronger organizational actors who are held accountable for their outputs, which aligns with managerial logic (Brunsson and Sahlin-Andersson, 2000, Drori et al., 2006). Formalization is also a prerequisite for strategic positioning in a globalized market (Krücken and Meier, 2006). I thus conceptualize rule following as part of managerial logic, despite its bureaucratic aspects.¹

Academic and managerial logics are based on differently rooted metaphors as organizing principles, offer different views of recruitment, perceive different parties as legitimate evaluators, and assign different levels of importance to evaluation criteria (see Table 1). These are presented as ideal logics, but in reality these logics often overlap (Noordegraaf,

¹ In the first paper, however, I separate these two logics to elaborate on the nuances.

2007, Meyer et al., 2014). The degree of alignment between these logics is thus investigated in the analyses and discussed in the results.

Table 1. Academic and managerial logics in academic recruitment.

	Academic Logic	Managerial Logic
Focus of Academic Recruitment	<ul style="list-style-type: none"> • Peers evaluate who is most academically qualified based on disciplinary standards 	<ul style="list-style-type: none"> • Universities act strategically to attract candidates with competitive advantages in a global market who can cover organizational needs • Recruitment must follow specific rules and regulations
Legitimate Evaluators	<ul style="list-style-type: none"> • Tenured professors 	<ul style="list-style-type: none"> • Academics, but also leaders
Evaluation Criteria	<ul style="list-style-type: none"> • Research contributions • Bibliometrics and success in grant proposals signal academic prestige 	<ul style="list-style-type: none"> • Research output measured by bibliometrics and success in grant proposals • Social and administrative skills to integrate into the work environment • Cover departmental teaching needs • Criteria established by the organization

* Based on (Thornton et al., 2012) definition of institutional logic and operationalized according to studies on universities (Deem, 1998, Gumpert, 2000, Bleiklie et al., 2017) and academic recruitment (Paisey and Paisey, 2018, Van den Brink et al., 2013, Van den Brink and Benschop, 2011, Bleiklie et al., 2017, Musselin, 2010).

2.1.1 Organizational Response to Multiple Institutional Logics

That many organizations experience multiple sometimes conflicting institutional logics is well established (Greenwood et al., 2011); the question is *how* they handle different and sometimes conflicting logics (Pache and Santos, 2010). Kraatz and Block (2008) lay out four responses to hybridity: organizations can resist pluralism by *eliminating* new and conflicting logics, *compartmentalize* logics into separate units, integrate and *balance* logics within the organization, or *create* a new institutional order. While both eliminating and establishing new

logics imply less hybridity, compartmentalization and balancing suggest tolerating hybridity but segregating it or allowing simultaneously operating logics. The literature offers examples of how actors employ identity work to handle this hybridity in balanced models (Aschhoff and Vogel, 2019, McGivern et al., 2015, Shams, 2020, Teelken, 2015, Reissner, 2019), while some scholars have provided examples of how organizations separate logics to avoid tensions between them (Colyvas and Powell, 2006, Dunn and Jones, 2010, Fossetøl et al., 2015, Reay and Hinings, 2009).

Despite the many studies examining hybridity, there has been less attention to how hybridity is handled in recruitment as core organizational processes, which in both universities and other public organizations are faced with different institutional logics (Lægreid and Wise, 2015, Nigro and Kellough, 2008, Sundell, 2014, Lavigna and Hays, 2004, Poocharoen and Lee, 2013).

Academic recruitment are sequential decision-making processes, with candidates passing through multiple forms of judgment (Musselin, 2010, Hamann, 2019). Seminal work on decision-making suggest that competing demands that cannot be handled simultaneously instead could be handled sequentially attending to one problem at a time since organizational attention is limited (Brunsson, 2003, Ocasio, 1997, Cyert and March, 1963). Institutional logics have also been discovered to have varying influence over different phases in academic recruitment (Pietilä and Pinheiro, 2020). Thus, sequential decision-making may help reconcile different logics in academic recruitment, but that demands further investigation.

Expectation 1: Academic recruitment involves both academic and managerial logics that are reconciled with sequential decision-making as a compartmentalization strategy.

2.2 Change and Stability in the Balance between Institutional Logics

While a historical institutional perspective (Thelen, 1999, Pierson, 1993) provides a framework to understand the stability and resilience of academic logic, a neo-institutional perspective such as isomorphism (DiMaggio and Powell, 1983, Meyer and Rowan, 1977) describes how we can understand change and the increased influence of managerial logic.

2.2.1 Historical Institutionalism

Organizations are perceived as stable and robust to change (March and Olsen, 1989, Selznick, 1957, Hannan and Freeman, 1984). To explain stability, historical institutionalists emphasize how organizations emerge in concrete temporal processes (Thelen, 1999) that create path dependency and feedback mechanisms (Pierson, 1993, Pierson, 2004), which lead to increasingly positive returns through resource effects like distributing resources, power, and authority and promoting incentives to conform that gives actors both interests in and abilities to maintain organizations (Pierson, 1993). Change also involve extra costs: setup costs and sunk costs that occur when past investments are abandoned possibly also endangering the organization's legitimacy (Pierson, 1993, Hannan and Freeman, 1984).

Feedback mechanisms can also be more cognitively and culturally expressed through interpretive effects (Pierson, 1993), with actors tending to follow what is perceived as appropriate behavior (March and Olsen, 1989) and organizations leaning on preexisting practices, making adjustments only at the margins (Lindblom, 1959). As organizations evolve, they become more resistant to change because change implies a sense of personal loss. Organizational maintenance becomes a goal in itself (Selznick, 1949).

Not all organizations and organizational processes are, however, equally robust. Organizational stability depends on age where older organizations developing a stronger resistance to change than newer ones (Hannan and Freeman, 1984). Similarly, core organizational processes are more robust than peripheral processes (Pache and Santos, 2010), and are often protected from change with symbolic adaptations to new demands (Meyer and Rowan, 1977). Founding logics, the dominant logic when the organization was established, have also proven more resilient than other logics (Oertel, 2018).

Organizational stability does not imply that organizations never change. On the contrary, historical institutionalists argue that organizations are constantly changing, especially when confronted with complexity (Streeck and Thelen, 2005). However, change is not abrupt but incremental: it involves gradual transformation through, for instance, *displacement*, where new logics slowly gain salience relative to other logics, or through *layering effects*, in which parallel structures gradually gain increased importance (Streeck and Thelen, 2005).

According to historical institutionalism, universities are robust organizations; when change materializes in them, it proceeds incrementally (Clark, 1983). As recruitment is a core process at complex and often old universities strongly dominated by an academic logic, this framework holds that recruitment will remain dominated by academic logic. Based on the

historical institutional perspective, I do not expect an increased use of bibliometric indicators that diverges from traditional methods of candidate evaluation.

The literature on academic recruitment supports the historic institutional expectations and show how especially the judgment processes in these recruitment are less affected by increased managerialism and remain controlled by the academic profession and academic logic (Musselin, 2010). Academic recruitment are moreover associated with both structural and cultural feedback mechanisms. First, scholars' control over recruitment involves significant power, which is hard to remove. Second, scholarly evaluations are anchored in the core scientific value that scholars should be evaluated by their peers (Merton, 1973) and contribute much legitimacy to the process. Removing scholars' control over these processes would thus also involve legitimacy costs. The disciplines' evaluative cultures also provide actors with scripts that are taken for granted and thus hard to change.

Expectation 2: Academic recruitment is robust and dominated by academic logic.

2.2.2 Neo-Institutionalism

In contrast to historical institutionalists' focus on organizations' historical development, neo-institutionalists focus on how organizations are connected to an institutional field with similar organizations. In the face of uncertainty, organizations will adopt rules and myths from their surroundings, mimicking similar organizations to gain legitimacy; this is known as isomorphism (Meyer and Rowan, 1977, DiMaggio and Powell, 1983, Bromley and Meyer, 2015).

DiMaggio and Powell (1983) describe how this process of homogenization occurs through three mechanisms: *coercive isomorphism* results from both formal and informal pressures exerted on organizations by other organizations. *Mimetic isomorphism* happens when organizations model more successful and legitimate organizations in their field. *Normative isomorphism* results from professionalization through formal education of professionals or professional networks spanning organizational boundaries.

The institutional logic perspective is derived from this theoretical framework, with the macro-level templates that organizations adopt are conceptualized as institutional logic. The major difference between isomorphism (DiMaggio and Powell, 1983) and the institutional logic approach (Thornton et al., 2012, Friedland and Alford, 1991) is that the latter focuses on the myriad competing logics in the institutional environment.

While neo-institutional scholars describe how adopting institutional scripts results in homogenization between organizations, translation scholars deny that organizations passively adopt myths and institutional logics in their surroundings. Instead, they claim that organizations will actively translate and edit these scripts and logics to fit their own contexts; adoption of myths can thus lead not to homogenization but to diversity (Czarniawska and Sevón, 1996, Wedlin and Sahlin, 2017, Sahlin-Andersson, 1996). The translation perspective is thus a middle way between isomorphism and path dependency, with the outcome of institutional change depending on both the logics in the institutional field and the organizational context in which these logics are adopted.

Neo-institutional scholars have described how organizations worldwide converge as they adapt to global scripts (Drori et al., 2006). Universities are embedded in highly institutionalized settings with a long history of well-established, shared beliefs and rules (Meyer et al., 2007), and universities worldwide have been found to seek to meet global university standards in their field (Ramirez, 2006, Krücken and Meier, 2006, Ramirez and Christensen, 2013, Buckner, 2017). In this regard, managerialism has gained increased influence as universities adopt elements of it (Deem, 1998, Gumpert, 2000, Musselin, 2013a, Slaughter and Leslie, 1997). This increased managerialism in higher education is related to the proliferation of managerialism in public organizations generally (Hood, 1991, Pollitt and Bouckaert, 2017, Christensen and Lægheid, 2010) and must be understood as a powerful logic in universities that has become taken for granted by many actors in the field.

Even though recruitment processes are highly robust and more resistant to institutional pressure, they are not completely isolated from that pressure. They are also associated with a high degree of uncertainty and substantial legitimacy expectations. Recent studies on academic recruitment also found that managerialism has gained increased influence over these processes through the adoption of general HR and talent management ideas (Van den Brink et al., 2013, Paisey and Paisey, 2018). This may indicate that universities are mimicking more general recruitment trends in other organizations (Lægheid and Wise, 2015, Nigro and Kellough, 2008, Sundell, 2014, Lavigna and Hays, 2004, Poocharoen and Lee, 2013).

The establishment of larger HR departments could also suggest that HR staff are becoming a profession with its own view of how recruitment should be conducted and transmitted through a network as an example of normative isomorphism. The national implementation of performance-based funding models could also create financial incentives to focus more on candidates' bibliometrics (Aagaard, 2015), as a form of coercive isomorphism. Hence, following the neo-institutional perspective could indicate that these processes will

eventually adopt managerial logic, even though they are more resistant in this regard than other organizational processes.

A translation perspective further defies how isomorphism can cause homogenization. For example, when new HR practices are adopted in universities, they will be translated to fit the university context. Similarly, when bibliometrics are mimicked, these measures will be translated to fit established evaluation cultures. Disciplines would thus employ different types of bibliometrics to varying degrees.

Expectation 3: Managerial logic influences academic recruitment processes because it dominates the institutional environment of universities.

2.3 Disciplinary Evaluative Cultures

Universities have been described as collections of academic disciplines with their own professions that represent independent systems of knowledge production with their own epistemic traditions and scholarly work (Becher and Trowler, 1989, Clark, 1978, Whitley, 1984, Lamont, 2009, Välimaa, 1998). The various disciplines have their own evaluation cultures with different notions of research quality and excellence (Polanyi, 2000, Lamont, 2009, Langfeldt et al., 2019).

Professors are hired as members of independent disciplines rather than university-wide assets. Applicants for academic work are thus evaluated according to discipline-specific evaluation criteria (Musselin, 2010); for example, the natural sciences assign more value than the social sciences to international experience (Herschberg et al., 2018). According to a historical institutional perspective, these evaluative cultures represent robust institutional layers which make recruitment robust to change. These evaluative cultures also explain why academic logic unfolds differently in the various disciplines.

Some disciplinary logics may align more with managerialism than others; for instance, some elements of NPM reforms have been shown to resonate better with the way of thinking in the natural than in the social sciences (Christensen, 1991). Similarly, economics has long heavily emphasized metrics (Fourcade, Ollion and Algan, 2015, Hylmö, 2018, Lee, Pham and Gu, 2013), reflecting a quantitative conception of research quality that is more in tune with managerialism.

Universities have moreover evolved in widely diverse national traditions and governance arrangements (Clark, 1978, Sivertsen, 2017, Whitley, 2003). Following a historical

institutional perspective these national differences could create differences in recruitment. For instance, the American custom of inviting candidates to visit the recruiting institution provides greater opportunities to evaluate a candidate's social skills than the European approach, which is limited to formal interviews (Musselin, 2010). However, even though national context shapes academic recruitment processes (Alfonso, 2016, Sanz-Menéndez and Cruz-Castro, 2019, Musselin, 2005), academics strive primarily for peer recognition and prestige within their disciplines (Clark, 1983, Driori, Meyer and Ramirez, 2003, Hessels et al., 2019), which are internationally oriented (Lamont, 2009, Langfeldt et al., 2019).

Expectation 4: Academic logics are disciplinary-dependent and to some extent country-dependent logics.

3 Research Design

Studying academic recruitment processes is challenging since they are far from transparent, and information about what happens inside them is largely unavailable for research. The thesis also poses somewhat complex research questions aiming both to examine the internal dynamics of recruitment and to generalize across disciplinary and national borders.

To answer these questions, I adopted a mixed method that argue that multiple methods are needed to understand complex social phenomena, and pragmatically combines the quantitative and qualitative methods best suited to answer the research questions (Johnson and Onwuegbuzie, 2004, Creswell and Clark, 2017).

The mixed methods tradition is situated between the quantitative and qualitative research traditions and views knowledge as both constructed align a qualitative research tradition and resulting from empirical discovery align a quantitative research tradition (Johnson and Onwuegbuzie, 2004, Johnson, Onwuegbuzie and Turner, 2007, Johnson and Gray, 2010, Creswell and Clark, 2017). The mixed method tradition thus assumes that the social reality in which we live can be studied. At the same time, research is not objective but influenced by the researchers themselves, the questions they ask, the methods they apply, and how they interpret results (Johnson and Onwuegbuzie, 2004).

Following the mixed methods approach, I applied a two-pronged research design. First, to examine what happens inside these black boxes, I examined confidential recruitment documents at three Norwegian universities and conducted interviews with participants in those processes. I then applied a cross-national survey to account for national variations.

Both a analyzes of confidential reports and survey analyses of academic recruitment are scarce. This novel approach enabled me to shed light on unclear items with which other methods have had only limited success.

Table 2 displays the methods, case selections, and theoretical approach applied for the four papers and how they relate to the research questions.

Table 2. The research questions and methods adopted in this thesis.

Papers	(1) Institutional Logics	(2) Use of Bibliometrics	(3) Metrics in Peer Assessments	(4) National and Disciplinary Evaluative Cultures
Main Research Question	X	X	X	X
RQ1) Institutional logics	X	X		
RQ2) Disciplinary differences; National differences	X	X	X X	X X
RQ3) Bibliometrics		X	X	
Method	Interviews and content analysis of recruitment documents	Content analysis of recruitment documents	Cross-national survey	Cross-national survey
Cases	Physics, informatics, sociology, and economics at three Norwegian universities (from 2000 to 2017)	Physics, informatics, sociology, and economics at the University of Oslo from 2000 to 2017	Physics, economics, and cardiology in Norway, Sweden, and the Netherlands	Physics and economics in Norway, Sweden, Denmark, the Netherlands, and the United Kingdom
Theoretical Framework	Institutional logics	Historical institutionalism; Neo-institutionalism; Translation perspective	Disciplinary evaluative cultures	Historical institutionalism; Neo-institutionalism

Mixed methods are flexible but can also be criticized from both from the quantitative and qualitative research traditions. While a quantitative research tradition , aiming to generate context-free generalizations, would scrutinize the study’s external validity, a qualitative research traditions, oriented to provide a deep understanding of social phenomena through rich observational data, and will thus focus on the study’s internal validity (Johnson and Onwuegbuzie, 2004, George and Bennett, 2005). After describing the methods applied in the

papers, I elaborate on the validity and reliability in terms of the philosophy of science. I conclude the chapter by discussing ethical considerations.

Before I proceed, I re-emphasize that this thesis focuses primarily on recruitment to permanent professorships (the top rank in academia) since my focus is core strategic processes. Recruiting for non-permanent positions or for lower-level scholars is of less strategic importance to universities and thus out of scope here. Keeping the type of recruitment constant also easier enables comparisons by time, discipline, and country.

I also want to accentuate that while Norway is part of the cross-national comparison, it was selected as the site of the qualitative approach for its unique data availability; Norway is one of a very few countries in which internal information regarding academic recruitment is accessible to researchers. The material is further relevant for the studies of universities in an internationalized world, as Norwegian universities are converging toward a globalized university model (Hansen et al., 2019, Ramirez and Christensen, 2013).

Some contextual information is needed. Norway has a largely public higher education system. The Ministry of Education and Research owns most universities and provides most of their funding; the recruitment of professors is consequently regulated by national laws and university regulations. The national mandates are largely general but do specify that candidates must possess a doctoral degree and relevant teaching experience to be appointed as associate professors. Evaluation criteria and recruitment procedures are further specified by the universities, which generally announce associated professor vacancies; only rarely are vacancies at the professor level advertised, as universities expect employees to later undergo an internal evaluation for promotion to that rank. In the findings, I further describe the sequential nature of these recruitment processes.

3.1 Mixed Methods

In this section I briefly describe the methods applied in this thesis; they are detailed in the papers.

3.1.1 Document Analysis

To document the internal dynamics of recruitment, I analyzed confidential recruitment reports obtained by applying to both Norway's Ministry of Education and Research and the universities. The documents covered 66 recruitments with over 1,000 candidates, including 65

announcement texts, 12 reports from selection committees, 66 expert committee reports, and 31 reports from interview committees in four disciplines at three Norwegian universities from 2000 to 2017. Reports from the departments and faculty councils were also included; in some cases, complaints and responses from the departments were also examined (see Appendix Table 1 for document overview).

The committee reports were written in formal language and include a short summary of each candidate, a final ranking, and concluding remarks. The names of candidates and committee members were included. The selection and expert committee reports were partly public, as they were distributed to candidates and sometimes to department staff. Applicants were also able to object to their evaluation and ranking. The interview reports were, however, not distributed outside the committee and included more personal evaluations of the candidates.

After obtaining the documents, I stored them in a private file with a security code per the European General Data Protection Regulation (GDPR) and labelled the documents by institution, discipline, year, and type of recruitment. I first conducted a thorough document analysis (Krippendorftt, 2013) of 59 recruitment processes in four disciplines at the University of Oslo. In this operation, I listed the number of candidates applying for the position, their gender, and their eventual Norwegian work and education background. I also documented whether the interview committee or department/faculty council changed the ranking of the expert committee and the names and affiliations of evaluators on those committees. I then coded the use of evaluation criteria after predefined codes (see Appendix Table 2) and evaluated their importance using the NVivo software package. In the next step, I analyzed the documents from sociology and physics at two other universities to see if I found the same patterns there. Due to workload concerns I was not able to conduct as thorough an analysis of these recruitments. The analyses of evaluation criteria are detailed in papers 1 and 2.

3.1.2 Interviews

Institutional logics are abstract phenomena but can be observed in actors' justifications (Meyer et al., 2014). To uncover the logics guiding recruitment actors, I thus conducted semi-structured interviews (Martin, 2013) with participants in recruitment processes.

The interviews were conducted in sociology and physics at three Norwegian universities. Altogether, I completed 26 interviews with academics serving on evaluation committees, heads of departments, heads of administration, and HR staff. I transcribed the interviews and analyzed them with NVivo. The total length of the transcribed text was 133,108

words, varying from 2,168 to 9,716 words per interview (see appendix for details of the interviews).

To systematically anchor these theoretical concepts to the data, I adopted a two-step approach. First, I coded the documents using informant-centric terms (first-order analysis). I then categorized these codes into research-centric terms (second-order analysis) according to the operationalization of the three institutional logics (Gioia, Corley and Hamilton, 2013). This rigorous method enabled me to be more transparent in how I understand the logics.

3.1.3 Cross-National Survey

To account for country variance, I applied a *cross-national survey*, which also uncovered discipline differences, and controlled for the evaluators' backgrounds variables, including their own bibliometric indicators. The survey was developed together with my research group, R-Quest; we invited participation from researchers in three disciplines (economics, physics, and cardiology) at the most relevant research institutions (as defined by Web of Science (WoS) data and journal classifications) in five countries (Norway, Sweden, Denmark, the Netherlands, and the United Kingdom).

Collecting contact information was tedious, especially for cardiologists, who often had double affiliations, one at a university and the other at a hospital. We thus decided to exclude cardiologists in Denmark and the United Kingdom. To obtain country samples with equal disciplinary representation, I thus applied different samples in the two survey papers: one with three disciplines in three countries, and one with two disciplines in five countries (See Appendix Table 3 and 4). The data were analyzed using logistic regression (Agresti, 2013) in the R software package. More information about the survey is provided in the appendix.

3.2 Validity and Reliability

External validity

External validity concerns whether a study's results can be generalized and relates to case selection (Yin, 2018, George and Bennett, 2005). Disciplinary variation is a major difference in higher education (Becher and Trowler, 1989), and to be able to generalize, I aimed to cover a wide set of disciplines. The three disciplines included in the survey (physics, economics, and cardiology) were predefined by the R-Quest project as representing three different research traditions (natural sciences, social sciences, and medicine and health).

These disciplines, however, also have commonalities. All are highly internationalized, dominated by males, and represent more positivist research traditions than, for instance, those found in the humanities (Becher and Trowler, 1989). With its heavy emphasis on quantitative methods, economics is not a typical discipline in the social sciences (Hylmö, 2018, Fourcade et al., 2015). To expand the selection of disciplines, I thus included sociology as a more typical representation of the social sciences (Christensen and Klemsdal, 2019); I also added informatics to represent technologies as a subgroup of the natural sciences, in contrast to physics, which is a pure science (Becher and Trowler (1989).

To generalize across countries, I employed a cross-national survey. The five Northern European countries in the survey (Norway, Sweden, Denmark, the Netherlands, and the United Kingdom) were also predefined by the research group. These countries are relatively similar, with well-developed and well-funded higher education systems, which could lead to more moderate country variance than if non-European countries had been included. Still, these countries also have important differences in terms of, for example, career systems (Frølich et al., 2018) and performance-based research funding systems (Sivertsen, 2017), which do suggest country variation.

In a quantitative research approach aiming for generalizability (Johnson and Onwuegbuzie, 2004), more countries and disciplines would be desirable, but would also be more time consuming to collect and analyze. I thus argue that my discipline and country selections cover a sufficient share of variance to account for country and discipline differences. Additionally, I underscore that the study is set in Northern Europe and does not claim to report on the global context.

The case study's external validity could also be criticized for since it involve only one country. Although Norway is relevant in a global context, this country also has special characteristics that could influence the results. For instance, Norway has implemented more moderate NPM reforms in its higher education system than many other countries (Christensen and Læg Reid, 2010, Bleiklie et al., 2011), which could lead to less confrontation between university leaders and academics (Sehested, 2002). These more modest reforms could also have led to a more limited use of bibliometrics.

The results from the cross-national studies, however, reveal that Norway does not deviate significantly from other northern European countries. Additionally, the case analysis was anchored in prior research and in paper 1 represented an example of a least likely case from which we can make more theoretical interferences (George and Bennett, 2005). This aspect

bolsters the assumption that lessons from this case study could be generalized to similar countries.

Internal validity

Internal validity is about whether the measuring instruments actually reflect the phenomena they are intended to detect (Mosley, 2013), which also relates to construct validity: whether the concepts have been correctly operationalized (Yin, 2018). To secure internal validity in the survey, the questionnaire was carefully anchored in prior research and pretested by colleagues and a test sample to see if the questions were appropriate.

To secure internal validity in qualitative research, I was more concerned about whether the tools were appropriate to capture the intended aspects. Institutional logics are theoretical and hard to detect but can be observed in actors' justifications (Meyer et al., 2014). I thus conducted interviews to uncover these logics. Additionally, I conferred with the informants to ensure their citations were correctly understood (Yin, 2018). The documents served to better detect the use of evaluation criteria. They were also valuable testimonies for tracing developments over time since they were contemporary records and free from recollection biases. The two data sources complemented and validated each other since they covered the same events (Yin, 2018).

From a qualitative research perspective (Johnson and Onwuegbuzie, 2004), the surveys internal validity might be criticized for not capturing enough details and only measured the individual evaluators' perceptions of the use of evaluation criteria. Additionally, the survey did not capture that candidate evaluations are a collective process. However, I argue that a better understanding of how individual evaluators regard different evaluation criteria when assessing candidates is important as it represents the micro-level foundation of the collective process. Additionally, the interview and document analyses provide a solid context for interpreting these results.

The internal validity of the document analyses can also be criticized as they could contain biases and did not reveal all the subtle dynamics within recruitment processes. For instance, distributing documents to applicants could have led to not documenting when candidates were selected on less legitimate criteria like personal acquaintance or gender and ethnicity. Indeed, both inbreeding effects (Tavares, Sin and Lança, 2019, Altbach, Yudkevich and Rumbley, 2015) and gender biases (Nielsen, 2016, Husu, 2000, Van den Brink, Brouns and Waslander, 2006, Wennerås and Wold, 1997) in academic recruitment have been reported.

The existence of rationalization was confirmed by expert committee members who described how they carefully crafted their argumentation for the candidate ranking, knowing that it could be read by all applicants. Committee members also stated that internal committee disagreements were often omitted from the reports, which thus appeared more unified than the actual evaluations. Still, the committee members believed that the reports largely reflected their actual discussions on and agreement of quality. I thus regard them as strategic documents reflecting the legitimate reasoning around evaluation criteria in academic recruitment by tenured scholars. The potential biases in the documents illustrate the need for triangulation of these two data sources to enhance internal validity (Lynch, 2013).

In sum, I argue that despite these critiques, the method chosen serves the purpose of answering the research questions: capturing institutional logics (paper 1), uncovering the legitimate reasoning around evaluative criteria (paper 2), and detecting national and disciplinary variance of evaluative cultures (papers 3 and 4). However, it is important to repeat that the inferences of internal mechanisms in recruitment are based on the Norwegian case study material.

Reliability

Both quantitative and qualitative research traditions emphasize the importance of avoiding biases and that the analyses would provide the same results if repeated, which is known as reliability (Mosley, 2013, Yin, 2018).

However, reliability concerns are different in the two traditions. In quantitative research, reliability refers to precise replicability of the analyses. In the survey analysis, I therefore checked response rates, tested whether the eventual weights gave other results, used several types of regression analysis, tested how much of my models' variance could be explained with ANOVA tests, examined which independent variables contributed the most explanations with AIC and BIC tests, (Agresti, 2013) and checked for eventual multicollinearity with VIF tests (Lin, 2008). Results were displayed with standard errors. I also controlled for evaluator position, gender, age, and own research output with bibliometric measures drawn from WoS. R scripts were carefully written for replicability purposes and are available on request.

Reliability is more challenging with qualitative methods, in which interpretation of the data can involve biases (George and Bennett, 2005, Mosley, 2013). Avoiding interpretative biases was especially important since anonymity concerns meant I was the only one who could read the documents (Martin, 2013, Mosley, 2013). I thus applied several measures to secure reliability. During analysis, I strictly followed predefined codes that are available to the reader

(see Appendix Table 2). I also analyzed the material using quantitative text analysis, more specifically topic models, in R to see if automatic analyses provided similar results (Grimmer, 2010). I coded the transcribed interviews using Gioia et al. (2013) method to anchor the theoretical aspects to the data. Additionally, I provided a large number of citations from both interviews and documents to provide transparency and actively discussed my interpretations of them with colleagues. I also had some research assistant to transcribe some of the interviews, which whom I discussed the interviews.

3.3 Ethical Considerations

Recruitment involves personal and sensitive information, and thus neither the documents nor the interviews are available for replication. Instead, these data have been carefully managed according to the GDPR and not shared with others. In presenting the results from the reports and interviews, I have altered genders, names, and bibliometric indicators to preserve anonymity. Additionally, the names of the three institutions included in the first paper were disclosed, but this was not considered necessary in the second paper.

I obtained active consent from survey and interviews participants. However, it was not possible to seek permission from candidates in the target recruitment period, which began in 2000. Instead, institutions gave consent on behalf of the candidates. Preserving the anonymity of these candidates has thus been of the highest priority. All studies have been approved by the Norwegian Centre for Research, which follows GDPR regulations; the data will eventually be deleted. The research is also conducted according to The National Committee for Research Ethics in the Social Sciences and the Humanities (NESH) guidelines.

4 Presentation of Papers

The four papers represent independent contributions that collectively answer the main research question. While the first two papers are based on exclusively Norwegian material and describe internal recruitment dynamics, the last two are based on a cross-national survey and account for national and disciplinary differences.

The papers also speak to different literatures. While the first paper is anchored in public administration theory, papers 2 and 4 speak to higher education research, and paper 3 to science evaluation literature and its narrower focus on bibliometrics.

In this section I elaborate briefly on how the papers contribute to answering the research questions. This section is relatively short, since I have chosen to gather the findings from the papers in the next section.

Reymert (submitted to an international journal): Handling Hybridity Through Sequential Decisions in Recruitment of Professionals

The first paper conceptualizes the institutional logics at play in academic recruitment processes, revealing how recruitment was faced with both academic and managerial logics that were handled through sequential decision-making: attending to one issue at a time. Consequently, each institutional logic dominated a different phase of the process. Sequential decision-making also proved flexible for satisfying new demands but altered the balance between the two logics, diminishing the importance of peer review and its academic logic.

Reymert (2020): Bibliometrics in Academic Recruitment: A Screening Tool Rather than a Game Changer

This paper describes assessment processes as largely stable, where bibliometrics were chiefly used as a screening tool to decrease the number of eligible candidates rather than to replace traditional peer review. These robust evaluation practices provided an empirical example of how academic recruitment processes are chiefly influenced by path dependency mechanisms and only moderately by isomorphism. The disciplines also adhered to different types of metrics, suggesting that each discipline translated metrics to fit its established evaluative culture, as the translation perspective suggests.

Langfeldt, Reymert, and Aksnes (2020): The Role of Metrics in Peer Assessments

The third paper provides insights into how academics apply bibliometrics in their research and candidate evaluations across different evaluative settings, disciplines, and countries. The paper shows that only a minority of scholars saw metrics as a reason for considering something to be the best research. Still, a large majority indicated that metrics were important or partly important in their review of grant proposals and candidate assessments. These findings show that the use of bibliometrics is often pragmatic but is also distinct from research quality, which strengthens the perception that these metrics correspond more to managerial than academic logic. The paper also reveals that reviewers with high bibliometric scores appeared to rely more heavily to metrics in their assessments.

Reymert, Jungblut, and Borlaug (2020): Are Evaluative Cultures National or Global? A Cross-National Study on Evaluative Cultures in Academic Recruitment Processes in Europe

The fourth paper offers a more theoretical explanation for country and disciplinary variance in evaluative cultures and shows how academic disciplines have different evaluative cultures and apply distinct criteria when assessing candidates. We also found that these field-specific preferences were to some extent mediated through national frameworks such as funding systems, but country differences were moderate. Evaluative cultures are thus primarily anchored in disciplinary cultures that are more supranational than nationally embedded.

5 Main Findings

The overarching question of this thesis is how different logics and their interplay influence academic recruitment processes, and how and why their importance varies across time, country, and discipline. In this section I first answer the three sub-questions: 1) explaining the institutional logics in recruitment, 2) explaining disciplinary and country differences, and 3) explaining the use of bibliometrics. I then use these three answers to help explain the dynamics between the two logics over time.

5.1 Institutional Logics in Academic Recruitment

The first sub-question was posed to reveal the dominant institutional logics in academic recruitment over time and was answered by papers 1 and 2. The institutional logics detected in the analyses prescribed different *understandings* of academic recruitment, which types of personnel were to be appointed according to which *criteria*, and who were perceived as legitimate *evaluators*. Figure 2 in paper 1 shows how these theoretical concepts were coded and anchored in the data from the interviews and document analyses using Gioia et al. (2013) coding strategies.

Professorial recruitment proved to be powerfully dominated by academic logic with the purpose of selecting the most academically qualified candidates, based on peer evaluations of their research contributions. The dominance of academic logic in recruitment aligns with prior research (Musselin, 2010) and expectations derived from a historic institutional perspective suggesting that funding logics in core organizational processes are not easily changed (Oertel, 2018).

These processes were also influenced by a managerial logic that regarded recruitment more as a strategic endeavor to secure organizational advantages and cover departmental needs; successful candidates should satisfy those needs and contribute comparative advantages. Here, university leaders and their adherence to managerial logic played a crucial role.

This influence of managerial logic aligns with recent studies (Van den Brink et al., 2013, Paisey and Paisey, 2018) and a neo-institutional perspective suggesting that universities will adapt to managerialism as a dominant logic in their sector. These findings reveal that even core university processes are influenced by managerial logic, although only moderately. I also observed that managerial logic had gained increased influence over these processes; before returning to this argument, I first present all the findings needed to draw that conclusion.

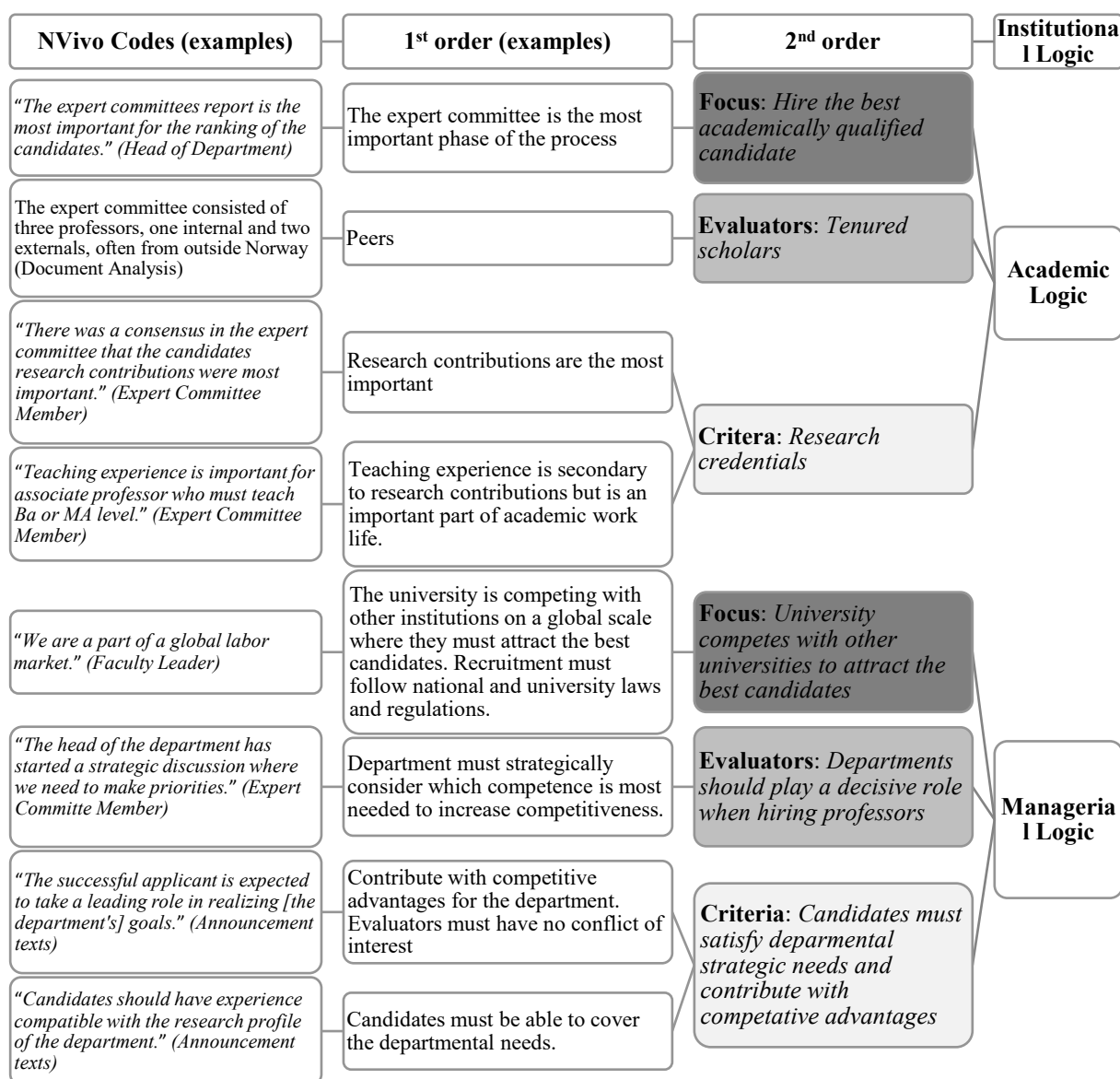


Figure 1. Institutional logics in academic recruitment processes with second-order categories and examples of first-order categories. Source: Paper 1.

Although academic and managerial logics have been described as conflicting (Gumport, 2000, Bleiklie et al., 2017), the analyses showed them to be more harmonious with overlapping elements, as is often found in institutional logics (Noordegraaf, 2007, Meyer et al., 2014). For instance, receiving grants as a signal of academic prestige was an important qualification by academic logic but also aligned with managerial logic by representing external funding and signaling comparative advantage.

Second, even though most actors were guided primarily by one logic, they were influenced by both. For instance, although university leaders were largely guided by managerial

logic, they often referred to the importance of the academic evaluation of the candidate, an aspect of academic logic. Similarly, while scholars were primarily guided by academic logic, they were also concerned with the department's strategic positioning, which reflects managerial logic.

There was still tension between actors and logics, as when departments head argued to overrule the academic ranking of the candidates and offered positions to candidates with better strategic profiles (managerial logic) instead of those ranked highest in peer review (academic logic).

5.1.1 The Sequential Nature of Recruitment Reconciles Institutional Logics

The second part of the first sub-question sought to explain why universities handled this tension the way they do. A major observation was how professorial recruitments were sequential. In these sequential processes the candidates in multiple judgment processes with different actors using different criteria adhering to different institutional logics.

First, the department councils' crafting of announcement texts was dominated by a managerial logic, with departments concerned about achieving strategic goals. If there were many applicants, some departments established a selection committee of internal professors to screen eligible candidates based on their CVs and research records. This was largely a pragmatic endeavor and reflected the bureaucratic approaches favored by managerial logic. After screening the candidates, the department council appointed an expert committee that constituted the backbone of the recruitment process and consisted of internal and external, often international professors. This committee conducted a more thorough evaluation of the candidates' research contributions and was dominated by academic logic. Finally, the highest-ranked candidates were often invited for interviews that included trial lectures. The composition of interview committees differed across departments but usually consisted of an internal professor, the department head, the head of administration, and sometimes a student representative. This committee assessed the candidates' teaching experience and social skills and was dominated by managerial logic, as the departments aimed to select the most strategic candidate who would best fit departmental needs. The job offer decision was finally made at the departmental or faculty level and followed managerial logic. Figure 2 displays the sequential structure of recruitment.

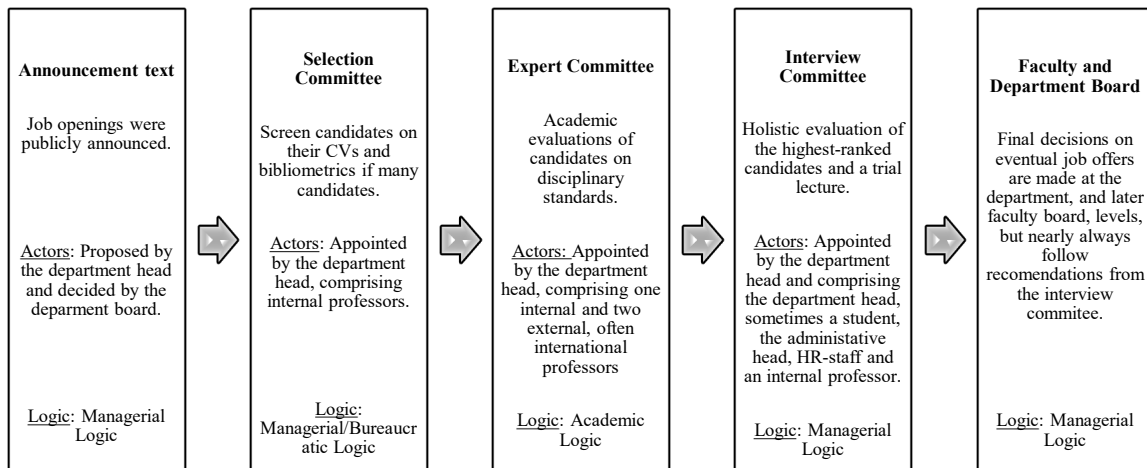


Figure 2. The sequential nature of professorial recruitment at Norwegian universities. Source paper 1.

Figure 3 further illustrates how the committees applied different assessment criteria and reveals the most important criteria used by the various committee types in four disciplines at the University of Oslo between 2000 and 2017. The figure shows that, while bibliometrics were the most important evaluation criterion in selection committees, qualitative research evaluations were most important in expert committees, and teaching skills and personality and administrative skills were most important to interview committees.

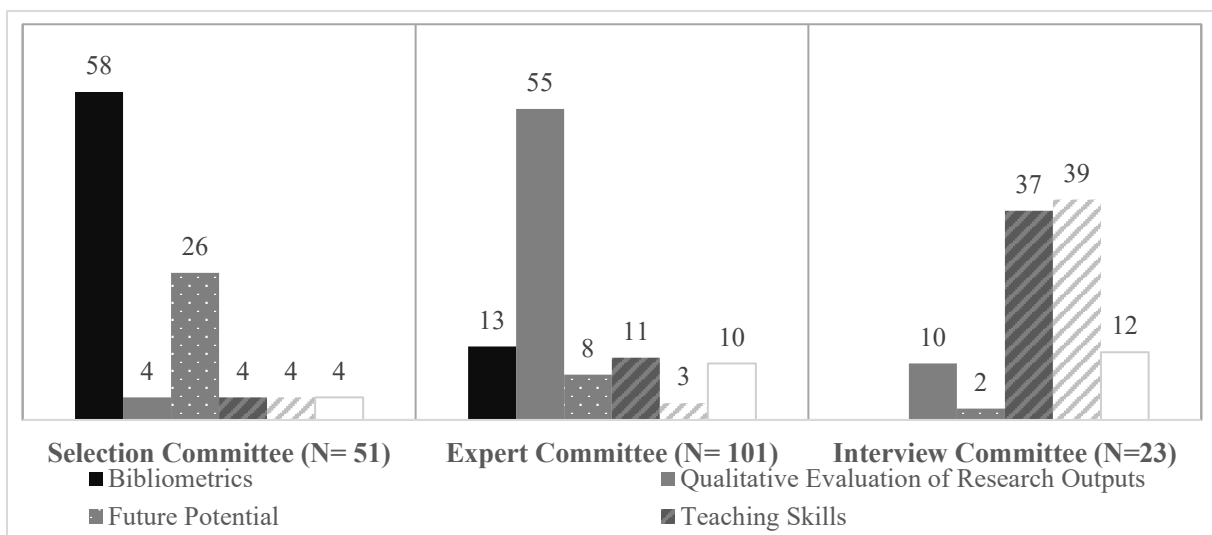


Figure 3. The most important criteria (in percentages) in three different committee types in four fields (economics, sociology, physics, and informatics) at the University of Oslo from 2000 to 2017. N refers to the number of the most important assessment criteria in the different committees. Source paper 2.

How the institutional logics influenced different phases of the process demonstrates that logics not only have unequal importance in various organizational processes (Pache and Santos, 2010, Reay and Hinings, 2009), but also unequal importance in different phases of an organizational process. Separating institutional logics into different phases further decreased the tension and illustrates how compartmentalized strategies (Kraatz and Block, 2008) can occur within organizational processes through sequential decision-making.

Even though sequential decision-making separated the logics, they were not totally isolated from each other; to some extent, they also operated simultaneously. To manage this hybridity within the processes' phases, the actors evoked identity work, balanced different identities, and negotiated the importance of the different logics, as examples of balancing strategies (Kraatz and Block, 2008). However, sequential decision-making decreased potential conflict as a whole, making such strategies less necessary, which represents a novel finding in the literature on hybrid organizations.

Sequential decision-making also proved flexible in satisfying new demands. The analyses revealed that universities were confronted with heightened requirements for external funding, an increased focus on grant success, and growing concerns over teaching quality. These demands sharpened the requirements for new appointees who increasingly had to meet multiple qualifications. These requirements were sometimes incompatible, as the candidate with the best research records were not always the most exemplary team players or best teachers.

Although the universities needed to satisfy these new demands, instructing the expert committee to deviate from traditional academic candidate evaluations was challenging. First, because of how the process is anchored in the cultural foundation of the disciplines and changing the expert committee would imply legitimacy losses. Second, because the expert committee was controlled by scholars as a powerful profession.

Instead of changing the expert committee I detected that the universities increased the emphasis on the new skills in the announcement text and in the final interviews, leaving the expert committees relatively untouched. In this way, they satisfied the new demands without changing the core recruitment processes. Figure 4 shows how sequential decision-making handled these requirements.

At the same time, the importance of the announcement and interview committees increased, resulting in a modest decline in the importance of the expert committee and its academic logic. This development illustrates the historic institutional argument that parallel

structures can gain increased importance at the expense of core processes, which is known as a layering effect (Streeck and Thelen, 2005).

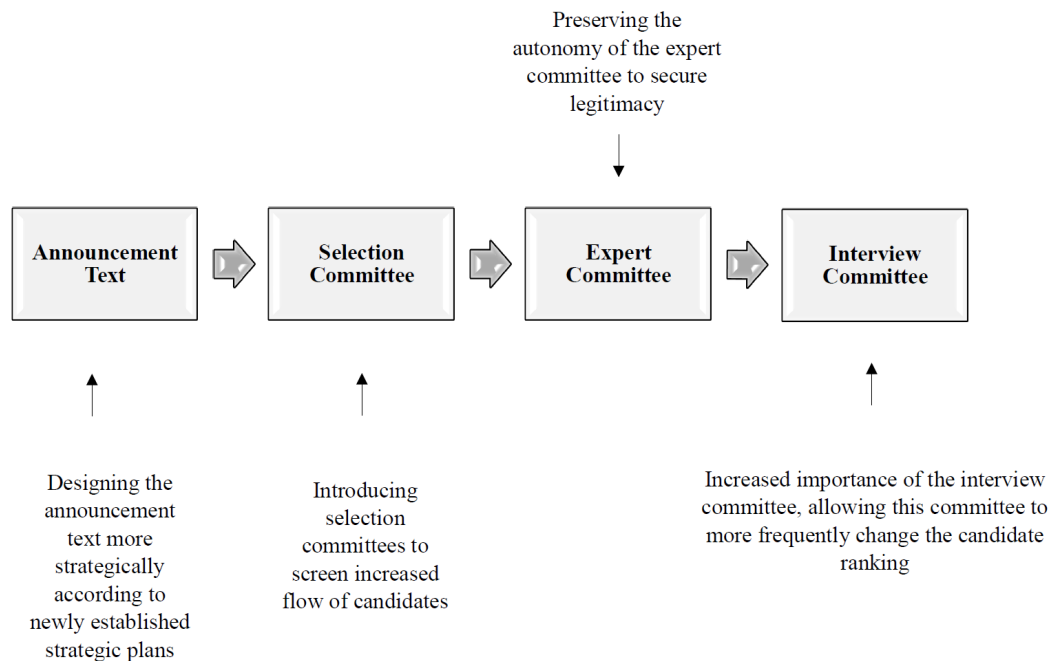


Figure 4 How organizational design is used to handle competing institutional logics. Source Paper 1.

5.2 Discipline-Dependent Recruitment and Moderate Country Differences

The second sub-question asked why these logics unfold differently in various disciplines and countries. While all papers covered disciplinary differences, only papers 3 and 4 discussed national differences.

Disciplinary differences are a major topic in the higher education literature (Becher and Trowler, 1989, Lamont, 2009). My analyses revealed that recruitment also proved to depend highly on discipline, with the various evaluative cultures applying different criteria that reflected their own notions of research quality. For example, whereas economic scholars primarily evaluated candidates on their publication records, physicists focused on their research profiles. Figure 5 illustrates these disciplinary differences by displaying the most important evaluation criteria used by expert committees in four disciplines at the University of Oslo from 2000 to 2017.

These evaluative cultures were not part of explicit guidelines that instructed the evaluators. Rather, they evolved in negotiations between the scholars within the disciplines and

must be conceptualized as cultural aspects of those disciplines. The stability of these evaluative processes could thus be explained by cultural and cognitive feedback mechanisms (Pierson, 1993, Selznick, 1957, March and Olsen, 1989), with evaluators rating the candidates and negotiating research quality within the context of these cultural scripts. The evaluative cultures were also maintained by structural feedback mechanisms (Pierson, 1993). For example, most announcement texts specified that the research contribution was superior to other criteria. Additionally, the very establishment of the expert committee as the backbone of these processes provided scholars with the formal power to decide who was appointed, a structural power that would be hard to remove.

Another implication of these disciplinary differences was how the disciplinary cultures aligned with managerialism; some were more closely tied to this logic than others. Most obvious was how the notion of research quality in economics resonated better with bibliometrics, which is not unexpected considering that NPM reforms and performance-based indicators derive from that field (Hood, 1991, Pollitt and Bouckaert, 2017). Another example of how the historic structure of an organization can explain contemporary decisions is the strategic allocation of new vacancies to specific research areas in the natural sciences, which aligned with established practices in those disciplines (Thelen, 1999).

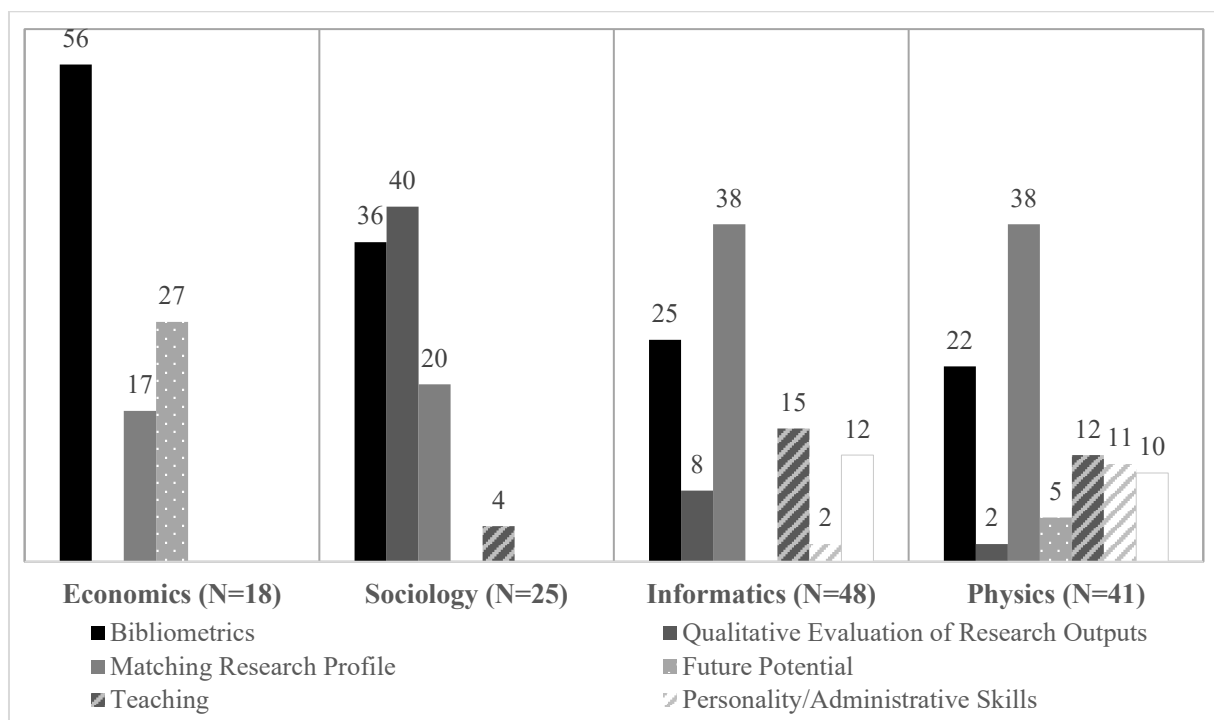


Figure 5. Most important assessment criteria (in percentages) among the expert committees in four academic disciplines at the University of Oslo from 2000 to 2017. N refers to the number of important assessment criteria detected in the expert committee. Source paper 2.

Academic recruitment processes have also been suggested to be nationally sensitive (Musselin, 2010), as both universities (Clark, 1978) and academic labor markets (Musselin, 2005) are nationally embedded. However, I found only moderate country-related differences, with disciplines across countries applying relatively similar evaluative criteria. The limited nature of country differences may be partly explained by the inclusion of only northern European countries, which are relatively similar by global standards. However, these countries do differ in terms of their higher education (Elken, Frølich and Reymert, 2016), funding (Sivertsen 2017), and career systems (Frølich et al., 2018). I also found some national differences; for example, Dutch scholars emphasized publication numbers less than others when asked to identify the most important evaluation criterion. The fact that bibliometrics are not included in the Dutch performance-based funding system (Sivertsen 2017) could explain this difference, as there are fewer incentives to ensure certain publication patterns among newly hired staff, which reflects a structural feedback mechanism (Pierson, 1993).

Figure 6 displays which criteria were most important for academics who has served as evaluators in recruitment in economics and physics in five different countries. The figure shows that overall, the country differences are moderate.

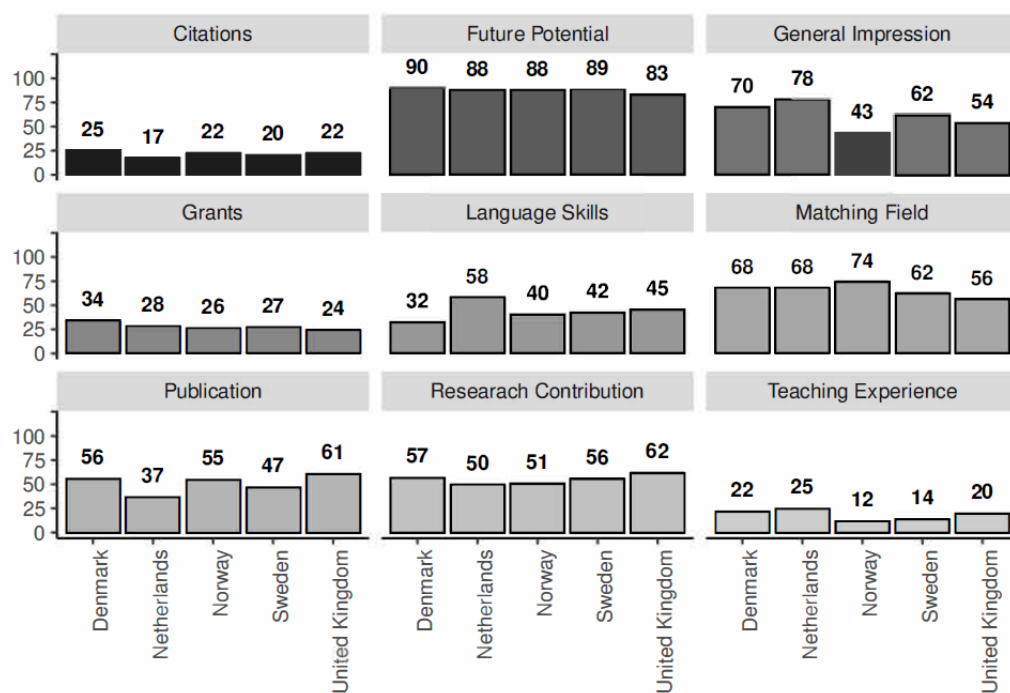


Figure 6. Highly important evaluative criteria in the most recent recruitment reported by physicists and economists by country; N = 168 per country. Source paper 4.

5.4 Bibliometrics as an Indication of Managerial Logic

The third sub-question asked how bibliometrics were applied in academic recruitment, how to explain this use, and which institutional logics these measures represent. These questions were mainly answered by papers 2 and 3.

Scholars have found that metrics are increasingly used in recruitments (Hicks et al., 2015, Stephan et al., 2017), but my study found a more moderate use of bibliometrics in evaluating candidates. Bibliometrics had not replaced traditional peer review but rather supplemented the established evaluation practices and were primarily used as a judgment tool to screen eligible candidates in processes with large numbers of applicants.

This pragmatic use of bibliometrics could be explained by how bibliometrics are becoming increasingly legitimate due to their general proliferation in science. Other scholars' use of bibliometrics in other recruitment or research evaluation processes could well have stimulated such use. Normative isomorphism (DiMaggio and Powell, 1983) is also a possible trajectory for the increased use of bibliometrics, as the expert committees included external professors who could have brought the use of bibliometrics with them.

Another explanation for the moderate increase in bibliometrics use could be the introduction of performance-based funding models that use those measures to calculate university funding. These funding models use of bibliometrics have shown to cause trickle-down effects and an increased use of bibliometrics (Hicks, 2012, Sivertsen, 2017, Aagaard, 2015, Maassen, Moen and Stensaker, 2011) and as an example of coercive isomorphism (DiMaggio and Powell, 1983). However, bibliometrics were applied in Norwegian recruitment before the implementation of performance-based funding models (Sivertsen, 2016), which makes the connection between bibliometrics and funding decisions more tenuous.

In the analyses I observed that the use of bibliometrics reflected both academic and managerial logics. First, bibliometrics resonated with academic logic's emphasis on individual research contributions. Especially in some disciplines, these measures matched established notions of research quality; for instance, publishing in top journals is regarded as an important indication of research quality in economics (Hylmö, 2018). However, when asked to think about the best research in their field, only a minority of the scholars mentioned bibliometrics as a reason for their view (see Figure 7)¹. This finding indicates that bibliometrics is not perceived as identical with research quality.

¹ This question was phrased as follows: Think about the research you consider to be the best in your specific field/specialty. Why do you consider this the best research?

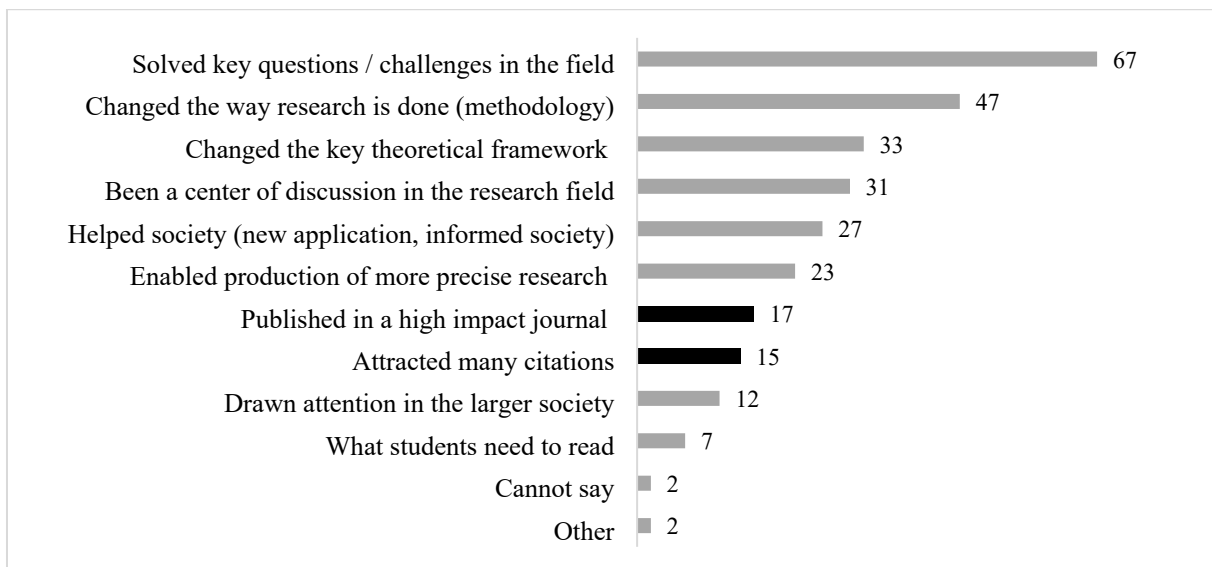


Figure 7 Reasons for considering something the best research in the field among cardiologists, economists, and physicists (percentages; N=1,621). Source paper 3.

The interviews also showed how these measures enabled non-academics to participate in quality evaluations and thus overcame the academic monopoly on such quality evaluations. The increased reliance on bibliometrics was not merely an attempt to reduce complexity in recruitment; it must also be seen in relation to the more general spread of objective standards in organizations (Jacobsson and Sahlin-Andersson, 2006, Power, 1999). Additionally, the use of these measures structures ways of thinking about research quality, directing attention to quantitative output instead content. Thus, despite its limited influence and reflecting both logics, the use of bibliometrics indicate a moderate increase in the importance of managerial logic in recruitment processes.

The analyses further revealed that different disciplines adhered to different types of metrics that resonated with their own notions of research quality. Whereas the social sciences chiefly preferred bibliometrics involving publication volume and journal quality, the natural sciences relied more on impact metrics and the number of conference proceedings. This demonstrates that disciplines have not passively adopted bibliometrics as standards, as isomorphism would suggest. Instead, they more actively translated bibliometrics to fit their established evaluation cultures and narrow any disparity between bibliometrics and their traditional notion of research quality, as a translation perspective would suggest; this shows that the use of bibliometrics is influenced by both external and internal forces.

5.5 The Increased Influence of Managerial Logic

The main aim of this thesis has been to investigate and explain the dynamics between academic and managerial logics in academic recruitment and how and why their importance varies across time, countries, and disciplines. While section 5.2 explained variance over disciplines and countries, I now concentrate on explaining development over time. These results are drawn from the two first papers and thus cover only the Norwegian case study.

As discussed, my analyses revealed largely stable recruitment processes, which are explained by their roots in robust evaluative traditions maintained by cultural and structural feedback mechanisms (Pierson, 1993). I also observed an increased influence of managerial logic, as suggested by a neo-institutional perspective on how university processes adapt to managerialism as a powerful logic in their sector.

The increased managerialism in professorial recruitment is based on four main observations. First, I observe in paper 1 that *sequential decision-making* altered the balance between the two logics by increasing the importance of the more strategic phases, which were dominated by managerial logic. Second, I also show in that paper how *all actors adhered to managerial logic*, including scholars. For example, when academics evaluated candidates, they also assessed who would most bolster departmental competitiveness in addition to their more scholarly evaluations. The way professionals adapt to managerialism have been observed by other scholars (Noordegraaf, 2020, Noordegraaf, 2007) and represent an example of change as displacement (Streeck and Thelen, 2005), by which managerial logic gradually emerges as more influential among all actors. Third, even though the *increased use of bibliometrics* was more moderate, as I observed in the second paper, this use still made the view of research quality more managerial. Finally, in the first paper, I documented an *increased formalization of recruitment*, with more frequent use of guidelines and protocols, reflecting how universities are becoming stronger organizational actors (Krücken and Meier, 2006, Drori et al., 2006, Ramirez, 2006). This formalization was, however, not opposed by scholars, as other studies have suggested (Paisey and Paisey, 2018, Van den Brink et al., 2013). Instead, this development was welcomed by the scholars, who perceived this attention to guidelines as valuable for ensuring progress in these sometimes tedious processes.

This development and the increased influence of managerial logic in Norwegian recruitment can be explained by external pressure in the institutional field, where universities were confronted with heightened requirements and simply had to adapt. The growing numbers of foreign applicants due to the internationalization of Norwegian recruitment required new

methods of screening candidates to reach a manageable number, and the increased influence of bibliometrics made these new measures both more able and more legitimate to solve that problem. The increased focus on teaching experience in Norwegian recruitment could similarly be explained by an increased focus on teaching skills in higher education generally (Levander, 2017). Moreover, the increased focus on guidelines followed a general rationalization tendency by which Norwegian universities and other public sector institutions emerged as stronger organizational actors with clearer goals, mission statements, and formalized procedures (Ramirez, 2006, Krücken and Meier, 2006, Ramirez and Christensen, 2013). Thus, the observed institutional changes are largely explained by changes in the institutional field, as neo-institutional scholars have proposed.

The Norwegian recruitment processes also proved highly stable, and the academic logic was still very influential, perhaps the most influential in recruitment. This stability can be attributed to the robust disciplinary traditions that were maintained by both cultural and structural feedback mechanisms. Hence these recruitment processes were neither radically changed nor immutably stable.

As these recruitments were constantly affected by stability and change, they evolved incrementally through both displacements processes, as all actors including scholars increasingly adhered to managerial logic, and layering effects, as the more strategic phases of the processes gained more influence. I also observed that even as disciplines adopted bibliometrics in their candidate evaluations, they translated them to fit their established cultures. Hence, this thesis has provided a rich and detailed account of the dynamics between change and stability forces and how even core processes that are most resistant are in fact changing; however, this change is affected by the robustness of these organizations.

6 Implications

This thesis has provided inside knowledge of what occurs during professorial recruitment, which is often perceived as a black box from which little information can be extracted. Additionally, the thesis has accounted for national differences which has been scarcely covered by prior studies. The thesis has shown that academic recruitment processes are highly robust, but also incrementally adapting to changes in their environment. Thus, this thesis contributes a detailed and theoretical account of how robust core organizational processes respond to environmental pressure but are also preserved by feedback mechanisms.

The study illustrates that universities are special organizations (Musselin, 2006) that are still typical examples of professional bureaucracies (Mintzberg, 1983) strongly dominated by professionalism. At the same time, the study reveals that universities are increasingly emerging as managerial institutions (Deem, 1998, Gumpert, 2000) and that this transformation even affects their core processes, though only moderately.

Universities are thus robust but also flexible. For instance, faced with increased demands for teaching quality, universities demonstrated their ability, or at least willingness, to strengthen their teaching experience requirements when appointing professors. This development could be especially favorable for students if it results in better education. I also observed that more professionalized recruitment processes have decreased the time spent on them, which was warmly welcomed by both scholars and university leaders.

The changes also pose some challenges. An increased focus on candidates' grant records in recruitment, for example, could spur increased inequality among researchers through a Matthew effect (Merton, 1968), whereby prior grant successes not only increase scholars' future chances of grant successes (Madsen and Aagaard, 2020) but also their employment advancement opportunities. The increased importance of bibliometrics in recruitment could exacerbate the favoring of a small group of successful researchers who lead in bibliometric parameters (Barrier and Musselin, 2009). This development could also spur epistemic homogenization since having high bibliometrics scores not only depends on being an excellent researcher but also on whether one's research topics and methods are favored by high-ranked journals (Lee et al., 2013). An increased focus on bibliometric output could also stop researchers from pursuing riskier ideas out of fear of losing the publications needed to excel in future recruitment, which could be especially challenging for young untenured scholars who face uncertain career prospects (Laudel, 2017).

On the other hand, relying on grant success and bibliometrics in recruitment could simply represent an effective method of selecting the best candidates and a more objective way of evaluating candidates that helps avoid the bias in qualitative evaluations. The effects of bibliometrics and grant success in recruitment are not obvious and should be more closely investigated. Yet, in the absence of clear knowledge of such effects, universities should be alert to eventual academic homogenization and ensure diversity in epistemic traditions as precaution.

This thesis began by discussing the organizing principles in public organizations. Many public organizations have been regarded as professional bureaucracies (Mintzberg, 1983) but have in recent decades been confronted with managerial logic (Hood, 1991) and have evolved into hybrid organizations displaying both managerial and professional logics (Bleiklie et al., 2017, Noordegraaf and Van der Meulen, 2008). The thesis shows how the interplay between these two logics plays out in practice and shows that even though these logics conflict at an institutional level they cause less tension on the shop floors of these organizations.

The relationship between managers and professionals is more nuanced than just a dualistic rivalry. First, professionals have taken managerial positions in which they balance managerial and professional logics, as other scholars note (Noordegraaf and Van der Meulen, 2008, Spyridonidis, Hendy and Barlow, 2015, Sartirana, Currie and Noordegraaf, 2019, Croft, Currie and Lockett, 2015, Pietilä and Pinheiro, 2020, Shams, 2020). Managerial logic not only guides managers but also influences professionals, and the two logics contain overlapping elements. We should thus move beyond a crude understanding of managerial-professional opposition in public organizations to a more nuanced and interwoven view of how these logics relate.

Despite the many studies that document multiple institutional logics in recruitment in public organizations (Lægreid and Wise, 2015, Nigro and Kellough, 2008, Sundell, 2014, Lavigna and Hays, 2004, Poocharoen and Lee, 2013), less attention has been paid to how these logics are handled. Thus, a major contribution of this thesis is to help fill this research gap and show how sequential decision-making reconciles multiple logics.

This thesis has only studied the interplay between professional and managerial logics in academic recruitment, limiting the range of generalizations. However, as discussed, these recruitment processes represent the least likely in which to expect increased managerialism (George and Bennett, 2005). Thus, this study's findings do have some general implications for the proliferation of managerialism throughout the public sector.

Even though I detected only a moderate and incremental reliance on managerialism in these processes and academic logic remained dominant, managerialism still enjoyed an

increased influence, which suggests that managerialism has penetrated even the strongholds of professionalism in the public sector and provides another example of declining professional logic. If the least likely case is influenced by managerialism, the same is almost certainly true of most processes in public organizations.

There is however the question of whether pockets of purely professional control ever existed, as most organizations have always been subjected to competing logics (Greenwood et al., 2011). Hence, the major implication of these findings is the spread of managerialism in public organizations, not that organizations are managing multiple logics.

In any case, academic logic proved very strong in recruitment, and another key result in the thesis is that professionalism is far from dead but is still a highly influential logic in the public sector, despite NPM reforms. This finding also says something about the limits of proliferation of managerialism and NPM reforms when they are faced with strong professions in public organizations.

Finally, the thesis also offers implications for practice. First, decision-makers should be aware that professor recruitment unfolds very differently in the disciplines, each of which has its own criteria, needs, and labor markets. Moreover, to successfully implement new requirements like increased teaching experience into professorial recruitment, policymakers should integrate the new policies into announcement texts and interview committees instead of altering peer review processes that are resistant to change; excessive changes to peer review could also cause legitimacy losses. Similarly, the academic profession and its unions should be aware that changes in announcement texts and interview committees, where their profession has less control, could dilute their organizational power even if peer review remains untouched.

7 Suggestions for Future Research

This thesis has contributed important findings to different research streams and posed new questions that require further investigation. I now address four strands of future research that I encourage scholars to pursue.

First, exploring the dynamics between professional and managerial logics in recruitment has proven to be a rich empirical lens for examining how the dynamics between these logics play out and are handled on the shop floor of public organizations. I thus urge future scholars to study this interplay in recruitment in other public organizations like hospitals and explain differences and similarities. I especially encourage scholars to investigate how these logics are handled, which has been scarcely covered by the literature, and determine whether sequential decision-making also reconciles logics in other recruitment.

Second, recruitment has proved to be a very good empirical lens on organizational culture and should be more frequently used by scholars aiming to uncovering institutional cultures. When describing others, we expose ourselves, and choosing strategic personnel reveals an institution's culture.

Third, in relation to the higher education literature, governmental steering of academic recruitment should be more closely scrutinized. Although academic career structures are heavily debated among politicians and interest groups, with policymakers seeking to increase the emphasis on teaching recruitments (Frølich et al., 2018, Christiansen, 2016, Benner, 2016), my analyses suggest that national regulations are an explanation for the developments observed since regulations are often implemented after such developments are detected. These national regulations could of course reinforce these ongoing developments but may not be their cause. Hence, a pressing issue is the effect these national policies have on academic recruitment, which is the universities' jurisdiction.

Finally, it remains difficult to detect the consequences of using bibliometrics and the focus on grant proposals in recruitment, especially as to epistemological homogenization and academic elites. I thus urge scholars to investigate these effects, despite the methodological challenges.

8 References

- Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*. The University of Chicago Press.
- Agresti, A. (2013). *Categorical data analysis*. New Jersey: John Wiley & Sons.
- Aksnes, D. W., Langfeldt, L. and Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*. 9(1).
<https://doi.org/10.1177/2158244019829575>
- Alfonso, A. (2016). Varieties of academic labor markets in Europe. *PS: Political Science & Politics*. 49(4), pp. 816-821.
- Altbach, P. G., Yudkevich, M. and Rumbley, L. E. (2015). Academic inbreeding: local challenge, global problem. *Asia Pacific Education Review*. 16(3), pp. 317-330. <https://10.1007/s12564-015-9391-8>
- Aschhoff, N. and Vogel, R. (2019). Something old, something new, something borrowed: Explaining varieties of professionalism in citizen collaboration through identity theory. *Public Administration*. 97, pp. 703-720. <https://doi.org/10.1111/padm.12589>
- Barrier, J. and Musselin, C. (2009). The rationalization of academic work and careers: Ongoing transformations of the profession and policy challenges. In: Kehm, B. M., Huisman, J. and Stensaker, B. (eds.) *The European Higher Education Area*. Brill Sense. pp. 203-221.
https://doi.org/10.1163/9789087907143_012
- Becher, T. and Trowler, P. R. (1989). *Academic Tribes and Territories*. USA/UK: The Society for Research into Higher Education and Open University Press.
- Benner, M. (2016). Karriärstrukturen vid svenska universitet och högskolor - en historik [The career structure at Swedish universities and colleges - a history]. In: SOU 2016:29. *Trygghet och attraktivitet - en forskerkarriär för framtiden [Security and attractiveness - a research career for the future]*. Stockholm, Sweden: Ministry of Education.
- Bleiklie, I., Enders, J. and Lepori, B. (2013). Introduction: Transformation of universities in Europe. *Higher Education*. 65(1), pp. 1-4. <https://doi.org/10.1007/s10734-012-9577-5>
- Bleiklie, I., Enders, J. and Lepori, B. (2017). *Managing Universities*. Palgrave Macmillan.
<https://doi.org/10.1007/978-3-319-53865-5>
- Bleiklie, I., Enders, J., Lepori, B. and Musselin, C. (2011). New public management, network governance and the university as a changing professional organization. In: Christensen, T. and Lægreid, P. (eds.) *The Ashgate Research Companion to New Public Management*. Farnham, U. K.: Ashgate. pp. 161-176.
- Bovens, M. (2005). Public Accountability. In: Ferlie, E., Lynn, L. E. and Pollitt, C. (eds.) *The Oxford Handbook of Public Management*. Oxford: Oxford pp. 183-208.
- Bromley, P. and Meyer, J. W. (2015). *Hyper-organization. Global organizational expansion*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199689859.001.0001>

- Brunsson, N. (2003). Organized Hypocrisy. In: Czarniawska, B. and Sevón, G. (eds.) *Northern Lights*. Oslo: Liber, Abstrakt, Copenhagen Business School Press. pp. 201-222.
- Brunsson, N. and Sahlin-Andersson, K. (2000). Constructing Organizations: The Example of Public Sector Reform." *Organization Studies*, 21(4), 721-746. *Organization Studies*. 21(4), pp. 721-746. <https://doi.org/10.1177/0170840600214003>
- Buckner, E. S. (2017). The changing discourse on higher education and the nation-state, 1960–2010. *Higher Education*. 74(3), pp. 473-489. <https://doi.org/10.1007/s10734-016-0056-2>
- Christensen, T. (1991). *Virksomhetsplanlegging: myteskaping eller instrumentell problemløsning? [Organizational planning: myth-making or instrumental problem-solving]*. Oslo: Universitetsforlaget.
- Christensen, T. and Klemsdal, L. (2019). Social Science Disciplines in Complex Development Contexts - the Professional Dimension of Reputation Management. In: Christensen, T., Gornitzka, Å. and Ramirez, F. (eds.) *Universities as Agencies*. Palgrave Macmillan. pp. 117 - 154.
- Christensen, T. and Lægreid, P. (2010). Introduction. In: Christensen, T. and Lægreid, P. (eds.) *The Ashgate Research Companion to New Public Management*. USA: Ashgate Publishing. pp. 1-17.
- Christiansen, F. V. (2016). Stillingsstrukturens betydning for samspillet mellem forskning og undervisning [The significance of the position structure course for the interplay between research and teaching]. *Dansk Universitetspædagogisk Tidsskrift* 21, pp. 1-43.
- Clark, B. (1978). Academic Differentiation in National Systems of Higher Education. *Comparative Education Review*. 22(2), pp. 242-258. <https://www.jstor.org/stable/1187674>
- Clark, B. (1983). *The higher education system*. Berkeley, California: University of California Press.
- Clark, B. (1998). The Entrepreneurial University: New Foundations for Collegiality, Autonomy, and Achievement. *Higher Education*. 13(2), pp. 9-24.
- Colyvas, J. A. and Powell, W. W. (2006). Roads to institutionalization: The remaking of boundaries between public and private science. *Research in Organizational Behavior*. 27, pp. 305-353. [https://doi.org/10.1016/S0191-3085\(06\)27008-4](https://doi.org/10.1016/S0191-3085(06)27008-4)
- Creswell, J. W. and Clark, W. L. P. (2017). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, United States: SAGE Publications Inc.
- Croft, C., Currie, G. and Lockett, A. (2015). Broken ‘two-way windows’? An exploration of professional hybrids. *Public Administration*. 93(2), pp. 380-394. <https://doi.org/10.1111/padm.12115>
- Cyert, R. M. and March, J. G. (1963). *A Behavioral Theory of the Firm* Prentice-Hall: Englewood Cliffs, N.J. Available at SSRN: <https://ssrn.com/abstract=1496208>
- Czarniawska, B. and Sevón, G. (1996). *Translating Organizational Change*. New York: Walter de Gruyter. <https://doi.org/10.1515/9783110879735>

- Deem, R. (1998). 'New managerialism' and higher education: The management of performances and cultures in universities in the United Kingdom. *International Studies in Sociology of Education*. 8(1), pp. 47-70. <https://doi.org/10.1080/0962021980020014>
- Denis, J.-L., Ferlie, E. and Van Gestel, N. (2015). Understanding Hybridity in Public Organizations. *Public Administration*. 93(2), pp. 273-289. <https://doi.org/10.1111/padm.12175>
- DiMaggio, P. J. and Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organization fields. *American Sociological Review*. 48(2), pp. 147-160.
- Driori, G., Meyer, J. W. and Ramirez, F. (2003). *Science in the modern world polity: institutionalization and globalization*. Stanford: Stanford University Press.
- Drori, G. S., Meyer, J. W. and Hwang, H. (2006). *Globalization and Organization, World Society and Change*. Oxford University Press.
- Dunn, M. B. and Jones, C. (2010). Institutional logics and institutional pluralism: The contestation of care and science logics in medical education, 1967–2005. *Administrative Science Quarterly*. 55(1), pp. 114-149. <https://doi.org/10.2189/asqu.2010.55.1.114>
- Elken, M., Frølich, N. and Reymert, I. (2016). Steering approaches in higher education. *NIFU Report*. 35. <http://hdl.handle.net/11250/2432050>
- Ferlie, E. and Geraghty, K. (2005). Professionals in public services organizations. Implications for public sector "Reforming". In: Ferlie, E., Lynn, L. E. and Pollitt, C. (eds.) *The Oxford Handbook of Public Management*. United Kingdom: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199226443.003.0019>
- Fossestøl, K., Breit, E., Andreassen, T. A. and Klemsdal, L. (2015). Managing Institutional complexity in Public Sector Reform: Hybridization in Front-line service organizations. *Public Administration*. 93(2), pp. 290-306. <https://doi.org/10.1111/padm.12144>
- Fourcade, M., Ollion, E. and Algan, Y. (2015). The Superiority of Economists. *Journal of Economic Perspectives*. 29(1), pp. 89-114. <https://doi.org/10.1257/jep.29.1.89>
- Freidson, E. (2001). *Professionalism, the third logic: On the practice of knowledge*. University of Chicago press.
- Friedland, R. and Alford, R. (1991). Bringing Society Back In: Symbols, Practices, and Institutional Contradictions. In: Powell, W. W. and DiMaggio, P. J. (eds.) *The New Institutionalism in organizational Analysis*. Chicago, IL: University of Chicago Press. pp. 232-263.
- Frølich, N., Wendt, K., Reymert, I., Tellmann, S., Elken, M., Kyvik, S., Vabø, A. and Larsen, A. (2018). Academic career structures in Europe: Perspective from Norway, Denmark, Sweden, Finland, the Netherlands, Austria and the UK. *NIFU Report*. 4. <http://hdl.handle.net/11250/2487666>
- George, A. L. and Bennett, A. (2005). *Case Studies and Theory Development in the Social Sciences*. Cambridge, Massachusetts: Belfer Center for Science and International Affairs.

- Gioia, D. A., Corley, K. G. and Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational research methods*. 16(1), pp. 15-31. <https://doi.org/10.1177/1094428112452151>
- Greenwood, R., Raynard, M., Kodeih, F., Micelotta, E. R. and Lounsbury, M. (2011). Institutional complexity and organizational response. *The Academy of Management Annals*. 5(1), pp. 317-371. <https://doi.org/10.1080/19416520.2011.590299>
- Grimmer, J. (2010). A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases. *Political Analysis*. 18(1), pp. 1-35. <https://doi.org/10.1093/pan/mpp034>
- Gumport, P. J. (2000). Academic Restructuring: Organizational Change and Institutional Imperatives. *Higher Education*. 39(1), pp. 67-91. <https://doi.org/10.1023/A:1003859026301>
- Hamann, J. (2019). The making of professors: Assessment and recognition in academic recruitment. *Social Studies of Science*. 49(6), pp. 919-941. <https://doi.org/10.1177/0306312719880017>
- Hammarfelt, B. and Rushforth, A. (2017). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation*. 26(3), pp. 169-180. <https://doi.org/10.1093/reseval/rvx018>
- Hannan, M. T. and Freeman, J. (1984). Structural Inertia and Organizational Change. *American Sociological Review*. 49(2), pp. 149-164. <https://doi.org/10.2307/2095567>
- Hansen, H. F., Geschwind, L., Kivistö, J., Pekkola, E., Pinheiro, R. and Pulkkinen, K. (2019). Balancing accountability and trust: university reforms in the Nordic countries. *Higher Education*. 78(3), pp. 557-573. <https://doi.org/10.1007/s10734-019-0358-2>
- Herschberg, C., Benschop, Y. and Van Den Brink, M. (2018). Selecting early-career researchers: the influence of discourses of internationalisation and excellence on formal and applied selection criteria in academia. *Higher Education*. 76, pp. 807-825. <https://doi.org/10.1007/s10734-018-0237-2>
- Hessels, L. K., Franssen, T., Scholten, W. and De Rijcke, S. (2019). Variation in Valuation: How Research Groups Accumulate Credibility in Four Epistemic Cultures. *Minerva*. 57(2), pp. 127-149. <https://doi.org/10.1007/s11024-018-09366-x>
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*. 41(2), pp. 251-261. <https://doi.org/10.1016/j.respol.2011.09.007>
- Hicks, D., Wouters, L., Waltman, S. D. R. and Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*. 22(April). <http://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351>
- Hood, C. (1991). A Public Management for all Seasons? *Public Administration*. 69(1), pp. 3-19. <https://doi.org/10.1111/j.1467-9299.1991.tb00779.x>
- Husu, L. (2000). Gender Discrimination in the Promised Land of Gender Equality. *Higher Education in Europe*. 25(2), pp. 221-228. <https://doi.org/10.1080/713669257>

- Hylmö, A. (2018). *Disciplined Reasoning: Styles of reasoning and the mainstream-heterodoxy divide in Swedish economics*. Doctoral Dissertation. Lund: Lund University.
- Jacobsson, B. and Sahlin-Andersson, K. (2006). Dynamics of soft regulations. In: Sahlin-Andersson, K. and Djelic, M.-L. (eds.) *Transnational Governance: Institutional Dynamics of Regulation*. Cambridge: Cambridge University Press. pp. 247-265.
<https://doi.org/10.1017/CBO9780511488665.013>
- Johnson, B. and Gray, R. (2010). A history of philosophical and theoretical issues for mixed methods research. In: Tashakkori, A. and Teddlie, C. (eds.) *SAGE Handbook of mixed methods in social and behavioral research*. SAGE. pp. 69-94.
- Johnson, R. B. and Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*. 33(7), pp. 14-26.
<https://doi.org/10.3102/0013189x033007014>
- Johnson, R. B., Onwuegbuzie, A. J. and Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*. 1(2), pp. 112-133.
 10.1177/1558689806298224
- Krippendorff, K. (2013). *Content Analysis And Introduction to Its Methodology*. USA: SAGE Publications Inc.
- Krücken, G. and Meier, F. (2006). Turning the University into an Organizational Actor In: Driori, G., Meyer, R. and Hwang, H. (eds.) *Globalization and Organization. World Society and Organizational Change*. New York: Oxford University Press.
- Kraatz, M. S. and Block, E. S. (2008). Organisational implications of institutional pluralism. In: Greenwood, R., Oliver, C., Lawrence, T. B. and Meyer, R. E. (eds.) *The Sage handbook of organisational institutionalism*. SAGE. pp. 243-275.
- Lamont, M. (2009). *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.
- Langfeldt, L., Nedeava, M., Sörlin, S. and Thomas, D. (2019). Co-existing Notions of Research Quality: A Framework to Study Context-specific Understandings of Good Research. *Minerva*. 58pp. 115-137. <https://doi.org/10.1007/s11024-019-09385-2>
- Laudel, G. (2017). How do National Career Systems Promote or Hinder the Emergence of New Research Lines? *Minerva*. 55pp. 341-369. <https://doi.org/10.1007/s11024-017-9314-4>
- Lavigna, R. J. and Hays, S. W. (2004). Recruitment and Selection of Public Workers: An International Compendium of Modern Trends and Practices. *Public Personnel Management*. 33(3), pp. 237-253. <https://doi.org/10.1177/009102600403300301>
- Lee, F. S., Pham, X. and Gu, G. (2013). The UK Research Assessment Exercise and the narrowing of UK economics. *Cambridge Journal of Economics*. 37(4), pp. 693-717.
<https://doi.org/10.1093/cje/bet031>

- Levander, S. (2017). *Den pedagogiska skickligheten och akademins väktare. Kollegial dedikering vid rekrytering av universitetslärare*. Doctoral Dissertation. Uppsala: Uppsala University.
- Levander, S., Forsberg, E. and Elmgren, M. (2019). The meaning-making of educational proficiency in academic hiring: a bland spot in the black box. *Teaching in Higher Education*. 25pp. 541-559. <https://doi.org/10.1080/13562517.2019.1576605>
- Lin, F. (2008). Solving Multicollinearity in the Process of Fitting Regression Model Using the Nested Estimate Procedure. *Qual Quant* 42pp. 417–26. <https://doi.org/10.1007/s11135-006-9055-1>
- Lindblom, C. E. (1959) The Science of "Muddling Through". *Public Administration Review*. 19(2), pp. 79-88. <https://doi.org/10.2307/973677>
- Lynch, J. F. (2013). Aligning Sampling Strategies with Analytic Goals. In: Mosley, L. (ed.) *Interview Research in Political Science USA*: Cornell University Press.
- Læg Reid, P. and Wise, L. R. (2015). Transitions in Civil Service Systems: Robustness and Flexibility in Human Resource Management. In: Van Der Meer, F. M., Raadschelders, J. C. N. and Toonen, T. a. J. (eds.) *Comparative Civil Service Systems in the 21st Century*. London: Palgrave Macmillan UK. pp. 203-222. https://doi.org/10.1057/9781137491459_11
- Madsen, E. B. and Aagaard, K. (2020). Concentration of Danish research funding on individual researchers and research topics: Patterns and potential drivers. *Quantitative Science Studies*. 1(3), pp. 1159-1181. https://doi.org/10.1162/qss_a_00077
- March, J. G. and Olsen, J. P. (1989). *Rediscovering institutions. The organizational basis of politics*. New York: Free Press.
- Martin, C. J. (2013). Crafting Interviews to Capture Causes and Effects. In: Mosley, L. (ed.) *Interview Research in Political Science*. USA: Cornell University Press.
- McGivern, G., Currie, G., Ferlie, E., Fitzgerald, L. and Waring, J. (2015). Hybrid Manager-professionals' Identity Work: The Maintenance and Hybridization of Medical Professionalism in Managerial Contexts. *Public Administration*. 93(2), pp. 412-432. <https://doi.org/10.1111/padm.12119>
- Merton, R. K. (1968). The Matthew Effect in Science. *Science*. 159(3810), pp. 56-63. <https://doi.org/10.1126/science.159.3810.56>
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. . University of Chicago Press.
- Meyer, J. W., Ramirez, F., Frank, D. J. and Schofer, E. (2007). Higher education as an institution. In: Gumpert, P. (ed.) *Sociology of Higher Education: Contributions and Their Contexts*. Baltimore, MD: Johns Hopkins University Press. pp. 187-221.
- Meyer, J. W. and Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*. 83(2), pp. 340-363. <https://doi.org/10.1086/226550>

- Meyer, R., Egger-Peitler, I., Höllerer, M. A. and Hammerschmid, G. (2014). Of bureaucrats and passionate public managers: Institutional logics, executive identities, and public service motivation. *Public Administration*. 92pp. 861-885. <https://doi.org/10.1111/j.1467-9299.2012.02105.x>
- Mintzberg, H. (1983). *Structure in fives: Designing effective organizations*. New Jersey: Prentice-Hall International Editions.
- Mosley, L. (2013). "Just Talk to People"? Interviews in contemporary political science. In: Mosley, L. (ed.) *Interviews research in Political Science*. Ithaca and London: Cornell University Press.
- Musselin, C. (2005). European academic labor markets in transition. *Higher Education*. 49(1), pp. 135-154. <https://doi.org/10.1007/s10734-004-2918-2>
- Musselin, C. (2006). Are Universities specific organisations? In: Krüken, G., Kosmützky, A. and Torka, M. (eds.) *Towards a Multiversity ? Universities between Global Trends and national Traditions*. Bielfeld: Transcript Verlag. pp. 66-84.
- Musselin, C. (2010). *The Market for Academics [Les marché des universitaires]*. New York, USA: Routledge.
- Musselin, C. (2013a). How peer review empowers the academic profession and university managers: Changes in relationships between the state, universities and the professoriate. *Research Policy*. 42(5), pp. 1165-1173. <https://doi.org/10.1016/j.respol.2013.02.002>
- Musselin, C. (2013b). Redefinition of the relationships between academics and their university. *Higher Education*. 65pp. 25-37. <https://doi.org/10.1007/s10734-012-9579-3>
- Maassen, P., Moen, E. and Stensaker, B. (2011). Reforming higher education in the Netherlands and Norway: the role of the state and national modes of governance. *Policy Studies*. 32(5), pp. 479-495. <https://doi.org/10.1080/01442872.2011.566721>
- Nielsen, M. W. (2016). Limits to meritocracy? Gender in academic recruitment and promotion processes. *Science and Public Policy*. 43(3), pp. 386-399. <https://doi.org/10.1093/scipol/scv052>
- Nigro, L. G. and Kellough, J. E. (2008). Personnel Reform in the States: A Look at Progress Fifteen Years after the Winter Commission. *Public Administration Review*. 68pp. S50-S57. <https://doi.org/10.1111/j.1540-6210.2008.00978.x>
- Noordegraaf, M. (2007). From "Pure" to "Hybrid" Professionalism: Present-Day Professionalism in Ambiguous Public Domains. *Administration & Society*. 39(6), pp. 761-785. <https://doi.org/10.1177/0095399707304434>
- Noordegraaf, M. (2020). Protective or connective professionalism? How connected professionals can (still) act as autonomous and authoritative experts. *Journal of Professions and Organization*. 7(2), pp. 205-223. <https://doi.org/10.1093/jpo/joaa011>

- Noordegraaf, M. and Van Der Meulen, M. (2008). Professional power play: Organizing management in health care. *Public Administration*. 86(4), pp. 1055-1069. <https://doi.org/10.1111/j.1467-9299.2008.00746.x>
- Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic management journal*. 18(S1), pp. 187-206. <http://www.jstor.org/stable/3088216>
- Oertel, S. (2018). The role of imprinting on the adoption of diversity management in German universities. *Public Administration*. 96(1), pp. 104-118. <https://doi.org/10.1111/padm.12384>
- Pache, A.-C. and Santos, F. (2010). When worlds collide: The internal dynamics of organisational responses to conflicting institutional demands. *Academy of Management Journal*. 35(3), pp. 455-476. <https://doi.org/10.5465/amr.35.3.zok455>
- Paisey, C. and Paisey, N. J. (2018). Talent management in academia: the effect of discipline and context on recruitment. *Studies in Higher Education*. 43(7), pp. 1196-1214. <https://doi.org/10.1080/03075079.2016.1239251>
- Pierson, P. (1993). When Effect Becomes Cause: Policy Feedback and Political Change. *World Politics*. 45(4), pp. 595-628. <https://doi.org/10.2307/2950710>
- Pierson, P. (2004). *Politics in time: history, institutions, and social analysis*. Princeton: Princeton University Press.
- Pietilä, M. and Pinheiro, R. (2020). Reaching for different ends through tenure track—institutional logics in university career systems. *Higher Education*. 81pp. 1197-1213. <https://doi.org/10.1007/s10734-020-00606-2>
- Polanyi, M. (2000). The republic of science: Its Political and Economic Theory. *Minerva*. 38pp. 1-32. <http://www.jstor.org/stable/41821153>.
- Pollitt, C. and Bouckaert, G. (2017). *Public management reform: A comparative analysis-into the age of austerity*. . United Kingdom: Oxford University Press.
- Poocharoen, O.-O. and Lee, C. (2013). Talent management in the public sector: A comparative study of Singapore, Malaysia, and Thailand. *Public Management Review*. 15(8), pp. 1185-1207. <https://doi.org/10.1080/14719037.2013.816525>
- Power, M. (1999). *The Audit Society: Rituals of Verification*. Oxford University Press.
- Ramirez, F. (2006). The Rationalization of Universities. In: Djelic, M. and Sahlin, K. (eds.) *Transnational Governance: Institutional Dynamics of Regulation* Cambridge University Press. pp. 225-244. <https://doi.org/10.1017/CBO9780511488665.012>
- Ramirez, F. and Christensen, T. (2013). The formalization of the university: rules, roots and routes. *Higher Education*. 65. pp. 695-708. <https://doi.org/10.1007/s10734-012-9571-y>
- Rao, H. and Kenney, M. (2008). New forms as settlements. In: Greenwood, R., Oliver, C., Sahlin, K. and Suddaby, R. (eds.) *The Sage handbook of organizational institutionalism*. London: Sage.: London: Sage. pp. 352–370. <http://dx.doi.org/10.4135/9781849200387.n14>

- Reay, T. and Hinings, C. R. (2009). Managing the rivalry of competing institutional logics. *Organisation studies*. 30(6), pp. 628-652. <https://doi.org/10.1177/0170840609104803>
- Reissner, S. C. (2019). 'We are this hybrid': Members' search for organizational identity in an institutionalized public-private partnership. *Public Administration*. 97(1), pp. 48-63. <https://doi.org/10.1111/padm.12333>
- Sahlin-Andersson, K. (1996). Imitating by Editing Success: The Construction of Organizational Fields. *Translating Organizational Change*. Berlin; New York: Walter de Gruyter. <https://doi.org/10.1515/9783110879735.69>
- Sanz-Menéndez, L. and Cruz-Castro, L. (2019). University academics' preferences for hiring and promotion systems. *European Journal of Higher Education*. 9(2), pp. 153-171. <https://doi.org/10.1080/21568235.2018.1515029>
- Sartirana, M., Currie, G. and Noordegraaf, M. (2019). Interactive identity work of professionals in management: a hospital case study. *Public Management Review*. 21(8), pp. 1191-1212. <https://doi.org/10.1080/14719037.2018.1549269>
- Sauermann, H. and Stephan, P. (2013). Conflicting logics? A multidimensional view of industrial and academic science. *Organisation studies*. 24(3), pp. 889-909. <https://doi.org/10.1287/orsc.1120.0769>
- Schildt, H. and Perkmann, M. (2017). Organisational settlements: theorizing how organisations respond to institutional complexity. *Journal of Management Inquiry*. 26(2), pp. 139-145. <https://doi.org/10.1177/1056492616670756>
- Scott, W. R. and Davis, G. F. (2007). *Organizations and organizing: Rational, natural, and open system perspectives*. Upper Saddle River, N.J: Pearson Prentice Hall.
- Scott, W. R., Ruef, M., Mendel, P. J. and Caronna, C. A. (2000). *Institutional change and healthcare organizations: From professional dominance to managed care*. Chicago: University of Chicago Press.
- Sehested, K. (2002). How new public management reforms challenge the roles of professionals. *International Journal of Public Administration*. 25(12), pp. 1513-1537. <https://doi.org/10.1081/PAD-120014259>
- Selznick, P. (1949). *TVA and the Grass Roots; A Study in the Sociology of Formal Organization*. Berkeley and Los Angeles: Cambridge University Press. <https://doi.org/10.2307/1950244>
- Selznick, P. (1957). *Leadership in Administration*. California: University of California Press.
- Shams, F. (2020). A Micro-Process Model of Institutional Complexity in Public Hybrid Organisations: Construal of Identity Threats and Mitigation Strategies. *Public Administration*. pp. 1-18. <https://doi.org/10.1111/padm.12704>
- Sivertsen, G. (2016). Publication-Based Funding: The Norwegian Model. In: Ochsner, M., Hug, S. E. and Daniel, H. D. (eds.) *Research Assessment in the Humanities. Towards Criteria and Procedures*. Zürich: Springer Open. pp. 79-90.

- Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications*. 3.
<https://doi.org/10.1057/palcomms.2017.78>
- Skelcher, C. and Smith, S. R. (2015). Theorizing Hybridity: Institutional Logics, Complex Organizations, And Identities: The Case of Nonprofits. *Public Administration*. 93(2), pp. 433-448. <https://doi.org/10.1111/padm.12105>
- Slaughter, S. and Leslie, L. L. (1997). Academic capitalism: Politics, policies, and the entrepreneurial university.
- Smets, M., Morris, T. and Greenwood, R. (2012). From Practice to Field: A Multilevel Model of Practice-Driven Institutional Change. *The Academy of Management Journal*. 55pp. 877-904.
<http://www.jstor.org/stable/23317618>
- Spyridonidis, D., Hendy, J. and Barlow, J. J. P. A. (2015). Understanding hybrid roles: The role of identity processes amongst physicians. *Public Administration*. 93(2), pp. 395-411.
<https://doi.org/10.1111/padm.12114>
- Stephan, P., Veugelers, R. and Wang, J. (2017). Reviewers are blinkered by bibliometrics. *Nature*. 544pp. 411-412. <https://doi.org/10.1038/544411a>
- Streeck, W. and Thelen, K. (2005). Introduction: Institutional Change in Advanced Political Economies. In: Streeck, W. and Thelen, K. (eds.) *Beyond Continuity: Institutional Change in Advanced Political Economies*. New York: Oxford University Press.
- Sundell, A. (2014). Are formal civil service examinations the most meritocratic way to recruit civil servants? Not in all countries. *Public Administration*. 92(2), pp. 440-457.
<https://doi.org/10.1111/padm.12077>
- Talbot, C. (2005). Performance Management. In: Ferlie, E., Lynn, L. E. and Pollitt, C. (eds.) *The Oxford Handbook of Public Management*. United Kingdom: Oxford University Press pp. 491-517.
- Tavares, O., Sin, C. and Lança, V. (2019). Inbreeding and Research Productivity Among Sociology PhD Holders in Portugal. *Minerva*. 57(3), pp. 373-390. <https://doi.org/10.1007/s11024-019-09378-1>
- Teelken, C. (2015). Hybridity, Coping Mechanisms, and Academic Performance Management: Comparing Three Countries. *Public Administration*. 93(2), pp. 307-323.
<https://doi.org/10.1111/padm.12138>
- Thelen, K. (1999). Historical institutionalism in comparative politics. *Annual Review of Political Science*. (2), pp. 369-404. <https://doi.org/10.1146/annurev.polisci.2.1.369>
- Thornton, P. H., Ocasio, W. and Lounsbury, M. (2012). *The Institutional Logics Perspective. A New Approach to Culture, Structure, and Process*. Oxford, UK: Oxford University Press.

- Van Den Brink, M. and Benschop, Y. (2011). Gender practices in the construction of academic excellence: Sheep with five legs. *Organization*. 19(4), pp. 507-524.
<https://doi.org/10.1177/1350508411414293>
- Van Den Brink, M., Brouns, M. and Waslander, S. (2006). Does excellence have a gender? . *Employee Relations*. 28(6), pp. 523-539. <https://doi.org/10.1108/01425450610704470>
- 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), pp. 180-195. <https://doi.org/10.1111/j.1748-8583.2012.00196.x>
- Välilmaa, J. (1998). Culture and identity in higher education research. *Higher Education*. 39(2), pp. 119-138. <https://doi.org/10.1023/A:1003248918874>
- Wedlin, L. and Sahlin, K. (2017). The Imitation and Translation of management Ideas. In: Greenwood, R., Oliver, C., Lawrence, T. B. and Meyer, R. (eds.) *The SAGE Handbook of Organizational Institutionalism*. 55 City Road, London: SAGE Publications Ltd pp. 102-127.
<https://doi.org/10.4135/9781526415066>
- Wennerås, C. and Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*. 387pp. 341-343.
<https://doi.org/10.1038/387341a0>
- Whitley, R. (1984). *The Intellectual and Social organization of the Sciences*. Oxford: Clarendon Press.
- Whitley, R. (2003). Competition and pluralism in the public sciences: the impact of institutional frameworks on the organisation of academic science. *Research Policy*. 32(6), pp. 1015-1029.
[https://doi.org/10.1016/S0048-7333\(02\)00112-9](https://doi.org/10.1016/S0048-7333(02)00112-9)
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J. and Johnson (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. In: Hefce (ed.). <https://doi.org/10.13140/RG.2.1.4929.1363>
- Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods*. Thousand Oaks, United States: SAGE Publications Inc.
- Aagaard, K. (2015). How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy*. 42pp. 725-737. <https://doi.org/10.1093/scipol/scu087>

Part II: Papers

Paper 2: Bibliometrics in academic recruitment: A screening tool rather than a game changer

Ingvild Reymert (single-authored)

Publication status: Published in *Minerva*, 2021, 59, 53–78.

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



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

Ingvild Reymert¹

Published online: 10 September 2020
© 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

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 (Mussetlin 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 disciplines 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, Hylmö's (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 sub-units) 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

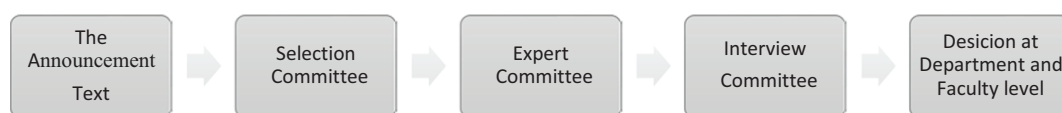


Fig. 1 The Stages in the Recruitment Processes

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 peripheral 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			Total
		2000–2005	2006–2012	2013–2017	
Social sciences	Economics	3	4	4	11
	Sociology	2	4	5	11
Natural sciences	Physics	4	7	5	16
	Informatics	8	3	10	21
<i>Total</i>		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 time². 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	
<i>Qualitative Research Evaluations</i>	Qualitative research evaluations of a candidate's research
<i>Matching Research Profile</i>	Assessment of the match between the candidate's research profile and the specific call
Metrics	The quantitative analysis of scientific and scholarly publications
<i>Future Potential</i>	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
<i>Metrics on journal publications</i>	
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
<i>Metrics on impact and citations</i>	
Citations and h-index	Citation numbers and h-index
<i>Other types of publications</i>	
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.

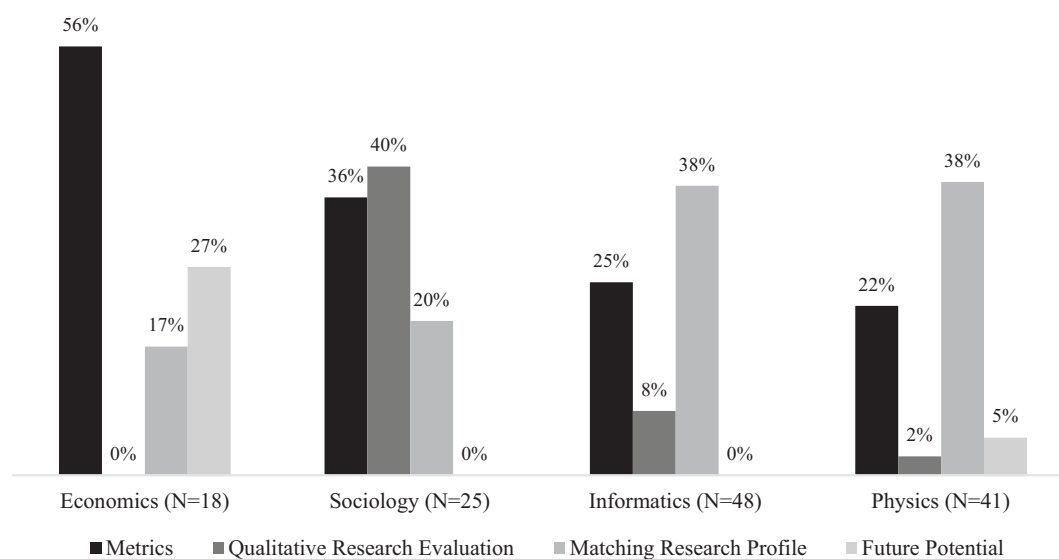


Fig. 2 Most important assessment criteria in the expert committees by academic discipline (percentage). N refers to the number of important assessment criterion/a detected in the expert committee

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*

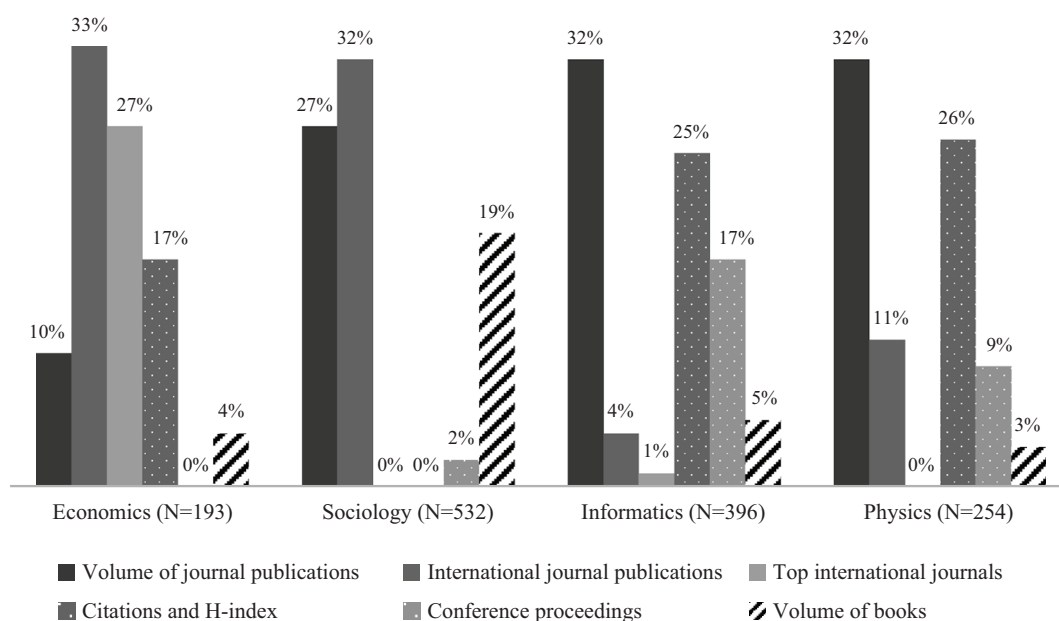


Fig. 3 Types of metrics applied by expert committees in each discipline (percentage). N refers to the number of arguments coded as metrics in the expert committee

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.

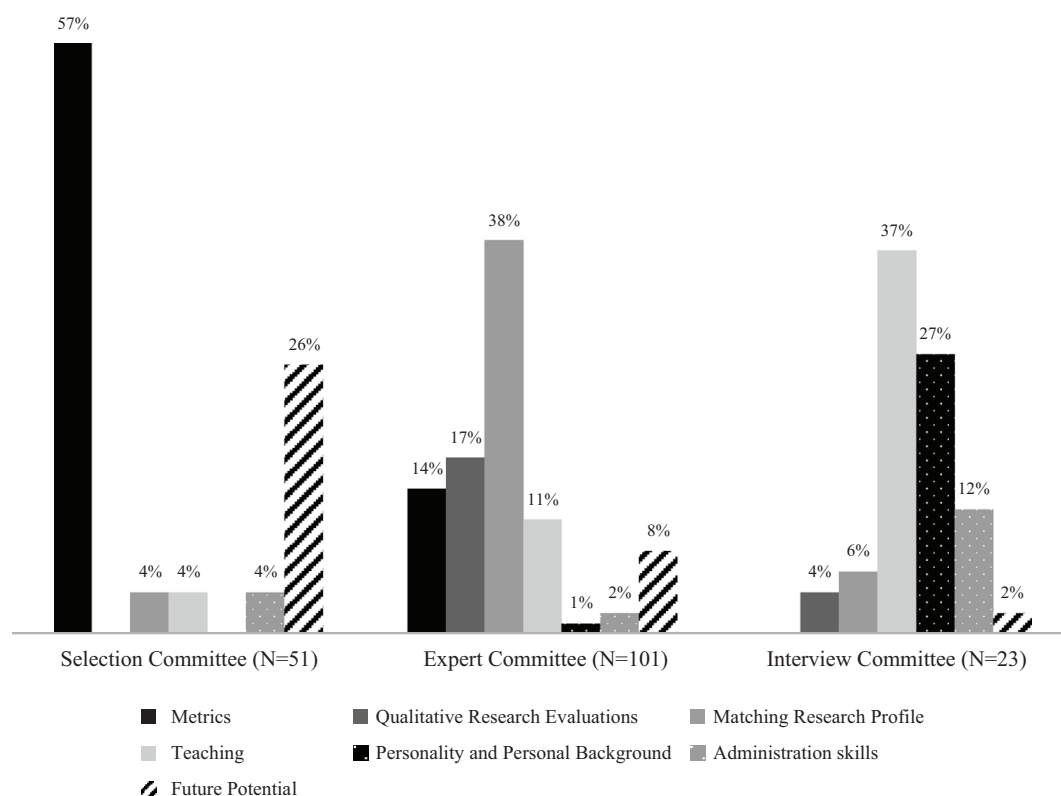


Fig. 4 Most important criterion in the three different committee types (percentage). N refers to the number of the most important assessment criterion/a in the different committees

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	<ul style="list-style-type: none"> Announcement texts emphasized the candidate's publications in international journals The expert committees used recognized and top international journals as the most important criteria 	<ul style="list-style-type: none"> Announcement texts also included <i>future potential</i> as an assessment criterion Selection committees were introduced The expert committees expanded their use of different types of metrics 	<ul style="list-style-type: none"> 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 <i>recognized</i> international journals instead of the volume of publications in international journals
Sociology	<ul style="list-style-type: none"> 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 	<ul style="list-style-type: none"> The expert committees expanded their use of different types of metrics 	<ul style="list-style-type: none"> The announcement texts mentioned metrics
Physics	<ul style="list-style-type: none"> 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 	<ul style="list-style-type: none"> The expert committees expanded their use of different types of metrics 	<ul style="list-style-type: none"> The announcement texts mentioned metrics
Informatics	<ul style="list-style-type: none"> 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 	<ul style="list-style-type: none"> The expert committees expanded their use of different types of metrics 	<ul style="list-style-type: none"> 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 (Välmaa 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; Välmaa 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](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 Hokyung 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 Academia – 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 heterodoxy 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 Hokyung 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/10.1080/13562517.2019.1576660>.
- 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 Thoening. 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.

Paper 3: The role of metrics in peer assessments

Liv Langfeldt, Ingvild Reymert and Dag W.Aksnes (Co-authored)

Publication status: Published in *Research Evaluation*,
2020, <https://doi.org/10.1093/reseval/rvaa032>

The role of metrics in peer assessments

Liv Langfeldt *, Ingvild Reymert and Dag W. Aksnes

Nordic Institute for Studies in Innovation, Research and Education (NIFU), P.O. Box 2815 Tøyen, N- 0608 Oslo, Norway

*Corresponding author. Email: liv.langfeldt@nifu.no.

Abstract

Metrics on scientific publications and their citations are easily accessible and are often referred to in assessments of research and researchers. This paper addresses whether metrics are considered a legitimate and integral part of such assessments. Based on an extensive questionnaire survey in three countries, the opinions of researchers are analysed. We provide comparisons across academic fields (cardiology, economics, and physics) and contexts for assessing research (identifying the best research in their field, assessing grant proposals and assessing candidates for positions). A minority of the researchers responding to the survey reported that metrics were reasons for considering something to be the best research. Still, a large majority in all the studied fields indicated that metrics were important or partly important in their review of grant proposals and assessments of candidates for academic positions. In these contexts, the citation impact of the publications and, particularly, the number of publications were emphasized. These findings hold across all fields analysed, still the economists relied more on productivity measures than the cardiologists and the physicists. Moreover, reviewers with high scores on bibliometric indicators seemed more frequently (than other reviewers) to adhere to metrics in their assessments. Hence, when planning and using peer review, one should be aware that reviewers—in particular reviewers who score high on metrics—find metrics to be a good proxy for the future success of projects and candidates, and rely on metrics in their evaluation procedures despite the concerns in scientific communities on the use and misuse of publication metrics.

Key words: peer review; research quality; bibliometric indicators; metrics; research fields

1. Introduction

Research organizations, funding agencies, national authorities and other organizations rely on peer assessments in their research evaluations. Peer assessments, in turn, may (partly) rely on metrics on scientific publications and their citations. In recent decades, such bibliometric indicators have become more easily accessible and have been used more in the evaluation of research. This raises the question of how such metrics impact what is perceived as good research, i.e. the notions of research quality. This paper addresses whether metrics are considered a legitimate and integral part of the assessment of research, explore the role of metrics in different review contexts and fields of research, and discuss implications for research evaluation and policy.

The use of metrics has a long history, dating back more than 100 years (De Bellis 2009). With the creation of the Science Citation Index by Eugene Garfield in 1961, new possibilities for quantitative studies of scientific publishing emerged, including analyses of how

often the articles had been referred to or cited in subsequent scientific literature. Initially, the potential of bibliometrics within science policy was only seen by a few individuals (Martin 1996). Later, research evaluation became an important area of application of bibliometric analyses. Today, indicators or metrics are applied for a variety of purposes and have permeated many aspects of the research system (Abbott et al. 2010; Aksnes, Langfeldt and Wouters 2019). For example, metrics have long been provided to peer reviewers in research evaluations, such as in national research assessments and institutional reviews (Lewison, Cottrell and Dixon 1999; Wilsdon et al. 2015). Nowadays, individual applicants may be requested to provide standardized Curriculum Vitae (CVs) that include citations rates when applying for grants¹, and metrics may also play an important role in hiring and promotion processes (Stephan, Veugelers and Wang 2017).

The use of bibliometric indicators has been more common in the natural and medical sciences than in the social sciences and humanities (Moed 2005). This may be due to the fact that the latter areas

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Table 2. Descriptive statistics

Variable/value	Count value	% value	n ^a
Age: 39 and younger	404	28	1435
Age: 40 to 49 years old	369	26	1435
Age: 50 to 59 years old	302	21	1435
Age: 60 years and older	360	25	1435
Gender: Female	325	23	1432
Gender: Male	1107	77	1432
Position: Assistant Professor	463	29	1611
Position: Associate Professor	391	24	1611
Position: Leader	77	5	1611
Position: Other	195	12	1611
Position: Professor	485	30	1611
Recruiting Juniors	552	71	774
Recruiting Seniors	222	29	774
Grant specification: Open Call	450	70	639
Grant specification: Target Research Call	189	30	639
Grant type: Fellowship	83	13	643
Grant type: Large Grants/Centre	78	12	643
Grant type: Research Project	482	75	643

Respondents' bibliometric performance	Mean	St. Dev.	Min	Max	n
Number of publications	27.86	63.835	0.250	781.00	1355
Log of number of publications	2.17	1.598	-1.39	6.66	1355
Have cited publications (dummy MNCS)	0.96	0.202	0.00	1.00	1355
MNCS	1.46	2.262	0.00	30.84	1355
Log of MNCS ^b	-0.03	0.913	-2.30	3.43	1297
MNJS	1.34	0.985	0.10	18.88	1355
Log of MNJS	0.15	0.517	-2.30	2.94	1355
Having publications in top percentile (dummy)	0.61	0.488	0.00	1.00	1355
Share of publications in top percentile	13.94	20.796	0.00	100.00	1355
Log of share of publications in top percentile ^b	2.73	0.933	0.00	4.61	828

^aSmaller n on reviews of grant proposals and candidates for positions, as these questions were posed only to those who reported to have participated in such reviews the last 12 months.

^bThe log of MNCS/Share of publication in top percentile (including those who have scores above 0 on the MNCS indicator/publications in top percentile).

corresponds to the world average. As an additional citation indicator, the proportion of articles that are among the 10% most cited articles in their fields has been included (the share of publications in the top percentile can be found in Table 2).

We included these metrics in binary logistic regression analyses, investigating the relation between the respondents' bibliometric performance and their emphases on metrics when characterizing the best research in their fields, assessing grant proposals and assessing candidates for positions. Equations are attached in the note.¹⁴ Apart from the factors included in the model, respondents' institutional affiliation and their specific research fields may influence their emphases in the assessment of research. Institutional affiliation has proven to influence researcher's evaluation at least in recruitments (Musselin 2010) and there may be large differences within research fields regarding notions of research quality and use of metrics (Lamont 2009; Hylmö 2018). Due to low numbers of respondents per institution, and insufficient data on subfields, we have not been able to control for these factors.

The bibliometric variables were skewedly distributed among the respondents, and thus the binary logistic regression analyses were conducted with log-transformed bibliometric variables, which ANOVA and AIC-tests showed improved our models. We settled on models with the log-transformed variables when displaying field

differences. Still for graphic illustration of results the original variables are used to ease the interpretation for the reader. Table 2 displays the distribution of the original and log-transformed metrics variables.

It should be noted that the Web of Science database does not equally cover each field's publication output. Generally, physics and cardiology are very well encompassed, while the coverage of economics is somewhat less so, due to different publication practices (Aksnes and Sivertsen 2019). In addition, not all respondents had been active researchers during the entire 2011–2017 period, and for 16% of the respondents in the sample no publications were identified in the database. The latter individuals were not included in the bibliometric analyses. Despite these limitations, the data provides interesting information on the bibliometric performance of the researchers at an overall level.

3.4 Methods

We used the programming software 'R' when analysing the data and 'RMarkdown' for visualization. The RMarkdown file can be provided upon request.

Weighted Results: As sample sizes vary by fields and country, the bivariate analyses were weighted so that each field in each country

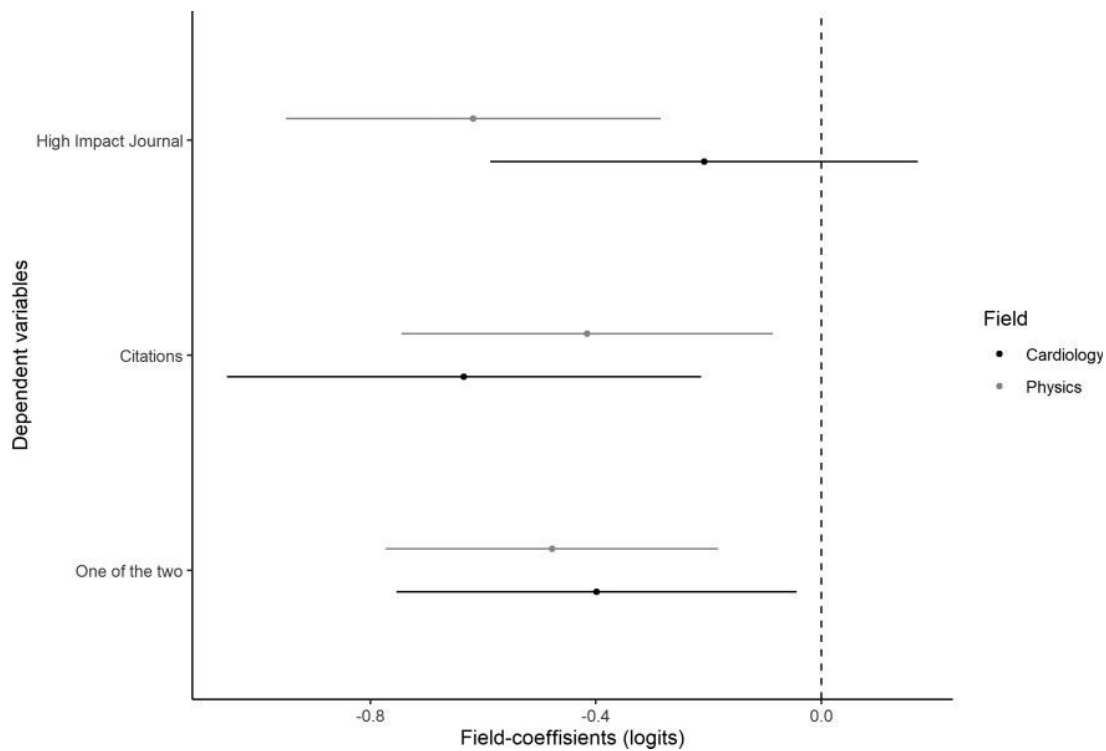


Figure 1. Journal impact factor and citations as reasons for considering something the best research in the field. Field coefficients from binary logistic regression analyses (Dot-And-Whiskers Plots). Economics as the baseline category represented by the dotted line.

Table 5. Aspects identified as “highly important” in grant assessments (Percent. Weighted results).

Reply	Cardiology	Economics	Physics	Total	Total n
Project description: research question/problem selection	98	87	94	94	678
Project description: methods/research plan	90	82	81	85	670
Track record of the research team: important prior contributions in the relevant research field (assessed independently of citation scores and source of publication)	46	35	55	46	673
Track record of the research team: number of publications/productivity	41	50	44	45	674
The research environment: resources and facilities for performing the proposed research	59	15	41	41	671
Track record of the research team: citation impact of past publications	18	29	25	23	676
Track record of the research team: experience with risk-taking research	19	14	21	19	670
Communication / dissemination plan for scientific publications	9	10	9	9	673
Other, please specify	10	8	9	9	664
Communication/dissemination plan addressing user groups outside academia	6	6	5	6	666

important compared to several other aspects (Table 5). They still appear far below the ‘research question’ (94%) and the ‘methods/research plan’ (85%), which came up as the most important in the assessments. Nonetheless, including those who replied ‘somewhat important’ (48% for number of publications and 59% for citation impact), the great majority replied that such metrics impacted their assessments of which proposal was the best (Supplementary Appendix Table A12).

The binary logistic regression analysis shows that emphases on metrics were field-dependent, as shown in Figure 2 (full regression models are shown in Supplementary Appendix Tables A4–A7). Compared to cardiologists, the economists (dotted line) were significantly more inclined to identify the number of publications and citation impact as ‘highly important’. Conversely, the physicists were significantly more inclined to emphasize important research contributions (assessed independently of metrics) than were the

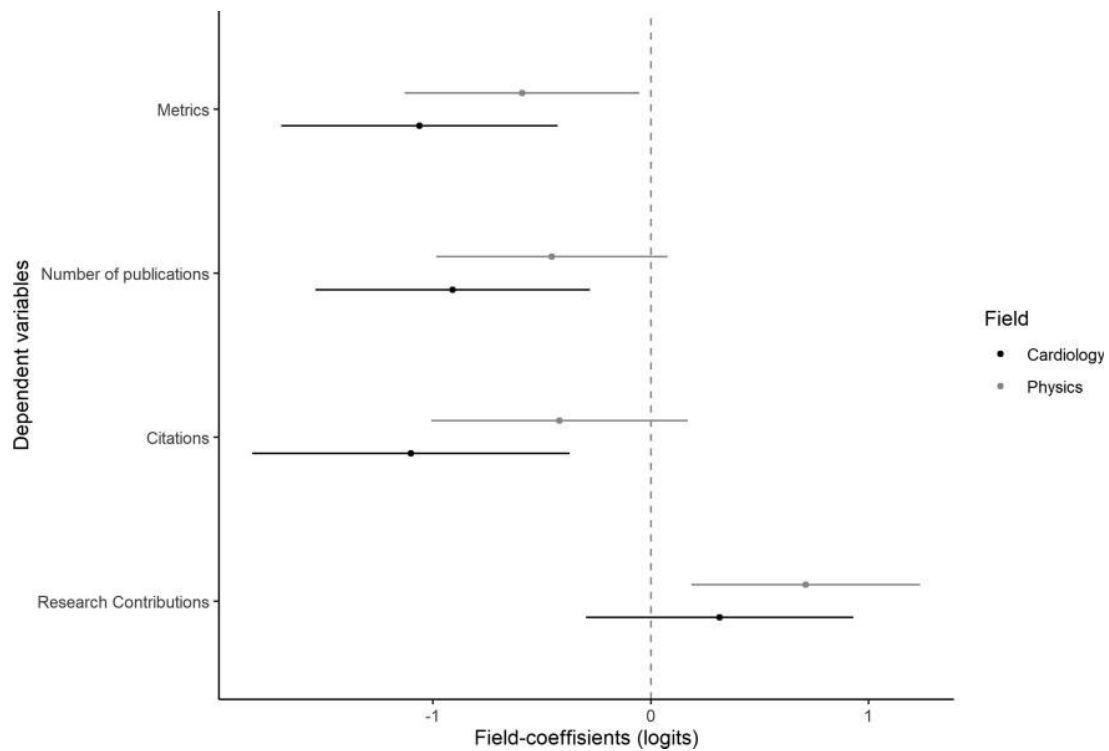


Figure 2. Aspects identified as ‘highly important’ in assessing grant proposals. Field coefficients from binary logistic regression analyses (Dot-And-Whiskers Plots). Economics appears as the baseline category, represented by the dotted line.

economists. The analysis uncovered less difference between physicists and cardiologists, but still, the physicists were significantly more inclined than the cardiologists to emphasize citations.

The regression coefficients imply that the probability of Swedish economist professors with an average share of top publications and number of publications, of identifying the number of publications or/and citations as ‘highly important’ in their assessment of proposals to open calls is 50%, whereas the probability of the similar groups of physicists and cardiologists to do so is substantially lower (39% for physicists and 29% for cardiologists). Conversely, the economists in this group (professors with average bibliometric scores) were less inclined to emphasize ‘important prior research contributions assessed independently of metrics’ (59% for physicists, 49% for cardiologists, and 42% for economists).

Furthermore, the regression analyses indicated insignificant country-related effects, but significant effects of the respondents’ academic positions and the type of grants being reviewed. The probability of identifying citations or number of publications as highly important was lower when assessing project grants than when assessing fellowships or large/centres grants. Moreover, the probability of identifying the number of publications as highly important was lower when reviewing proposals to open calls rather than targeted calls (Supplementary Appendix Tables A4–A7). The replies also depended on the respondents’ bibliometric performance, as discussed in detail below.

4.3 Candidates for positions

Similar results to those for assessing grant proposals appear for the assessment of candidates for positions: quantitative measures appear more important than when identifying the best research in the field.

Forty-two percent answered that the number of publications/productivity was ‘highly important’ in their assessments of candidates (Table 6). Citations impacts appear to be less important (19% replied that this was highly important). Notably, research contributions assessed independently of citation scores and publication source appear more important than number of publications/productivity in cardiology (47% highly important) and physics (61% highly important). In economics, on the other hand, there is a higher percentage who find the number of publications/productivity to be highly important (54%) and a lower percentage who find that contributions assessed independently of metrics are highly important (45%).

When the respondents were asked to identify the most important among the aspects they had identified as highly important, the candidate’s ‘potential for future achievements’ and ‘expertise matching the group/unit/project’ prevail as the two most important aspects in all three fields, indicating that these have general high importance regardless of fields. The third most important aspect, however, varied greatly between the fields: whereas cardiology appears with ‘general impression from interview with candidate’ and physics with ‘important prior research contributions (assessed independently of citation scores and source of publication)’, in economics ‘number of publications/productivity’ appears as the third most important aspect (Supplementary Appendix Figure A1).

Binary logistic regression analysis confirms statistically significant differences between fields (documented in Supplementary Appendix Tables A8–A11 and illustrated in Figure 3 below). Economists were more inclined than both cardiologists and physicists to identify the number of publications as ‘highly important’ when assessing candidates for positions (the difference between the

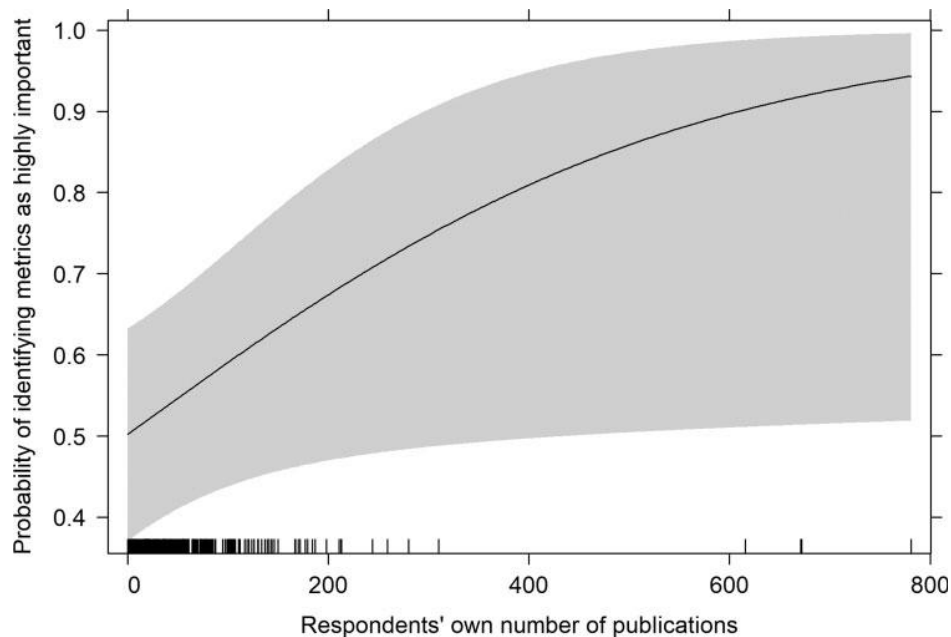


Figure 4. Assessment of grant proposals: The probability of identifying the number of publications and/or citation impact as 'highly important', by respondents' number of publications. The rug at the x-axis marks the number of observations.

The ANOVA tests showed that, in some of the models, the respondents' gender, age and country had significant effect on the respondents' emphases. The country-related effects were mostly insignificant, but being Dutch instead of Swedish decreased the possibility of identifying publication numbers as 'highly important'. Professors were less inclined than those in other kinds of positions to see the number of publications as 'highly important', while there was no significant effect on emphasis on citations. Furthermore, both quantitative measures and important prior contributions were more often seen as important when recruiting to senior rather than to junior positions.

4.4 The effects of the respondents' bibliometric performance

Looking further into the results, a key question is whether the respondents' emphases on publication metrics corresponded with their own bibliometric performance. For example, one might find that researchers with many publications would put more emphasis on this dimension in their assessments. Therefore, we compared the respondents' answers with their own scores on the relevant bibliometric indicators.

The regression analyses showed no effect of respondents' bibliometric performance on their reasons for considering something the best research. However, in practice, when assessing grants proposals and candidates for positions, their own performance was positively related to their use of metrics. When assessing grant proposals, the probability of identifying the number of publications and citation impact as 'highly important' increased along with the respondents' number of publications, whether they had top percentile publications, and their share of top percentile publications.¹⁹ Figure 4 displays this relationship, and shows how the probability of identifying citation impact and/or number of publications as highly important in assessments of grant proposals for Swedish professors in

economics increases with their own number of publications.²⁰ Respondents with high bibliometric performance scores more frequently considered such indicators as important in their assessments. On the other hand, the respondents' bibliometric performance did not affect whether they found prior contributions 'highly important' (Supplementary Appendix Table A7).

Similarly, metrics in the assessment of candidates for academic positions depended on the respondents' bibliometric performance, but less so than for grant assessments. The respondent's own number of publications did not significantly affect the probability of identifying candidates' citation impact or number of publications as 'highly important', yet the log of the number of publications increased with respondents' publication output. Moreover, a respondent having top cited publications increased the probability of identifying candidates' citations as 'highly important'. For Swedish economics professors who recruited for senior positions, having top cited publications increased this probability from 28 to 40%. However, neither the respondents' MNCS, MNJS nor share of top cited publications had a significant effect on the use of metrics in the assessment of the candidates. Hence, the importance of metrics in these assessments was lower than that for the assessment of grant proposals.

Moreover, the respondents' MNCS (log of) and MNJS increased the probability of identifying prior research contributions as highly important when assessing candidates for positions, but did not increase the probability of identifying such contributions as highly important when assessing grant proposals.

4.5 Divergent opinions and perspectives

Insofar as this attempt to conclude on whether metrics are considered a legitimate and integral part of assessments of research, the results indicate conflicting views as well as differences between review contexts and the type of metrics. A large majority of

Table 7. Free text replies—what was important for your assessment of:

The best grant proposal	
1	'I evaluate research according to the value of proposed research. For the highest scores, there has to be an outstanding problem to address. There has to be a realistic plan outlining how this is possible. [...] I don't pay much attention to the rate of publications, but rather to the lasting impact of these. I also do not care much about citations, as this varies profoundly between different topics. Rather, impact must be assessed based on an actual understanding of previous research' (Research project grant, physicist in Sweden who selected 'not important' both on number of publications and citations.)
2	'Solves relevant questions. Science at excellent level considering modern perspectives in research evaluation (NOT publication and citations numbers as primary component).'
3	(Fellowship grant, cardiologist in Norway who selected 'not important' both on number of publications and citations) 'Concrete yet ambitious proposal, novel methodology, included a Plan B if the risky plan A failed, collaboration with industry to obtain interesting field data, productive researcher with high H-index.'
4	(Research project grant, social sciences/other in the Netherlands) 'Important research problem, High-quality candidate(s), Track record in terms of publications in top international journals (e.g. Nature, Science), assessed independently of citation records.' (Large grant/centre, physicist in the Netherlands)
The best candidate for a position	
5	'The candidate was expected to do research, to teach, and in particular to build a Research Group within [subfield]. Communication skills, Cultural competence and networking ability are crucial, in addition to number and "weight" of publications.'
6	(Junior/early career position, cardiologist in Norway) 'Much of the selection is based on the impact factors of the journals a candidate has published in, and potentially the network of the candidate (I do not necessarily believe these are the best criteria per se, but they are generally used).'
7	(Junior/early career position, economist in the Netherlands) 'It's a mix. Citation impact without productivity indicates a few very highly cited papers, which is not what I mean. It should be a combination of high productivity of high-quality papers that also have attracted citations. So number of high-quality papers, overall citations, h-index, and where the work was published all matter.' (Senior/tenure position, physicist in Sweden)

respondents confer to metrics in their reviews and seem to find it a legitimate and ordinary basis for reviews. This goes particularly for the number of publications/productivity in the review of grant proposals; only six percent replied that this was not important (Supplementary Appendix Table A12). Still, a substantial proportion (33%) indicated that citation scores were not important regarding assessments of candidates for academic positions (Supplementary Appendix Table A13).

The free text replies concerning the main positive characteristics of the best proposals illustrate the divergent opinions and perspectives. Some grant reviewers emphasized that metrics were not important (illustrated by #1 and #2 in Table 7). Others emphasized publication and citation rates as key characteristics of the best proposal, or simply publications in top international journals (#3 and #4 in Table 7).

Several of those who had reviewed candidates for positions seemed to find publications in major/top journals a basic or objective criterion, and then added other important characteristics that would befit the particular research group or the tasks of the position (#5 in Table 7). Others indicated that the ranks of the journals the candidate had published in—or a combination of relevant metrics—was important in the selection process. Still, views on the adequacy of such criteria varied (#6 and #7 in Table 7).

5. Discussion and implications

In this paper we have explored whether metrics are part of researchers' notion of good research, and whether metrics are used when

reviewing research. Concerning the first issue, only a minority of the respondents reported metrics as a reason for considering something to be the best research. Thus, the empirical support for such an association is generally weak. On the second question, we find strong supportive evidence as a large majority indicated that metrics were important or partly important in their review of grant proposals and assessments of candidates for academic positions.

Notably, drawing conclusions on researchers' notions of research quality is difficult. Research quality is a multidimensional concept; what are seen as the key characteristics of good research may differ largely between contexts and fields (Langfeldt et al. 2020). Metrics such as citations, publication counts or journal impact factors may be perceived as relating to different characteristics of research quality, e.g. according to bibliometric studies, citations reflect (to some extent) the scientific value and impact of research, but not its originality, plausibility/soundness or societal value (Aksnes, Langfeldt and Wouters 2019). Our data indicate that the respondents distinguish between quantitative indicators as proxies for success when assessing the potential of future projects or candidates for positions and what they hold to be the characteristics of good research. A large majority of the respondents reported metrics as highly or somewhat important in their reviews of grant proposals and of candidates for positions, whereas about one-fifth of them indicated that one of their reasons for concluding on what was the best research in their field was that it was published in a journal with high impact factor or that it had attracted many citations. Hence, for one-fifth of the researchers in the survey, metrics seem to be a judgement device when identifying good research within one's

own field. This does not necessarily imply that they hold metrics, as such, to be characteristics of good research. Very few respondents indicated the journal impact factor or high citation rates as sole indicators of the best research in their field, and there is little indication that respondents view quantitative indicators as being a sufficient basis for concluding on eminent science. Nevertheless, some have suggested that publishing in high-impact journals has become an independent measure of scientific quality (Wouters 1999; Rushforth and de Rijcke 2015).

Moreover, the analysis indicates significant field differences in the use of publication metrics: The economists were more inclined to indicate journal impact factor and many citations as reasons for concluding that something is the best research in their field, and they were more inclined to emphasize the applicants' number of publications when assessing grant proposals and candidates for positions. Physicists and cardiologists, on the other hand, were less inclined to emphasize metrics and more inclined to emphasize prior research contributions assessed independently of metrics. These differences go along with differences in how research is organized and valued in these fields. In economics, research is mostly performed by individuals and organized around a theoretical core and key journals of high importance for individual reputation (Whitley 1984; Hammarfelt and Rushforth 2017; Hylmö 2018). Herein, high reliance on metrics may be explained by the combination of an explicit journal hierarchy and organization of research that makes it easier to attribute research performance to individuals. Physics consists of highly collaborative fields, some with hyper-authorship (Birnholtz 2008), and using publication metrics to attribute research performance to individual researchers is more difficult. Similarly, cardiology is a field within medical research with specialized tasks and skills, highly dependent on collaboration, resources and facilities for performing research (Whitley 1984), which may explain the lower emphasis on publication metrics than in economics, as well as far stronger emphases on research resources and facilities when assessing grant proposals. Notably, there is also much variation in replies within the fields: for example, a substantial proportion of the physicists and the cardiologists indicate the applicant's number of publications as highly important when assessing grant proposals and candidates for academic positions, whereas others find it unimportant or somewhat important. In sum, this points to the importance of understanding how epistemic and organizational differences—both between and within research fields—generate different bases for assessing research and research performance, and thereby different use of metrics.

Despite our comparative point of departure, along with the inclusion of countries with different use of metrics in national research funding, we found only limited country-specific differences. The lack of country-related differences indicates that notions of research quality are more connected to general field differences than to national context (Lamont 2009; Musselin 2010). Still, even if our sample of three countries in the northern corner of Europe represents variety in research funding and research evaluation, a larger sample of more diverse countries might have exposed greater differences in the use of metrics in peer assessments.

The findings have policy importance for multiple aspects of the evaluation of research. Below, we discuss implications relating to (1) how research agendas and research activity adapt to research evaluations, (2) the policies for restraining the (mis)use of metrics in research evaluation, and (3) the design and organization of research evaluations.

First, an emphasis on metrics may impact research activity and research agendas. Researchers—at least young and non-tenured ones—cannot disregard what gives acclaim in the academic career system and what is needed for attracting research funding. They need to take into consideration what kind of research will help them qualify for grants and positions (Müller and de Rijcke 2017). Notably, in our data, economists seem to put less emphasis (compared with the other groups) on expertise, matching the needs of the research group/unit, and they seem to be more apt to use metrics (Supplementary Appendix Figure A1). This may imply that, rather than making explicit decisions about the kind of researchers to employ (their topics and methods), the researchers who are able to do the kind of research that are most easily published in (top) economics journals are hired (Lee, Pham and Gu 2013). Hence, the ways in which researchers adapt to metrics come up as a key topic for studies in research evaluation and, more generally, for research policy.

Second, despite increasing concerns in the scientific communities on the use and misuse of research metrics (Wilsdon et al. 2015), the results herein indicate that researchers rely on the three types of metrics addressed in the survey: journal impact factors, number of publications and citation impact. Close to one-fifth of the respondents reported high impact factor as a reason for something being the best research in their field. As discussed in the introduction, journal impact factors and journal rankings have been widely used, particularly in medicine and economics, for assessing scientific performance. With the launch of the DORA-declaration in 2012, the problem with this practice has received more attention.²¹ As a response, policies and practices of many funding organizations, scientific societies, institutions and journal publishers have changed, according to Schmid (2017). Nevertheless, others report that journal impact factors are still used for purposes that conflict with the DORA-declaration (Bonnell 2016). Notably, the DORA-declaration has led to an increased focus on other ways to assess research. This includes the development of alternative paper-based metrics (Schmid 2017). Indicators of number of publications and citation impact do not have similar problems to those associated with the journal impact factor. Nevertheless, it is well known that these indicators also have various limitations and shortcomings as performance measures, particularly when applied at micro levels (Wildgaard, Schneider and Larsen 2014), and our survey indicates extensive use of these indicators at micro levels when reviewing grant proposals or candidates for academic positions. Moreover, the field differences found in the survey point to a need for a better understanding of why and how metrics are used in different fields as well as a need to consider field-adjusted policies for the use of metrics in research evaluation.

Finally, there are implications regarding the design and organization of research evaluation. Publication-based metrics seem to be perceived as good proxies for research quality and performance, at least for the majority of the researchers in the fields studied. This may be because they trust the review processes of the scholarly journals and publishers in their field, and metrics make sense as a proxy for quality. From this perspective, the editors and reviewers of the major journals end up high on the list of those controlling the gate-keeping criteria, not only for scientific publishing, but also for academic positions and research grants. At the end of the 'review chain', we will often find the criteria, review processes and publication policy of the major journals in the field. Hence, the researchers complying with the topics, perspectives/methods and formats of these journals can be expected to have the highest chances of success in competitions for grants and academic positions. Still, the above

analysis indicates deviant views among reviewers on the use of metrics in research evaluation. So even if certain topics, perspectives or methods dominate a field, the outcome of review processes may vary by the panel members' views on metrics. Consequently, when it comes to the 'luck of the reviewer draw' (Cole, Cole and Simon 1981), not only the panel members' scholarly profile and competences, but also their preferences for metrics may be decisive. This implies that in order to provide fair and well-grounded review processes, there is a need for insight into how panels use metrics in their assessments and to encourage explicit discussions about the use of metrics²². If the role of metrics is not openly discussed in review panels, nor understood by those organizing the reviews and acting upon them, we risk concealed review criteria.

Notes

1. When applying for advanced grants from the European Research Council (ERC), applicants have been asked to provide a ten-year track record including publications in leading journals and 'indicating the number of citations (excluding self-citations) they have attracted (if applicable)'. https://erc.europa.eu/sites/default/files/document/file/ERC_Work_Programme_2015.pdf. We find this formulation in the ERC work programmes for 2008–2016. For 2017, 2018, and 2019 the wording is: '(properly referenced, field relevant bibliometric indicators may also be included)'. http://ec.europa.eu/research/participants/data/ref/h2020/wp/2018-2020/erc/h2020-wp19-erc_en.pdf.
2. <http://www.ascb.org/dora/>
3. Likewise, in Norway, the research council regularly conducts peer evaluations of disciplines and subjects as well as institutes and programmes, and bibliometric indicators are used as one source of information whenever relevant (Sivertsen 2017).
4. Such perceptions may in turn be formed by/related to extensive use of e.g. journal rankings or citation measures in the field (Espeland and Sauder 2007: 16).
5. The use of metrics in peer review is also part of the more general story about how information technology impacts our evaluative practices (Lamont 2012: last section).
6. For example, the Research Council of Norway requires applicants to use a CV template that includes citation counts for applicants for regular researcher projects in all research fields. Up to 2018, the RCN template was named after its role model 'ERC track record description'.
7. A majority of those who provided input to the 'Metric Tide' report were sceptical to the role of metrics in research management while a significant minority were more supportive of the use of metrics (Wilsdon et al. 2015: viii).
8. The minimum number of publications (5 for economics and 10 for cardiology and physics) was selected based on analyses of individual publication output during the 2011–2016 period. By applying these thresholds we aimed at including the more active researchers within the fields and leaving peripheral researchers out. A higher number was applied for cardiology and physics because of the higher publication frequencies (and co-authorship) in these fields.
9. The survey is part of a larger research project and was launched in five countries in 2017–2018. The present analysis is based on replies from higher education institutions in three of these countries (1,621 replies). The full survey included

2,587 replies, and is also comprised of replies from economics and physics in Denmark and the UK, as well as replies from researchers affiliated with independent research institutes. Replies from Denmark and the UK are excluded from the present analyses, as cardiology was not sampled in these countries. We checked for the impacts of excluding the UK and Danish samples by conducting the analyses on Economics and Physics in all five countries, and did not find any significantly deviant results. Moreover, replies from independent research institutes are excluded as they constitute a small sample (in total 111 replies in the three countries) and research settings which may differ substantially from those at higher education institutions.

10. Table 1 shows response rates by field as identified in the sampling process, whereas our analyses are based on field as identified by survey responses. Respondents who replied other fields of research, rather than 'Cardiac/cardiovascular systems/diseases', 'Economic' or 'Physics' are not included in the analysis. Hence, the analyses are based on a smaller sample (1,621 respondents) than that which prevails in Table 1 (1,942 respondents).
11. Consequently, the analyses are based on the full sample for the first question, and different subsamples for the two latter questions. We checked for impacts of sample variation by additional analyses of those included in both subsamples (451 respondents stated that they had reviewed both grant proposals and candidates for positions the last 12 months). These analyses did not give deviant results. Hence, differences between the two review settings appearing from our data are not due to different subsamples.
12. We have data on gender for 92% of the invited respondents. Of these, 39% of the female and 35% of male respondents replied. Of those without information on gender, we have replies from 4%.
13. We have excluded minor contributions such as editorials, meeting abstracts, and corrections. As letters usually do not represent full scientific contributions, they are weighted as 0.25 of an article; this is in accordance with principles often applied by the Centre for Science and Technology Studies (CWTS) of Leiden University (for further discussion, see van Leeuwen, van der Wurff and de Craen 2007).
14. $Y1 = X1Country + X2Field + X3Bibliometricsb + X4Age + X5Gender + X6Position + e$
 $Y2 = X1Country + X2Field + X3Bibliometricsb + X4Age + X5Gender + X6Position + X7Call + X8Type + e$
 $Y3 = X1Country + X2Field + X3Bibliometricsb + X4Age + X5Gender + X6Position + X7Vacancy + e$
 Dependent variables: why they considered something to be the best research in their field (Y1), what was important for their assessment of grant proposals (Y2) and candidates for positions (Y3).
 a = type of assessment criteria
 b = type of bibliometrics
 e = error term
15. As an extra control, the regression analyses were run with weights. Results were not altered.
16. We also conducted ordinal logistic regressions for the best suited models with assessment of grant proposals and candidates for academic positions as dependent variables. These

- Lamont, M. (2009) *How Professor Think: Inside the Curious World of Academic Judgment*, Cambridge, MA: Harvard University Press.
- Lamont, M. (2012) 'Toward a Comparative Sociology of Valuation and Evaluation', *Annual Review of Sociology*, 38: 201–21.
- Langfeldt, L., Nedeve, M., Sörlin, S., and Thomas, D. A. (2020) 'Co-Exiting Notions of Research Quality: A Framework to Study Context-Specific Understandings of Good Research', *Minerva*, 58: 115–37.
- Langfeldt, L., and Scordato, L. (2016) *Efficiency and Flexibility in Research Funding. A Comparative Study of Funding Instruments and Review Criteria. NIFU Report 9/2016*. Oslo: NIFU Nordic Institute for Studies Innovation, Research and Education.
- Lee, F. S., Pham, X., and Gu, G. (2013) 'The UK Research Assessment Exercise and the Narrowing of UK Economics', *Cambridge Journal of Economics*, 37: 693–717.
- Lewis, G., Cottrell, R., and Dixon, D. (1999) 'Bibliometric Indicators to Assist the Peer Review Process in Grant Decisions', *Research Evaluation*, 8: 47–52.
- Lin, F. (2008) 'Solving Multicollinearity in the Process of Fitting Regression Model Using the Nested Estimate Procedure', *Quality and Quantity*, 42: 417–26.
- Loomba, R. S., and Anderson, R. H. (2018) 'Are we Allowing Impact Factor to Have Too Much Impact: The Need to Reassess the Process of Academic Advancement in Pediatric Cardiology?', *Congenital Heart Disease*, 13: 163–6.
- Martin, B. R. (1996) 'The Use of Multiple Indicators in the Assessment of Basic Research', *Scientometrics*, 36: 343–62.
- Moed, H. F. (2005) *Citation Analysis in Research Evaluation*. Dordrecht: Springer.
- Müller, R., and de Rijcke, S. (2017) 'Thinking with Indicators. Exploring the Epistemic Impacts of Academic Performance Indicators in the Life Sciences', *Research Evaluation*, 26: 157–68.
- Musselin, C. (2010) *The Market for Academics*, New York: Routledge.
- Piro, F. N., Aksnes, D. W., and Rorstad, K. (2013) 'A Macro Analysis of Productivity Differences Across Fields: Challenges in the Measurement of Scientific Publishing', *Journal of the American Society for Information Science and Technology*, 64: 307–20.
- Rushforth, A., and de Rijcke, S. (2015) 'Accounting for Impact? The Journal Impact Factor and the Making of Biomedical Research in the Netherlands', *Minerva*, 53: 117–39.
- Schmid, S. L. (2017) 'Five Years post-DORA: Promoting Best Practices for Research Assessment', *Molecular Biology of the Cell*, 28: 2941–4.
- Sivertsen, G. (2017) 'Unique, but Still Best Practice? the Research Excellence Framework (REF) from an International Perspective', *Palgrave Communications*, 3: 17078.
- Söderlind, J., and Geschwind, L. (2020) 'Disciplinary Differences in Academics' Perceptions of Performance Measurement at Nordic Universities', *Higher Education Governance & Policy*, 1: 18–31.
- Sousa, C. A. A., and Hendriks, P. H. J. (2007) 'That Obscure Object of Desire: The Management of Academic Knowledge', *Minerva*, 45: 259–74.
- Stephan, P., Veugelers, R., and Wang, J. (2017) 'Blinkered by Bibliometrics', *Nature*, 544: 411–2.
- van der Wall, E. E. (2012) 'Journal Impact Factor: Holy Grail?', *Netherlands Heart Journal*, 20: 385–6.
- van Leeuwen, T. N., van der Wurff, L. J., and de Craen, A. J. M. (2007) 'Classification of 'Research Letters' in General Medical Journals and Its Consequences in Bibliometric Research Evaluation Processes', *Research Evaluation*, 16: 59–63.
- Weingart, P. (2005) 'Impact of Bibliometrics upon the Science System: Inadvertent Consequences?', *Scientometrics*, 62: 117–31.
- Whitley, R. 1984. *The Intellectual and Social Organization of the Sciences*. Oxford: Clarendon Press.
- Wildgaard, L., Schneider, J. W., and Larsen, B. (2014) 'A Review of the Characteristics of 108 Author-Level Bibliometric Indicators', *Scientometrics*, 101: 125–58.
- Wilsdon, J., Allen, L., Belifiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Johnson, B. (2015) *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. HEFCE. DOI: 10.13140/RG.2.1.4929.1363, <https://responsiblemetrics.org/the-metric-tide/>.
- Wouters, P. (1999) 'Beyond the Holy Grail: From Citation Theory to Indicator Theories', *Scientometrics*, 44: 561–80.

Paper 4: Are evaluative cultures national or global? A cross-national study on evaluative cultures in academic recruitment processes in Europe

Ingvild Reymert, Jens Jungblut, Siri B. Borlaug (Co-authored)

Publication status: Published in *Higher Education*, 2020, <https://doi.org/10.1007/s10734-020-00659-3>



Are evaluative cultures national or global? A cross-national study on evaluative cultures in academic recruitment processes in Europe

Ingvild Reymert¹ · Jens Jungblut² · Siri B. Borlaug¹

Accepted: 17 November 2020
© The Author(s) 2020, corrected publication 2021

Abstract

Studies on academic recruitment processes have demonstrated that universities evaluate candidates for research positions using multiple criteria. However, most studies on preferences regarding evaluative criteria in recruitment processes focus on a single country, while cross-country studies are rare. Additionally, though studies have documented how fields evaluate candidates differently, those differences have not been deeply explored, thus creating a need for further inquiry. This paper aims to address this gap and investigates whether academics in two fields across five European countries prefer the same criteria to evaluate candidates for academic positions. The analysis is based on recent survey data drawn from academics in economics and physics in Denmark, the Netherlands, Norway, Sweden, and the UK. Our results show that the academic fields have different evaluative cultures and that researchers from different fields prefer specific criteria when assessing candidates. We also found that these field-specific preferences were to some extent mediated through national frameworks such as funding systems.

Keywords Evaluative cultures · Recruitment · Cross-national study

Introduction

Academia has always been an international endeavor as disciplines transcend national borders, and scholars collaborate internationally. However, this trend has increased in recent years, and academia has become even more globalized, with flows of international students, researchers, and an international academic job market in which universities compete for the best researchers. Not only are universities actively recruiting foreign faculty to build their international reputation, but also individual researchers are actively using their international network to recruit highly qualified postdocs and PhDs to fill their research and teaching needs outside their country (Ortiga et al. 2020). In Europe, the Bologna Process

✉ Ingvild Reymert
ingvild.reymert@nifu.no

¹ Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway

² Department of Political Science, University of Oslo, Oslo, Norway

was initiated to integrate national research and higher education systems into one European entity, and most national research systems encourage and support international research mobility (Chou and Gornitzka 2014). Additionally, universities around the world are moving toward organizational similarities, and national university characteristics are decreasing (Krücken and Meier 2006; Ramirez 2006). These global trends have also affected national career systems as countries have introduced tenure tracks (Henningsson et al. 2017), and universities worldwide have implemented more professional approaches to staff management partly as a response to these global developments (Van den Brink et al. 2013).

However, higher education institutions are still embedded in national systems, as they are primarily nationally funded and serve national needs; additionally, most researchers and students still work and study in their home countries. Furthermore, national research systems differ in both their research priorities and evaluative systems, and universities have different degrees of control over resources (Clark 1978; Sivertsen 2017; Whitley 2003). According to Musselin (2005), the most striking national patterns in academia are the labor market structures in which salaries, recruitment procedures, career patterns, and promotion rules differ between countries. Thus, despite the increasingly globalized nature of academic careers, they are still shaped in a national context (Finkelstein 2015).

Recruitment processes are critical junctures in the academic job market that determine the future of university performance through staff composition and the futures of individual researchers. They are complex processes shaped by both the cultural and regulatory contexts of countries, institutions, and fields (Musselin 2010; Sanz-Menéndez and Cruz-Castro 2019). The evaluation and ranking of candidates are central parts of these processes and include peer review of qualifications and strategic concerns (Langfeldt and Kyvik 2011; Musselin 2010). Peer review involves negotiations of different quality notions, and candidates are evaluated on multiple criteria, such as their scientific publications, citation numbers, teaching experience, administration and leadership skills, experience attracting grants, or social skills. Nevertheless, studies have shown that candidates' research output seems to be the most important criterion (Hamann 2019; Levander et al. 2019; Van den Brink and Benschop 2011). These evaluative processes are further embedded in each academic field's evaluative culture, assessing candidates by criteria that reflect the epistemology of the field and its notions of research quality (Herschberg et al. 2018; Levander et al. 2019), which could be both nationally and internationally anchored (Lamont 2009; Langfeldt et al. 2019). Even though national recruitment regulations are very different, Musselin (2010) argued that candidate evaluations are less sensitive to national regulations. However, apart from her study on recruitment processes in France, the USA, and Germany, most studies on peer reviews in recruitment processes focus on a single country and do not account for national differences (e.g., Hamann 2019; Hylmö 2018; Levander et al. 2019), nor are field peculiarities described. Some studies only cover one field (Hamann 2019; Hylmö 2018), while others only distinguish between social and natural science and neglect to disentangle specific disciplinary differences (e.g., Herschberg et al. 2018; Levander et al. 2019); still others have primarily utilized disciplines as control variables (e.g., Van den Brink and Benschop 2011).

Given the contemporary higher education landscape and studies indicating somewhat contrary tendencies in national convergence and divergence, there is a need for a comparative view of the preferences regarding evaluative criteria in recruitment. This paper investigates which evaluative criteria individual academics prefer in recruitment processes, and we ask to what extent scholars from different fields and countries prefer the same evaluative criteria when assessing candidates in academic recruitment processes. To investigate this, we drew on two somewhat contradicting theoretical perspectives: (i) institutional

isomorphism, which suggests that the internationalization of universities and the evaluation processes preferred in recruitment are embedded in academic fields and lead to national similarities in evaluative criteria preferences, and (ii) path dependency, which suggests that higher education institutions are embedded in a national context that generates a country-specific culture regarding researchers' preferences for different evaluative criteria.

To examine these elements, we applied new and original survey data from 2017/2018, including responses from economics and physics academics in universities in five different European countries (Norway, Sweden, Denmark, the Netherlands, and the UK). In the following sections, we first review prior studies on academic recruitment and then formulate expectations based on arguments about the role of national contexts and isomorphism. We then present data and methods before turning to the analysis. We summarize our results in the “[Discussion](#)” and “[Conclusion](#)” sections and offer avenues for further research.

Evaluative criteria in recruitment

Recruitments are fundamental organizational processes as well as gatekeepers that control and determine organizational membership and social boundaries. These processes are salient in organizations that must recruit the right kind of people for both technical and symbolic reasons (Scott and Davis 2007). In these processes, the application of evaluative criteria is pivotal, since they represent core peer-review processes (Langfeldt and Kyvik 2011) that are controlled and conducted by the academic profession (Musselin 2010). In these evaluations, candidates are assessed on multiple criteria, such as teaching experience (Levander et al. 2019), international experience and language skills (Herschberg et al. 2018), administrative skills (Hamann 2019), or social skills (Musselin 2010); therefore, the desired candidates are often referred to as “the sheep with five legs” or “jack-of-all-trades” (Van den Brink and Benschop 2011). Nevertheless, research output is often the most salient criterion (Van den Brink and Benschop 2011), although teaching experience has recently been gaining importance (Levander et al. 2019).

However, research quality and academic qualifications are not fixed entities but socially constructed and negotiated among academics in peer-review processes (Langfeldt et al. 2019). Each field has its own evaluative culture with its own understanding of academic qualifications and research quality that is tightly linked to its identity, epistemology, and academic work (Becher and Trowler 1989; Lamont 2009). In these different evaluative cultures, qualities are valued differently. For instance, humanities scholars have been found to define interpretative skills as highly important, while this quality has a more negative connotation in the social sciences (Lamont 2009).

Different evaluative cultures have also been identified in the evaluation of candidates in recruitment processes. Researchers in economics, for example, place more emphasis than researchers in biomedicine on how many publications a candidate has secured in top journals (Hammarfelt and Rushforth 2017). Furthermore, international experience is more highly valued in the natural sciences than the social sciences (Herschberg et al. 2018), and teaching experience generally seems to be more strongly emphasized in Science Technology Engineering and Mathematics (STEM fields) than in humanities or social science (Levander et al. 2019). Yet, despite prior studies addressing these field-based characteristics of evaluative criteria, the empirical evidence is still scarce.

The use of evaluative criteria is not only regulated by academic fields but also national academic career structures with different types of positions and diverse obligations, career

paths, and recruitment procedures (Alfonso 2016; Sanz-Menéndez and Cruz-Castro 2019). In Musselin's (2010) study, she found that the national context, including different formal and informal procedures, to some extent shapes the use of evaluative criteria. The American custom of inviting candidates to a visit, including lunch and dinner, provides greater opportunity for evaluating their personalities, and contrasts with the European approach of recruiters and candidates having more limited social encounters through more formal interviews. Similarly, the German custom of hiring candidates with long careers increases recruiters' expectations of scholarly output compared with the French or American traditions of hiring younger applicants (Musselin 2010). At the same time, the national differences in evaluative cultures are somewhat less sensitive to external logics. This paper does not aim to provide a complete description of national systems and regulations, but their diversity is an important premise that shapes national contexts in this study and, thus, potentially influences the way evaluative criteria are applied. We will return to this in the "Methods" section, in which we elaborate upon the selection of countries and fields for our study.

Hitherto, we have discussed how different evaluative criteria play decisive roles in academic recruitment. However, recruitment processes are complex undertakings in which peer evaluation and evaluative criteria are only two of many factors influencing the final selection. Studies have shown that recruitment is not always meritocratic (Nielsen 2016) where candidates are selected based on a set of fixed evaluative criteria (Musselin 2010). Academic inbreeding is common in many countries (Altbach et al. 2015; Tavares et al. 2019), and social networks enhance both academic careers (Pezzoni et al. 2012; Rossier 2020) and candidates' chances of success in recruitment (Combes et al. 2008; Lutter and Schröder 2016). The literature has also shown that recruitment processes can be gender biased (Husu 2000; Nielsen 2016; Wennerås and Wold 1997). Finally, recruitment processes are social processes in which the ranking of candidates may be strategic (Musselin 2010), and evaluators need to legitimize their conclusions (Hamann 2019). In particular, the final selection of the highest-ranked candidates is often hard to explain in terms of evaluative criteria (Musselin 2010). Although these criteria may not explain why a candidate was ranked highest, they reflect how evaluation committees select a shortlist of candidates, as they offer a basis for peers' discussions of candidates and the criteria used in arguing for them, even in cases in which this is window dressing (Musselin 2010). Moreover, the emphasis on evaluative criteria sends strong signals to the research community regarding which qualifications are more important than others (Tagliaventi et al. 2020), and it has thus been shown to be important in the selection of candidates (Herschberg et al. 2018; Van den Brink and Benschop 2011). Hence, we focus on evaluative criteria as more openly expressed factors affecting candidate selection. At the same time, we acknowledge that there are other more subtle or informal factors that also influence the selection of candidates; these must be studied with other methods and are therefore not part of our analysis.

Theoretical framework

National context

Despite common organizational features, universities are created and embedded in highly diverse national traditions and governance arrangements that have generated persistent differences between universities in different countries (Clark 1983). These systems, with their

rules, norms, and traditions, consistently shape universities (Whitley 2003). The national contexts underline the resilience of institutions in universities, and (Colyvas and Powell 2006), for example, showed that new procedures at universities must pass through several phases before gaining sufficient legitimacy to overcome initial resistance.

In a similar argument, the historical institutionalist literature stresses the importance of historical developments on today's decisions and future organizational paths (Mahoney and Thelen 2009; Thelen 1999). This perspective highlights the fact that temporality and context matter for decision-making costs and assessments of alternatives, and it indicates that institutional structures lead to path dependency, feedback mechanisms (Pierson 1993, 2004), and lock-in effects (Sydow et al. 2009). Thus, higher education research often argues that change represents one of the primary challenges for universities, and when change materializes, it proceeds incrementally and mostly through organizational layering (Clark 1983). For our study, this understanding implies that even if there is more internationalization in academic labor markets and recruitment as well as an increasingly global disciplinary community, specific national norms, values, regulations, and structures should still matter in the way evaluative criteria are mobilized by academics.

In this perspective, national legal frameworks and funding arrangements create long-lasting differences between universities in different countries and between national academic labor markets, which in turn can be expected to influence the way how academics embedded in these environments approach evaluative criteria. For example, in many European countries academics in public universities are at least partly regulated by laws that govern public sector employment. Similarly, Aagaard (2015) has demonstrated that national performance-based evaluation systems affect researchers' assessments of peers. Thus, even though academic labor markets have become more international in recent years, it can be assumed that national contexts still matter for recruitment processes. Our first expectation is therefore that researchers in different countries prefer distinctive evaluative criteria due to the specific national, historical, and cultural environments in which universities operate.

Expectation 1: Researchers from similar fields in different countries have distinct preferences regarding evaluative criteria in recruitment processes due to the specific national context in which they operate.

Internationalization

Despite universities' national embeddedness, academia has always been an international endeavor. In the last decades, researchers in organizational studies have highlighted how global reforms have spurred universities to become more alike and linked this development to a process of bureaucratization in which organizations in the same field converge as the field matures (Bromley and Meyer 2015; DiMaggio and Powell 1983; Meyer and Rowan 1977; Ramirez 2006). DiMaggio and Powell (1983) describe this as a process of homogenization or isomorphism in which organizations are drawn toward compatibility with other organizations in the same field. In this understanding, isomorphism is linked to shared values, organizational structures and common beliefs that spread over time throughout the organizational field. The organizational field of universities has a long history of well-established and shared beliefs and rules (Meyer et al. 2007) that define what is perceived as appropriate and what signifies prestige and standing in the academic community. At the same time, especially in recent decades one can observe an increasing prominence of internationalization and a shift in the dominant discourse on higher education (Buckner 2017).

These newly highlighted shared values and norms connect universities across national boundaries and create isomorphic pressures that drive organizations to adapt in order to gain and retain legitimacy (DiMaggio and Powell 1983). Adaption may occur through reform alternating actor's authority as coercive isomorphism or through mimetic isomorphism without structural change (Marini 2020). One example on the latter is that university leaders are increasingly occupied by recognition through league tables, rankings, and other international evaluative instruments (Paradeise and Thoening 2015; Sauder and Espeland 2009). Given the global rationalization of universities (Ramirez 2010) and how recognition and prestige are related to a global organizational field in higher education (Driori et al. 2003; Krücken and Meier 2006), growing internationalization can be expected to give rise to converging preferences for the use of evaluative criteria, thus leading to a more integrated global academic labor market.

While evaluation processes in academic recruitment are conducted in the contexts of universities, they are also embedded in their disciplinary fields (Clark 1978; Lamont 2009) and controlled by peers (Musselin 2010). The fields have their own evaluative cultures closely tied to their epistemological traditions and academic work (Becher and Trowler 1989; Lamont 2009; Välimaa 1998). These are also found in evaluation processes in recruitment in which the fields employ specific evaluative criteria when assessing candidates (Herschberg et al. 2018). Moreover, academics strive primarily for peer-recognition and prestige within their fields (Clark 1983; Driori et al. 2003; Hessels et al. 2019). While disciplinary fields have always had an international orientation (Lamont 2009; Langfeldt et al. 2019), they have in the last decades experienced increased internationalization through expanded participation at international conferences, journals, and academic training arrangements (Whitley et al. 2010). Hence, due to evaluation processes' embeddedness in values and norms of internationally oriented fields, one could expect convergence and isomorphism of preferences regarding evaluative criteria within each field independent of the national context (Buckner 2019; Ramirez 2006).

Expectation 2: Due to increased internationalization and the related isomorphism of disciplinary fields, researchers in the same fields in different countries prefer similar evaluative criteria in recruitment processes.

Empirical context: fields and countries

We selected physics and economics as fields because we expected them to have different criteria and standards for evaluating candidates. As noted by others, economics has a special status in the social sciences; recruiters place considerable value on candidates' publications in top journals, and there is a high level of internal consensus on mainstream or neoclassical economics, which is sustained by a highly international knowledge community (Hylmö 2018; Lee et al. 2013). Furthermore, researchers in economics and social science in general are less functionally dependent on the work of their colleagues (Whitley 2000). In physics, on the other hand, we assume researchers to be dependent on the results and methods of others and research to be a more collaborative effort (Välimaa 1998; Whitley 2000). The two fields have furthermore different scientific publication practices where scientific publications in physics (particularly those building on large experiments) often have a high number of co-authors, while there are relatively few co-authors on publications in economics. The different ways of doing research may further have a bearing on the fields' notion of

quality and preferences for the use of evaluative criteria in hiring processes (Välilmaa 1998). Therefore, we assume that both fields apply different evaluative criteria where researchers in economics should place stronger emphasis on the importance of metrics, while researchers in physics, to a larger extent, emphasize candidates' research profiles in order to assess how they might fit into a research group. This difference between the two fields has also been documented by recent qualitative studies building on confidential recruitment reports (Reymert 2020). Finally, both fields are characterized by a high degree of internationalization. While this increases the likelihood of finding field-specific effects compared with, for example, selecting a more national field, such as history or literature, internationalization is also something of a prerequisite that enables us to distinguish between national effects and field effects as highlighted in expectation 2. At the same time, these fields could be described as least likely to have national differences; therefore, differences that can be found between countries are especially relevant.

Globally, academic recruitment is organized in very different ways. To ensure relatively similar research conditions, we chose to compare North-Western European countries, namely Denmark, the Netherlands, Norway, Sweden, and the UK. These countries have well-developed and well-funded higher education systems, and they have different national contexts and performance-based research funding systems, which we assume to affect the use of evaluative criteria. Sivertsen (2017: 2) identified four ideal types of such systems, and the five countries in our study cover all of them:

- 1 Funding allocation is combined with a research evaluation organized at intervals and based on expert panels' peer review. Bibliometrics inform the panels. The UK and the Research Excellence Framework (REF) is an example here.
- 2 The funding allocation is based on indicators representing research activities. These are used annually and are included directly in the funding formula. Sweden is an example here.
- 3 While similar to the second model, this system's indicators represent several aspects of universities' activities, such as teaching, rather than focusing strictly on research. Examples here are Denmark and Norway.
- 4 This type is similar to the third, but bibliometric indicators are not part of the set of indicators used to distribute funding. The Netherlands is an example here.

We assume that the varying emphasis on citations, publications, and education could impact academics' preferences for the use of evaluative criteria in hiring processes in line with expectation 1. One reason for this is that the national funding systems set differing incentives when considering which criteria are the most promising from a funding perspective.

Data and methods

To investigate our research question, we used a Web-based survey, which was distributed to researchers in 2017/2018. Our target population was academic staff that had been involved in recruitment processes. To generate the respondent list, we pursued a two-step strategy in which we combined journal classification (Web of Science (WoS)) and organizational units to delimit the sample. In total, 59% of respondents were identified from staff lists

Table 1 Number of respondents by field and country

Country	Economics	Physics	Total	Response rate	Economics/Physics
Sweden	57	242	299	27.8%	42.5%
Norway	60	82	142	57.3%	49.9%
UK	32	62	94	11.4%	12.8%
Netherlands	66	120	186	20.0 %	24.2%
Denmark	44	83	127	32.2%	33.1%
Total	259	589	848	31.4%	34.6%

and 41% from WoS data. We removed respondents who declined to participate, those who were outside the target group, and those with nonfunctioning email addresses. The survey achieved an overall response rate of 33.6%, varying from 11.4% for economics in the UK to 57.3% for economics in Norway (see Table 1). The survey was part of a larger international research project, and we have included more information on the survey and its representativity in the [Appendix](#).

Since we aimed to investigate academic's preferences of evaluative criteria in academic recruitment, we singled out respondents who reported participating in recruitment processes (848 of 1697 respondents). Most respondents included in our sub-sample were professors; 80% were male, and more than half were between 40 and 59 years old. Table 2 shows their background information.

Dependent variables

In the survey, we asked the respondents to think about their last assessed candidate and identify which type of position they had assessed for: junior, senior, or other. They were also asked to indicate the importance of 13 predefined evaluative criteria on a scale of "Not important," "Somewhat important," "Highly important," and "Do not remember/cannot answer." In the analyses, we combined "Somewhat important" and "Do not remember/cannot

Table 2 Descriptive statistics for control variables

Statistic	Number	Mean	St. dev	Min	Max
Age: 39 years and younger	848	0.212	0.409	0	1
Age: 40 to 49 years old	848	0.298	0.458	0	1
Age: 50 to 59 years old	848	0.261	0.439	0	1
Age: 60 years and older	848	0.198	0.399	0	1
Gender (female = 1)	823	0.196	0.397	0.000	1.000
Respondents own position: Professor	846	0.459	0.499	0.000	1.000
Respondents own position: Associate Professor	846	0.281	0.450	0.000	1.000
Respondents own position: Assistant Professor	846	0.178	0.383	0.000	1.000
Respondents own position: Leader	846	0.063	0.242	0.000	1.000
Respondents own position: Other	846	0.019	0.136	0.000	1.000
Recruiting to junior position	835	0.725	0.447	0.000	1.000
Recruiting to senior position	835	0.275	0.447	0.000	1.000

Table 3 Abbreviations for predefined evaluative criteria categories in the questionnaire

Short abbreviations	Full text from the survey
Citation numbers	Research achievements: citation impact of past publications
Diversity	Ensure diversity in the group/department (e.g., gender, ethnicity, age)
Future potential	The potential for future achievements
General impression	General impression from interview with candidate
Grants	Ability to compete for research grants
Group standing	Standing of the unit/group where the candidate is/has been working/trained
Language skills	Communication and language skills
Matching field	Matching field/expertise to the needs of the group/unit/project
Publication numbers	Research achievements: number of publications/productivities
Research contributions	Research achievements: important prior research contributions (assessed independently of citation scores and source of publication)
Teaching experience	Teaching experience/achievements (including supervision of students)
Third mission experience	Experience in interacting with the public/users/industry
Third mission work experience	Experience/achievements from work outside science, e.g., professional/clinical practice, industry or public administration

answer” into a neutral category, and for the regression analysis, we computed dummy variables in which “Highly important” was assigned a value of 1 and other answers a value of 0. Since prior studies have shown that some evaluative criteria are more important than others, we also asked the respondents to select the most important aspect of all the criteria they identified as “Highly Important.”

The questions in the survey were based on prior research on academic recruitment processes (Herschberg et al. 2018; Levander et al. 2019; Van den Brink and Benschop 2011) and literature on research quality evaluations (Lamont 2009). We also allowed respondents to define their own criteria if they desired, but only a few submitted self-defined criteria, so additional categories were not constructed. As shown in Table 3, we compressed the descriptions from the survey into shorter abbreviations, which we refer to in the “Discussion” section.

One limitation of surveys is that they do not offer detailed answers. In this study, the category *Future Potential* posed some interpretation challenges, as it may have included issues such as future potential in terms of research contributions, teaching, or being a good colleague. However, since the survey primarily addressed issues around research and conditions, we believe that most respondents associated this category with research-related practices. We thus understand *Future Potential* as an indirect reference to future research contributions.

We first used bivariate correlation patterns before applying logistic regression analysis with control variables.

Independent variables

In the logistic regression analysis, we controlled for the country in which the respondents worked and their field affiliation. We further controlled for whether the respondents

recruited for a junior or senior position and for background variables, such as gender, age, and their own academic positions. We refrained from analyzing institutional differences because we had relatively few respondents from each institution (only one institution had more than 17 respondents in one of the fields).

Methods

We analyzed the data using R.¹ To ensure equal field and country compositions in the bivariate data presentations, we developed and applied weights (see [Appendix Table 5](#)).

Binary logistic regression analyses

To analyze country and field effects on which evaluative criteria researchers prefer, we applied binary logistic regression analysis with the different evaluative criteria as independent variables. Before conducting the analyses, we investigated the Pearson r correlation to control that none of the independent variables were highly correlated (see [Fig. 1 in the Appendix](#)). We applied ANOVA tests to investigate whether the independent variables contributed significantly in explaining the variance of the dependent variable (Agresti 2013) and conducted AIC and BIC tests to detect the models that were most suited to explaining this variance (Agresti 2013). The best-suited models are shown in the paper, while the others are available in the [Appendix](#). All binary logistic regression models were conducted with the different countries as baseline categories to map country effects, but only the models with the Netherlands as a baseline category are shown since the Dutch respondents had the most deviant answers; hence, these models show most of the significant effects that we discovered. To investigate interaction effects between country and field, we conducted separate regression analyses for physics and economics instead of including interaction terms in the regression models because of the related problems of including interaction terms in logistic regression models with relatively low numbers of available observations (Mood 2009).²

Results

The respondents identified multiple evaluative criteria as important, and only a few criteria were classified as irrelevant. For instance, the only criterion that was identified as not important by more than half of the respondents was *Third Mission Activities*. However, some criteria were more important than others, and respondents placed the most value in *Future Potential*, *Matching Field*, *General Impression*, and *Important Research Contribution*.

¹ The RMarkdown file is available on request.

² Physics is a large and heterogenous field. Some researchers depend on large international infrastructure (such as ATLAS), while others primarily work by themselves and without large equipment. To test for these differences, we grouped the participants based on whether they depended on large infrastructures. We did not find significant differences between the two, and thus we treated physics as one field.

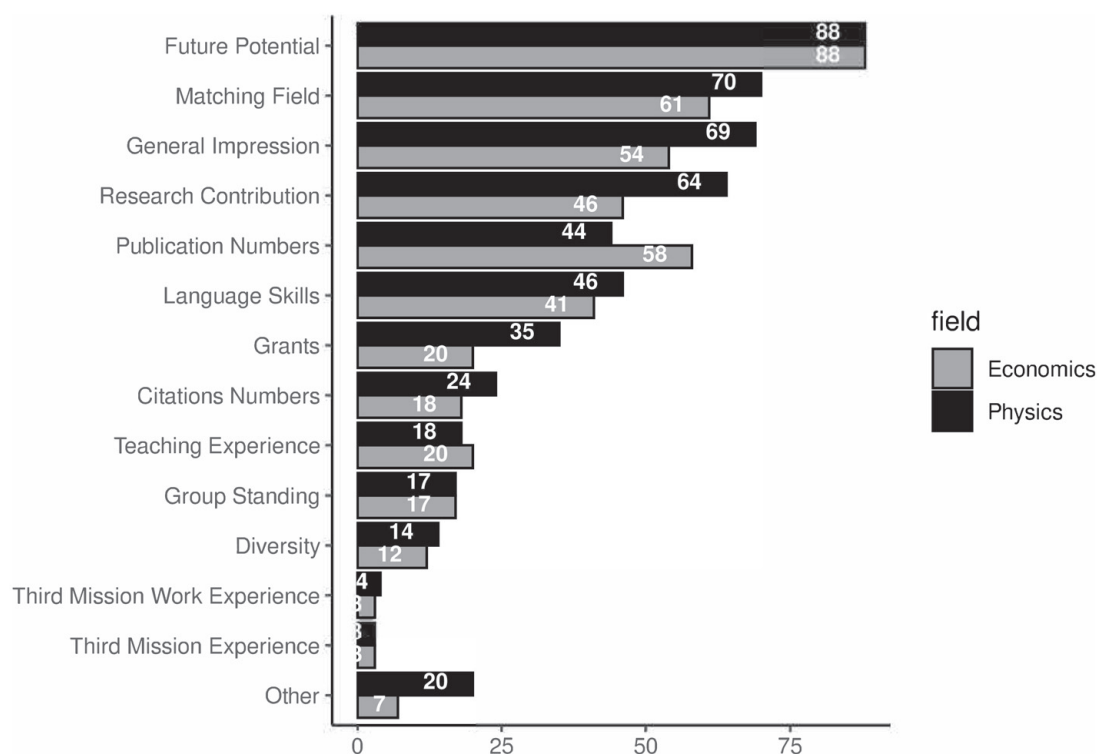


Fig. 1 Highly important evaluative criteria by field. Percent. $N = 419/419$

Candidates' *Future Potential* was the most important criterion in both fields. However, as shown in Fig. 1, we found differences between the fields. For instance, physicists more often identified *Matching Field*, *General Impression*, and *Important Research Contribution* as highly important, while economists more frequently valued *Publication Numbers* as highly important.

These academic field differences were confirmed by binary logistic regression analysis using the nine most common evaluative criteria as the dependent variables. The dot-and-whisker plot in Fig. 2 displays the field coefficients as part of the physics field—as opposed to economics—with standard errors (see Table 1.1 in the Appendix). The results indicated that being an economist instead of a physicist increased the probability of identifying *Publication Numbers* as a highly important evaluative criterion from 55 to 69% and decreased the probability of identifying *Research Contribution* as a highly important criterion from 74 to 59%.³ There were, however, no significant differences in how respondents from either field valued *Future Potential*, *Citation Numbers*, or *Teaching Experience*.

In the analyses, we only detected moderate country differences. Figure 3 shows the percentage of respondents who identified the nine most important criteria as highly important by country.

The binary logistic regression analysis further confirmed the moderate country differences (see Table 1.1 in the Appendix), and the ANOVA tests showed that there were no country differences for the criteria *Citations*, *Future Potential*, *Grants*, and *Research*

³ The effects were computed for male researchers between 50 and 59 years old recruiting a senior academic in Netherlands.

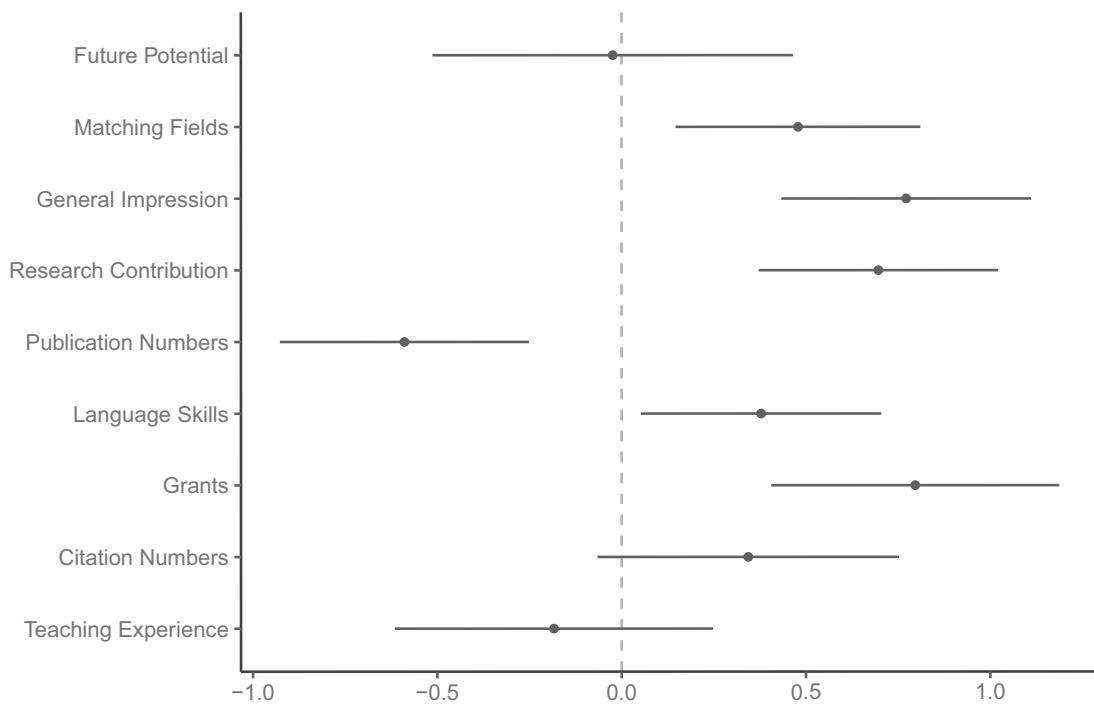


Fig. 2 Dot-and-whisker plots from regression analysis. Evaluative criteria. Field differences Physics. Coefficient with economics as baseline category. Coefficient from regression in [Appendix Table 1.1](#)

Collaboration. Moreover, the country effects were mainly due to the Dutch and, to some extent, the Norwegian respondents' answers. The Dutch respondents less frequently valued *Publication Numbers* as highly important criteria and more often valued *Language*

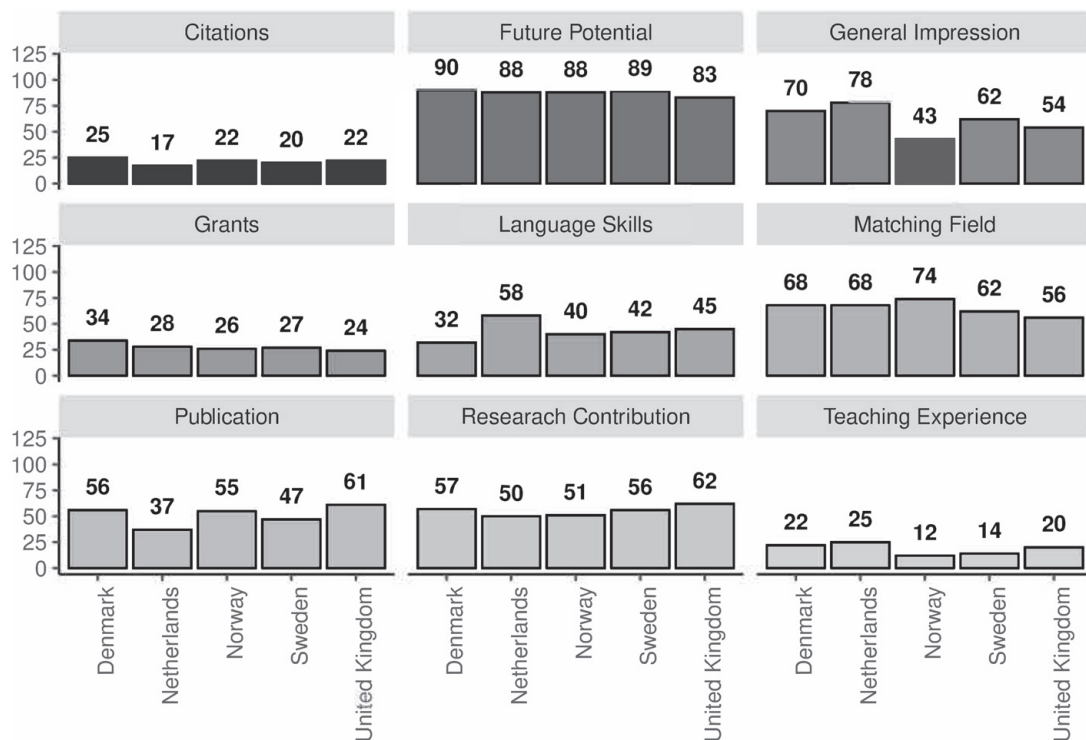


Fig. 3 Highly important evaluative criteria by country. Percent. $N = 168$ per country

Skills, *General Impression*, and *Teaching Experience* as highly important compared with many of their international colleagues.⁴ While there was only a 55% possibility for Dutch researchers to identify *Publication Numbers* as highly important, the probability for researchers from other countries was between 66 to 75%.⁵ Norwegian respondents also showed some deviant answers valuing *Matching Field* higher and *General Impression* lower than many of their international colleagues.⁶ The dot-and-whiskers plot in Fig. 4 displays the country coefficients with the Netherlands as the dotted baseline category.

We further checked for interaction effects between country and field by conducting separate regression analyses for physicists and economists (in the Appendix, see Table 3.1 for economics and Table 4.1 for physics). These analyses confirmed that there were only moderate country differences in both fields and that the country differences were similar (e.g., both the Dutch economists and physicists were less inclined to identify *Publication Numbers* as highly important criterion). At the same time, we found that the country variations played out differently in the two fields. For instance, Dutch physicists more often identified *Important Research Contribution* as a highly important evaluative criterion than their Norwegian colleagues, while there was no such significant difference between Dutch and Norwegian economists.

We also detected significant differences in applied evaluative criteria in relation to the type of position for which the respondents were recruiting. When recruiting seniors, the respondents were more inclined to emphasize *Important Research Contribution*, *Publication Numbers*, *Citations Numbers*, *Grants*, or *Teaching Experience*, whereas *Future Potential*, *Matching Field*, *General Impression*, and *Language Skills* were more important when recruiting juniors (see Table 1.1 in the Appendix). These effects had a further and quite substantial impact on the probability of valuing the different evaluative criteria. For example, recruiting to a senior position instead of a junior position increased the possibility of highlighting *Publication Numbers* from 39 to 69% and similarly raised the likelihood of highlighting *Teaching Experience* from 18 to 39%.⁷

We also controlled for the respondents' background variables, such as age, gender, and their positions (in the Appendix, see Tables 1.1–1.8). In these analyses, we observed that respondents over 40 years were more inclined to identify *Publication Numbers* as the most important evaluative criterion, and professors were more inclined to emphasize *Future Potential*, *Important Research Contribution*, *Publication Numbers*, *Grant*, and *Teaching Experience* than associate or assistant professors. However, the effects of these background variables were relatively small and did not alter the country or field effects.

Respondents were also asked to identify the single most important criterion of those criteria selected as “Highly Emphasized.” Despite the plethora of important criteria, only a

⁴ The logistic regression models with the different countries as baseline categories showed that Dutch respondents significantly valued *Publication Numbers* less than their Norwegian, Danish, and British colleagues. Moreover, they valued *Language Skills* more than Swedish, Danish, and Norwegian respondents, *General Impression* more than Norwegian, Swedish, and British respondents, and *Teaching Experience* more than Swedish and Norwegian respondents.

⁵ The effects were computed for male researchers between 50–59 years old in physics recruiting to senior positions.

⁶ The logistic regression models with different countries as baseline categories showed that the Norwegian respondents significantly valued *Matching Field* more than British and Swedish respondents and *General Impression* less than Danish, Swedish, and Dutch respondents.

⁷ The effects were computed for male economists between 50 and 59 years in the Netherlands.

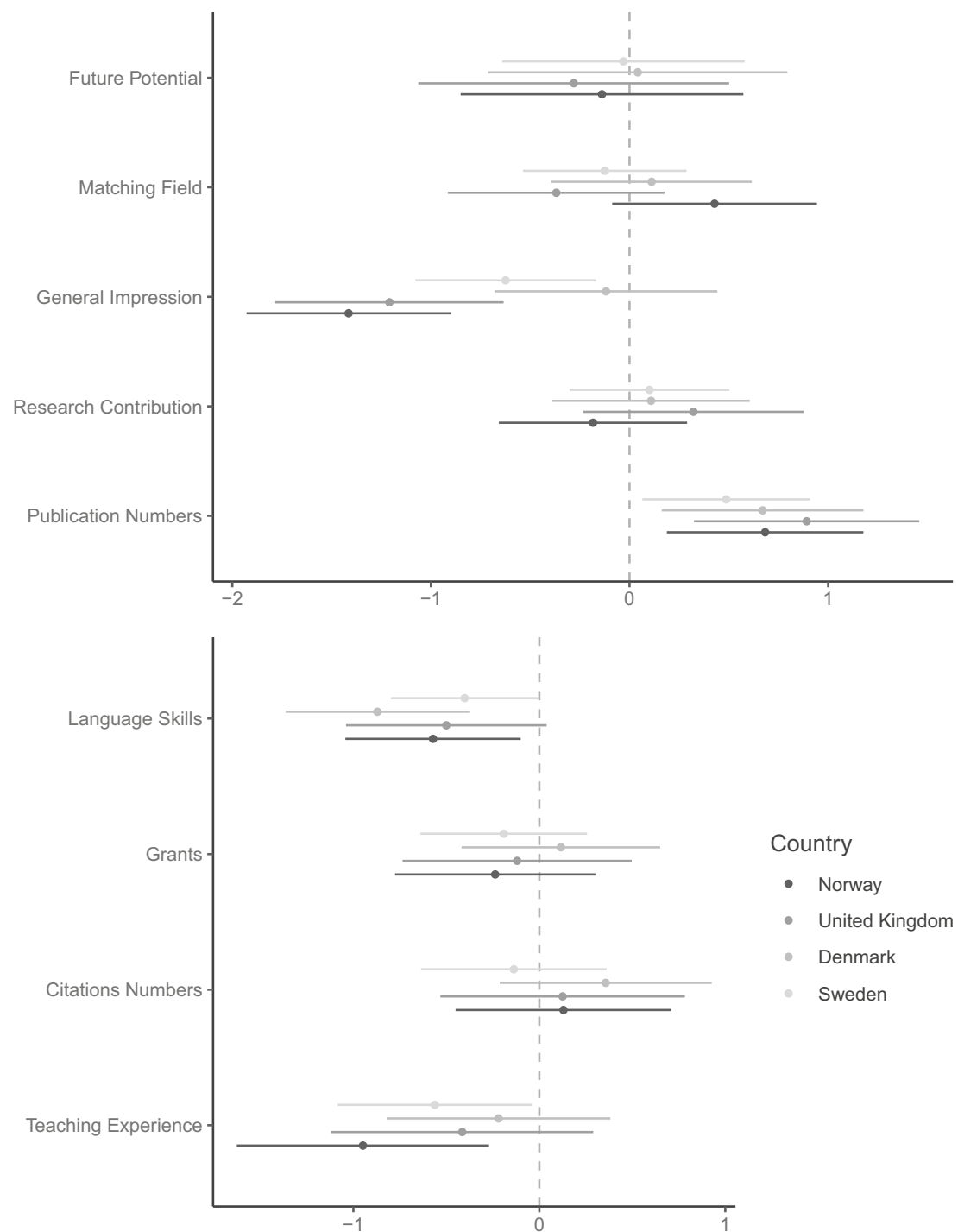


Fig. 4 (1/2) Dot-and-whisker plots from regression analysis. Evaluative criterias. Country differences. Netherlands as baseline category. Coefficient from regression in [Appendix Table 1.1](#)

few were identified as the most important, with 93% of respondents selecting either *Future Potential*, *Matching Field*, *Important Research Contribution*, *Publication Numbers*, or *General Impression* (see Fig. 5). Being important did not imply that the criterion was the single most important. For instance, although 44% of respondents identified *Language and Communication Skills* as a highly important, less than 1% identified it as the most important. Furthermore, it is noteworthy that four of the five most important criteria reflected

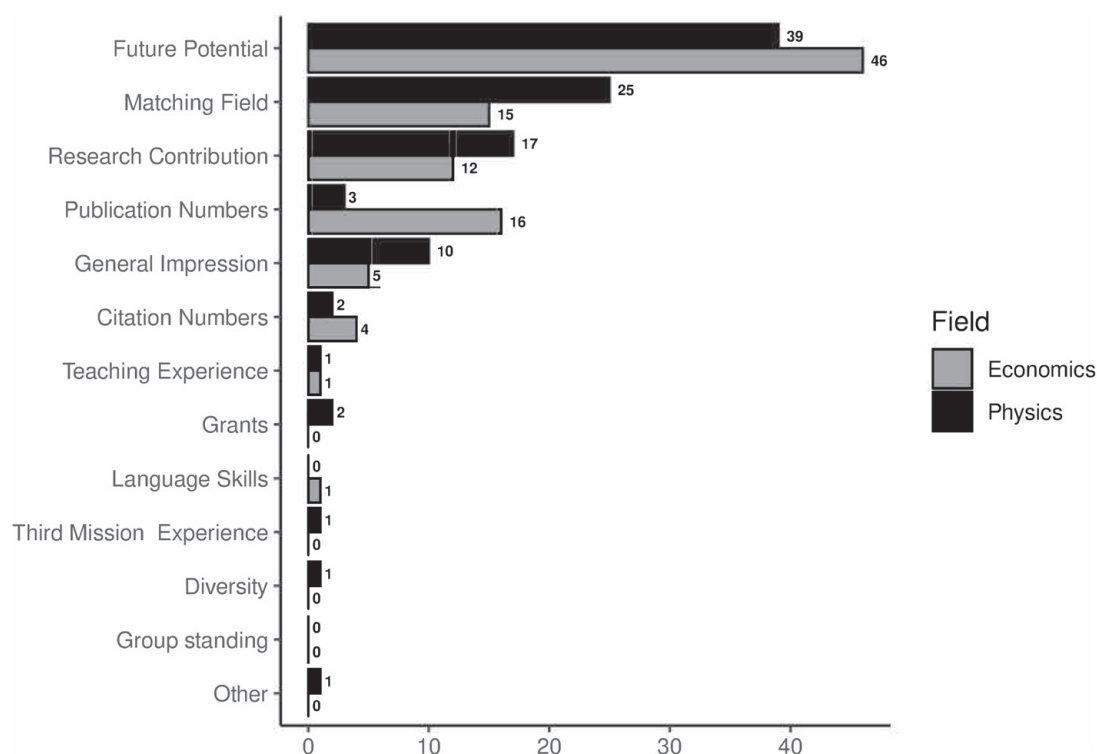


Fig. 5 Most important evaluative criteria by field. Percent. Economics ($N = 375$), Physics ($N = 399$)

research performance (*Future Potential*, *Matching Field*, *Important Research Contribution*, *Publication Numbers*). Hence, although candidates were evaluated on their teaching experience, language skills, grants experience, and third mission experience, their research performance was ultimately the most important.

The binary logistic regression analysis using the five most important evaluative criteria as dependent variables further confirmed the strong field differences and moderate country differences shown above (see [Table 2.1 in the Appendix](#)). For instance, being a physicist rather than an economist increased the possibility of identifying *Matching Field* as the most important criterion from 6 to 11% but decreased the possibility of identifying *Publication Numbers* as the most important criterion from 13 to 2%.⁸ Conversely, being an economist rather than a physicist increased the likelihood of identifying *Publication Numbers* as the most important criterion from 4 to 22%.⁹

However, the country effects were rather moderate. The ANOVA test showed that the country only contributed to the variance of *Matching Field* with significant explanations, whereas Norwegian respondents preferred *Matching Field* more than Swedish, Danish, and Dutch respondents. The separate regression analysis for the two fields (see [Tables 3.2 and 4.2 in the Appendix](#)) further confirmed the moderate country differences within the two fields, and to some extent, it revealed that there could be slightly greater significant differences in physics than in economics. However, there were more physicists than economists in the sample, so this result could be due to the larger number of available observations.

⁸ The effects were computed for male researchers between 50 and 59 years old recruiting for a senior position in the Netherlands.

⁹ The effects were computed for Swedes recruiting to a senior position.

The regression analysis moreover confirmed that the evaluative criteria depended strongly on the type of position for which respondents were recruiting, with senior positions relying more strongly on *Research Contribution* and *Publication Numbers*, while *Future Potential*, *Matching Field*, and the *General Impression* were more frequently preferred in junior recruitment. Additionally, we controlled for age, position, and gender differences (see [Tables 2.6–2.8 in the Appendix](#)), which turned out to have relatively small effects and did not alter the country or field differences.

Discussion

Initially, we suggested two somewhat contradictory expectations based on different strands of the literature. Our first expectation suggested that national differences exist in evaluative criteria preferences due to different national contexts (Clark 1983) and path dependencies (Thelen 1999). The second suggested that increasing internationalization of disciplines would lead to isomorphism and a prominence of field-dependent preferences (Lamont 2009). In line with the second expectation and based on prior studies, we expected that economists would more strongly value candidates' bibliometrics (Hylmö 2018), while physicists would place more value on their research profiles due to a larger degree of functional dependency among researchers (Reymert 2020; Whitley 2000). In line with our first expectation, we expected researchers in the Netherlands to be least concerned with bibliometrics because their funding model does not include bibliometric indicators (Aagaard 2015; Sivertsen 2017). Our results support both expectations, although the field differences were stronger than country differences.

We found moderate country differences, but considering our case selection, those that we found were especially relevant. The most striking country difference was that Dutch scholars placed a lower emphasis on publication, which may be explained by the fact that bibliometric indicators are not included in the Dutch performance-based funding system, and thus, there is less of an incentive to assure certain publication patterns in newly hired staff (Sivertsen 2017). This finding may suggest that the indicators in the performance-based research system have trickle-down effects on recruitment (Aagaard 2015). Moreover, while the Netherlands was a frontrunner in establishing teaching programs in many disciplines using English as language of instruction, recent policy debates have increasingly highlighted the importance of the Dutch language again, which could also explain the importance of language skills in Dutch responses (Duarte and van der Ploeg 2019). These country differences give some support to our first expectation of national embeddedness, especially considering that our field selection had a slight bias toward internationalized fields in which country differences should be less likely. However, the national differences were rather moderate, and we primarily discovered national similarities, as Musselin (2010) also observed, giving stronger support to our second expectation and the internationalization perspective (DiMaggio and Powell 1983; Ramirez 2006).

In addition, we found strong field differences, which supported our second expectation that evaluative criteria were field-specific (Musselin 2010; Van den Brink and Benschop 2011). For example, our results showed that economists assessed the candidates on their publication records, while physicists relied more on their important research contributions and the relevance of their research profiles. These results aligned with prior studies on academic recruitment processes, which have shown that economists emphasize publications in top journals (Hylmö 2018). The physicists put stronger emphasis on candidates' matching

research profiles and language skills aligned with prior studies (Reymert 2020); moreover, their emphasis on general impression could be understood in terms of how they work in research groups where individual researchers have a specific role (Whitley 2000), thereby emphasizing a more pressing need to select candidates with compatible profiles and competencies. These field differences were similar across national borders, indicating that despite national differences in recruitment (Alfonso 2016), the evaluation processes are embedded in their fields' epistemic traditions (Lamont 2009), which are less sensitive to national considerations (Musselin 2010). However, as Musselin (2010), we found moderate differences in criteria preferences across countries, indicating that evaluation processes are to some extent affected by the national context.

Our results thus show that the particular evaluative cultures in recruitment were primarily embedded in the fields and, to some extent, national contexts. This finding may imply that we should regard the international academic labor market as layered and multiple rather than singular. In it, processes are nationally regulated (Alfonso 2016), but the evaluation processes are particularly tied to different internationally oriented fields, with their evaluative cultures deeply embedded in their epistemic traditions and academic work (Lamont 2009; Välimaa 1998).

Finding partial support for both of our somewhat contradictory expectations opens the question of what kind of mechanism could be driving this development. In line with Christensen et al. (2014), one could argue that disciplines provide global norms regarding the preferences of evaluative criteria and that these global norms are then filtered when they are applied in a specific national context. In this understanding, disciplines are the main normative framework for academics, while national frameworks such as laws or funding systems mediate the application of these norms. Similar mechanisms that combine global and national factors have already been identified, for example, regarding questions of internationalization (Buckner 2019; Buckner 2020).

We further observed that despite the preference of multiple evaluative criteria, only a few were identified as most important, and these reflected primarily the candidates' research output, which other studies have also shown as the most decisive criteria in recruitment (Van den Brink and Benschop 2011). This criteria concentration could be understood by the stagewise nature of recruitment processes in which candidates are first met with formal standards or screened using their CVs before an expert committee undertakes a more thorough evaluation of their research (Hamann 2019; Musselin 2010). Additionally, we found that the evaluation of candidates for different types of positions required different evaluative criteria, which suggests that further studies are needed to identify how different evaluative criteria are used for different positions.

Conclusion

Academic recruitment and academics' emphasis on evaluative criteria in such processes are crucial for universities, as recruitment represents the basis for acquiring their key resource, namely talented academics. Over the past few decades, academic career structures and the academic labor market have become increasingly internationalized, with a growing number of international researchers and an increase in universities competing for the best scholars (Gornitzka and Langfeldt 2008). Still, local academic labor markets are embedded in national higher education systems and matching legal frameworks (Musselin 2005). Academics' preferences for different evaluative criteria in recruitment processes can

be seen as a key indicator of the degree of internationalization of an academic labor market. However, most studies on hiring processes have hitherto focused on a single country, and comparative studies are lacking; therefore, the question of whether fields apply similar or different evaluative criteria when evaluating candidates for academic positions has gone unanswered (e.g., Hylmö 2018; Levander et al. 2019; Van den Brink and Benschop 2011). This paper has targeted this knowledge gap.

In this study, we mainly observed field differences, but we also found moderate national differences. For instance, economists valued applicants' publication numbers higher than physicists, who emphasized candidates' research contributions. These findings align with prior studies addressing the fields' different evaluative cultures (Lamont 2009) and support the enduring importance of norms and values stemming from the research field's definition of peer recognition and prestige (Clark 1983; Driori et al. 2003; Hessels et al. 2019; Langfeldt et al. 2019). They further underline the differences in research practices and collaboration patterns in different research fields, and that these have to be taken into account in studies of recruitment processes. While other studies have claimed that the evaluation processes seem less sensitive to national considerations (Musselin 2010), our study found moderate but important country differences. It shows that having a bibliometric indicator in the performance-based research funding system seems to affect the emphasis on scientific publication when academics evaluate candidates for positions. Despite this, the findings first and foremost underline the internationalization of recruitment processes, showing that evaluative cultures are strongly embedded in internationally oriented research fields and their evaluative cultures, while being affected to a more limited degree by national frameworks. We argue that the mechanism behind this development is that disciplines provide global norms regarding the preferences of evaluative criteria and that these global norms are then mediated by national structures such as funding systems when they are applied in a specific context. These findings may imply that we should regard the international academic labor market not as singular but as layered and affected by multiple considerations.

We also note that our study only offers insights into one element of hiring processes, namely academics' preferred evaluative criteria. The ranking of candidates and their ultimate selection may be strategic (Musselin 2010), and other factors, such as academic inbreeding, or informal factors in candidate selection were not included in this study. The moderate country differences found in this study should also be seen in relation to the selection of relatively similar countries in North-Western Europe; studying countries in different corners of the world would probably have rendered stronger results. Moreover, only cross-sectional data were available and given the steady increase in the internationalization of higher education, tracing developments not only across countries but also over time would be highly desirable. In addition, the inclusion of more countries as well as a greater variety of fields would be an important avenue for future research. Finally, using other methods, such as interviews or participant observation, to include non-formal aspects of hiring processes would also help in the development of a more complete understanding of recruitment processes. However, the strength of this paper is that our approach enabled us to contribute with a systematic overview and comparison of evaluative preferences in different countries and fields.

Our findings have some policy implications. First, there seems to be a general understanding that bibliometric indicators have a dominant role in the evaluation of candidates. Our findings call for a more nuanced picture by accounting for field characteristics and national contexts. Moreover, removing bibliometric indicators from the performance-based research funding system seems to influence the preferences of such indicators in hiring processes. Furthermore, such indicators seem to be primarily preferred in fields such as

economics, which is often characterized as exhibiting a rather deviant and extreme behavior compared with other disciplines (Hammarfelt and Rushforth 2017; Hylmö 2018). Second, it seems that teaching and third mission activities are not seen as important when assessing candidates for positions, a finding that has also been put forward in previous studies (Levander et al. 2019). This has implications for the development of policies that emphasize the link between research and education and the importance of third mission activities, and it should be a subject of future study.

Supplementary Information The online version contains supplementary material available at (<https://doi.org/10.1007/s10734-020-00659-3>).

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, K. (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>
- Agresti, A. (2013). *Categorical Data Analysis*. New Jersey: John Wiley & Sons.
- Alfonso, A. (2016). Varieties of Academic Labor Markets in Europe. *PS: Political Science & Politics*, 49(4), 816–821.
- Altbach, P. G., Yudkevich, M., & Rumbley, L. E. (2015). Academic inbreeding: local challenge, global problem. *Asia Pacific Education Review*, 16(3), 317–330. <https://doi.org/10.1007/s12564-015-9391-8>
- Becher, T., & Trowler, P. R. (1989). *Academic Tribes and Territories. Intellectual enquiry and the culture of disciplines.*, USA/UK: The Society for Research into Higher Education and Open University Press
- Bromley, P., & Meyer, J. W. (2015). *Hyper-organization*. Global organizational expansion, Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199689859.001.0001>
- Buckner, E. (2019). The internationalization of higher education: National interpretations of a global model. 63(3), <https://doi.org/10.1086/703794>
- Buckner, E. (2020). Embracing the global: the role of ranking, research mandate, and sector in the internationalisation of higher education. *Compare: A Journal of Comparative and International Education*, 1–18. <https://doi.org/10.1080/03057925.2020.1753499>
- Buckner, E. S. (2017). The changing discourse on higher education and the nation-state, 1960–2010. *Higher Education*, 74(3), 473–489. <https://doi.org/10.1007/s10734-016-0056-2>
- Chou, M.-H., & Gornitzka, Å. (2014). *Building the Knowledge Economy in Europe*. Massachusetts: Edward Elgar Publishing Limited.
- Christensen, T., Gornitzka, Å., & Maasen, P. (2014). Global Pressures and National Cultures, In P. Mattei, (Ed.), *University Adaption in Difficult Economic Times*. (pp. New York: Oxford University Press
- Clark, B. (1978). Academic differentiation in national systems of higher education. *Comparative Education Review*, 22(2), 242–258.
- Clark, B. (1983). *The higher education system: academic organization in cross-national perspective*. Berkeley, California: University of California Press.
- Colyvas, J. A., & Powell, W. W. (2006). Roads to institutionalization: The remaking of boundaries between public and private science. *Res Organ Behav*, 27, 305–353
- Combes, P.-P., Linnemer, L., & Visser, M. (2008). Publish or peer-rich? The role of skills and networks in hiring economics professors. *Labour Economics*, 15(3), 423–441. <https://doi.org/10.1016/j.labeco.2007.04.003>
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organization fields. *American Sociological Review*, 48(147–160).

- Drori, G., Meyer, J. W., & Ramirez, F. (2003). *Science in the modern world polity: institutionalization and globalization*. Stanford: Stanford University Press.
- Duarte, J., & van der Ploeg, M. (2019). Plurilingual lecturers in English medium instruction in the Netherlands: the key to plurilingual approaches in higher education? *European Journal of Higher Education*, 9(3), 268–284. <https://doi.org/10.1080/21568235.2019.1602476>
- Finkelstein, M. J. (2015). How national contexts shape academic careers: A preliminary analysis, In U. Teichler and W. K. Cummings, (Ed.), *Forming, recruiting and managing the academic profession*. (pp. 317–328). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-16080-1_17
- Gornitzka, Å., & Langfeldt, L. (2008). *Borderless knowledge? Understanding the “new” internationalisation of research and higher education in Norway*. Dordrecht: Springer.
- Hamann, J. (2019). The making of professors: Assessment and recognition in academic recruitment. 49(6), 919–941. <https://doi.org/10.1177/0306312719880017>
- Hammarfelt, B., & Rushforth, A. (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>
- Henningsson, M., Jörnsten, A., & Geschwind, L. (2017). Translating tenure track into Swedish: Tensions when implementing an academic career system. *Studies in Higher Education*, 48(7), 1215–1226.
- Herschberg, C., Benschop, Y., & Van den Brink, M. (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>
- Hessels, L. K., Franssen, T., Scholten, W., & de Rijcke, S. (2019). Variation in valuation: How research groups accumulate credibility in four epistemic cultures. *Minerva*, 57(2), 127–149. <https://doi.org/10.1007/s11024-018-09366-x>
- Husu, L. (2000). Gender Discrimination in the Promised Land of Gender Equality. *Higher Education in Europe*, 25(2), 221–228. <https://doi.org/10.1080/713669257>
- Hylmö, A. (2018). *Disciplined reasoning: Styles of reasoning and the mainstream heterodoxy divide in Swedish economics*. Lund, Sweden: Lund University.
- Krücken, G., & Meier, F. (2006). Turning the University into and Organizational Actor In G. Drori, R. Meyer, and H. Hwang, (Ed.), *Globalization and Organization. World Society and Organizational Change*. (pp. New York: Oxford University Press
- Lamont, M. (2009). *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.
- Langfeldt, L., & Kyvik, S. (2011). Researchers as evaluators: Tasks, tensions and politics. *Higher Education*, 62, 199–212. <https://doi.org/10.1007/s10734-010-9382-y>
- Langfeldt, L., Nedeva, M., Sörlin, S., & Thomas, D. (2019). Co-existing notions of research quality: A framework to study context-specific understandings of good research. *Minerva*, 1-23. <https://doi.org/10.1007/s11024-019-09385-2>
- Lee, F. S., Pham, X., & Gu, G. (2013). The UK Research Assessment Exercise and the narrowing of UK economics. *Cambridge Journal of Economics*, 37(4), 693–717. <https://doi.org/10.1093/cje/bet031>
- Levander, S., Forsberg, E., & Elmgren, M. (2019). The meaning-making of educational proficiency in academic hiring: a blind spot in the black box. *Teaching in Higher Education*. <https://doi.org/10.1080/13562517.2019.157660>
- Lutter, M., & Schröder, M. (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>
- Mahoney, J., & Thelen, K. (2009). *Explaining institutional change: Ambiguity, agency and power*. Cambridge, UK: Cambridge University Press.
- Marini, G. (2020). Coercive and mimetic isomorphism as outcomes of authority reconfigurations in French and Spanish academic career systems. *Policy Reviews in Higher Education*, 1-20. <https://doi.org/10.1080/23322969.2020.1806726>
- Meyer, J. W., Ramirez, F., Frank, D. J., & Schofer, E. (2007). Higher education as an institution. In P. Gumpert (Ed.), *Sociology of Higher Education: Contributions and Their Contexts* (pp. 187–221). Baltimore, MD: Johns Hopkins University Press.
- Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, 83(2), 340–363. <https://doi.org/10.1086/226550>
- Mood, C. (2009). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, 26(1), 67–82. <https://doi.org/10.1093/esr/jcp006>
- Musselin, C. (2005). European academic labor markets in transition. *Higher Education*, 49(1), 135–154. <https://doi.org/10.1007/s10734-004-2918-2>

- Musselin, C. (2010). *The Market for Academics [Les marché des universitaires]*. New York, USA: Routledge.
- Nielsen, M. W. (2016). Limits to meritocracy? Gender in academic recruitment and promotion processes. *Science and Public Policy*, 43(3), 386–399. <https://doi.org/10.1093/scipol/scv052>
- Ortiga, Y. Y., Chou, M.-H., & Wang, J. (2020). Competing for academic labor: Research and recruitment outside the academic center. *Minerva*. <https://doi.org/10.1007/s11024-020-09412-7>
- Paradeise, C., & Thoenig, J.-C. (2015). *In search of academic quality*. UK: Palgrave Macmillan.
- Pezzoni, M., Sterzi, V., & Lissoni, F. (2012). Career progress in centralized academic systems: Social capital and institutions in France and Italy. *Research Policy*, 41(4), 704–719. <https://doi.org/10.1016/j.respol.2011.12.009>
- Pierson, P. (1993). When effect becomes cause: Policy feedback and political change. *World Politics*, 45(4), 595–628. <https://doi.org/10.2307/2950710>
- Pierson, P. (2004). *Politics in time: History, institutions, and social analysis*. Princeton: Princeton University Press.
- Reymert, I. (2020). Bibliometrics in academic recruitment: A screening tool rather than a game changer. *Minerva*. <https://doi.org/10.1007/s11024-020-09419-0>
- Ramirez, F. (2006). The rationalization of universities, In M. Djelic and K. Sahlin, (Ed.), *Transnational governance: institutional dynamics of regulation* (pp. 225–244). Cambridge University Press. <https://doi.org/10.1017/CBO9780511488665.012>
- Ramirez, F. (2010). Accounting for excellence: Transforming universities into organizational actors. In L. M. Portnoi, V. D. Rust, & S. S. Bagley (Eds.), *Higher education, policy, and the global competition phenomenon* (pp. 43–58). Basingstoke: Palgrave Macmillan.
- Rossier, T. (2020). Accumulation and conversion of capitals in professorial careers. The importance of scientific reputation, network relations, and internationality in economics and business studies. *Higher Education*. <https://doi.org/10.1007/s10734-020-00508-3>
- Sanz-Menéndez, L., & Cruz-Castro, L. (2019). University academics' preferences for hiring and promotion systems. *Eur J High Educ*, 9(2), 153–171. <https://doi.org/10.1080/21568235.2018.1515029>
- Sauder, M., & Espeland, W. N. (2009). The discipline of rankings: Tight coupling and organizational change. *American Sociological Review*, 74(1), 63–82. <https://doi.org/10.1177/000312240907400104>
- Scott, W. R., & Davis, G. F. (2007). *Organizations and organizing: Rational, natural, and open system perspectives*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications*, 3. <https://doi.org/10.1057/palcomms.2017.78>
- Sydow, J., Schreyögg, G., & Koch, J. (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>
- Tagliaventi, M. R., Carli, G., & Cutolo, D. (2020). Excellent researcher or good public servant? The interplay between research and academic citizenship. *Higher Education*, 79(6), 1057–1078. <https://doi.org/10.1007/s10734-019-00456-7>
- Tavares, O., Sin, C., & Lança, V. (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, K. (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, M., & Benschop, Y. (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, M., Fruytier, B., & Thunnissen, M. (2013). Talent management in academia: Performance systems and HRM policies. *Human Resource Management Journal*, 23(2), 180–195.
- Välilmaa, J. (1998). Culture and identity in higher education research. *Higher Education*, 39(2), 119–138. <https://doi.org/10.1023/A:1003248918874>
- Wennerås, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, 387, 341–343.
- Whitley, R. (2000). *The intellectual and social organization of the sciences*: Oxford University Press on Demand.
- Whitley, R. (2003). Competition and pluralism in the public sciences: the impact of institutional frameworks on the organisation of academic science. *Research Policy*, 32(6), 1015–1029.
- Whitley, R., Gläser, J., & Engwall, L. (2010) Reconfiguring knowledge production: Changing authority relationships in the sciences and their consequences for intellectual innovation, Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199590193.001.0001>

Appendix

Recruitment Reports

The recruitment reports were obtained from three Norwegian universities in economics, sociology, physics and informatics between 2000 and 2017. Appendix Table 1 display information of the documents included in the analyses.

Appendix Table 1. Information about the documents included in the analysis.

Explanation: N (Norwegian) and E (English) refer to the language and x indicates that documents do not exist

University A

		2000-2005		2006-2012				2013-2017				
		1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010
Economics	Case nr.											
	Announcement text	N	N	N	E	N	E	E	E	E	E	E
	Selection Committee	x	x	x	x	x	x	E	E	E	E	E
	Expert Committee	N	E	E	E	E	E	E	E	E	E	E
	Interview Committee	X	X	X	X	X	X	X	X	E	N	N

		2000-2005		2006-2012				2013-2017				
		1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111
Sociology	Case nr.											
	Announcement text	N	N	N	N	N	E	E	E	E	N	E
	Selection Committee	x	x	x	x	x	E	E	E	E	x	E
	Expert Committee	N	N	N	N	E	E	E	E	E	N	E
	Interview Committee	N	N	X	X	X	X	N	X	X	N	X

		2006-2012						2013-2017				2000-205					
		1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1212	1213	1214	1216	1217	1220
Physics	Case nr.																
	Announcement text	E	E	E	E	E	E	E	E	E	E	E	x	N	N	N	
	Selection Committee	x	x	x	x	x	x	x	x	x	E	x	x	x	x	x	
	Expert Committee	E	E	E	E	E	E	E	E	E	E	E	E	N	E	E	
	Interview Committee	N	N	N	N	N	N	x	N	N	N	N	N	x	x	x	N

		2000-2015						2006-2012				2013-2017										
		1314	1315	1318	1321	1322	1325	1327	1328	1331	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312
Informatics	Case nr.																					
	Announcement text	N	N	N	N	x	x	N	N	N	E	N	N	N	E	E	E	N	E	E	E	N
	Selection Committee	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	Expert Committee	E	E	E	E	E	E	E	N	E	E	N	N	E	N	E	E	E	E	E	E	N
	Interview Committee	N	N	x	N	N	x	x	x	x	x	x	N	x	x	N	E	x	E	N	N	x

University B

2013-2017		
Case nr.	2107	2106
Announcement text	N	N
Selection Committee	E	x
Expert Committee	E	E
Interview Committee	N	N

2013-2017		
Case nr.	2214	2213
Announcement text	A	A
Selection Committee	x	x
Expert Committee	A	A
Interview Committee	N	N

University C

2013-2017			
Case nr.	3101	3102	3103
Announcement text	N	N	N
Selection Committee	X	X	X
Expert Committee	E	E	N
Interview Committee	x	x	X

Appendix Table 2. General categories of assessment criteria in academic recruitment processes

Assessment criteria category	Definition
Research quality	
<i>Qualitative research evaluations</i>	Qualitative research evaluations of the candidate's research
<i>Matching research profile</i>	Assessment of the match between the candidate's research profile and the specifics of the announcement text
<i>Metrics</i>	Quantitative analysis of scientific and scholarly publications
<i>Future potential</i>	The candidate's potential for future research output
Teaching	Teaching experience: supervising, pedagogic courses, formal pedagogic competence, and pedagogic publications
Third mission activities	The diffusion and use of scientific and scholarly knowledge in society and the economy, broadly defined
International experience	International experience: work experience, networks, language skills, and connections to a specific country
Administration skills	Administrative and leadership experience
Personality and personal background	Social skills, personality, gender, and ethnicity
Grants	Records of and experience with attracting funding and grants

Interviews

I interviewed all types of actors involved in professor recruitment: department heads, administrative and elected leaders, and HR and administrative staff, and asked department heads for the names of internal and external experts who had recently served in evaluation committees. Altogether, 26 interviews were conducted with 10 males and 16 females; 10 were in the natural and 16 in the social sciences. Sixteen came from University A, six from University B, and four from University C. Among the respondents, 10 were academics serving on recruitment committees (including two external experts), six were department heads, seven were administrative heads, and three were HR staff. The interviews were conducted between August and December 2020 in person and via Microsoft Teams. The interviews lasted from 20 to 80 minutes and were audio-recorded, transcribed, and analyzed using the NVivo software package. The transcribed text totaled 133,108 words, with a range of 2,168 to 9,716 words per interview (see Appendix Tables 2 and 3 for more information about the interviews).

Appendix Table 3. Interview data: Respondent discipline and university affiliation.

	Number of interviews
Natural sciences	10
Social sciences	16
University A	16
University B	6
University C	4

Appendix Table 4. Interview data: Respondent position.

	Number of respondents
Internal committee member	8
External committee member	2
Head of Department	6
Head of Administration	7
HR–Administration	3

Survey

The survey was conducted under the auspices of my international research project, R-Quest. Invitations to the survey were sent to researchers in three disciplines (economics, physics, and cardiology) in five countries (Norway, Sweden, Denmark, the Netherlands, and the United Kingdom) at the most relevant research institutions (as defined by WoS data and journal classifications). We searched these institutions' websites for relevant organizational units to include in the survey and generated lists of personnel in relevant scientific and academic positions, including staff members, postdocs, and researchers but excluding Ph.D. students, adjunct appointments, guest researchers, and administrative and technical personnel. Some departments also had research groups in disciplines other than the one selected; we excluded the personnel in irrelevant groups. We then added people at the selected institutions with a minimum number of WoS publications in the field, regardless of the department or unit with which they were affiliated. In this way, we combined two sampling strategies to obtain a comprehensive sample. Based on the organizational units, we identified the full scope of researchers within a department or division. Using WoS categories, we identified those who publish in the field. This sample included respondents from eight Norwegian universities, seven Danish universities, eleven Swedish universities, eight U.K. Universities and ten Dutch universities.

In total, 10,442 respondents were invited to the survey. Among these, (a) 1,418 lacked functioning email addresses or were no longer researchers, (b) 30 were unwilling to participate, and (c) 36 were filtered out by the first survey question because they were outside the target group. Groups (a) and (c) were excluded before we calculated response rates. The updated sample thus included 9,060 researchers, of whom 6,443 did not answer, 218 partially answered, and 2,369 completed the questionnaire.

Response rates differed between the fields in a country, from 11.4 to 57.3 percent (see Appendix Table 4). Due to the low response rate, especially in the United Kingdom, we conducted a response analysis with Gender-api.com, which identified the gender of 92.4% of the invited respondents by their first names. The analysis showed that women were slightly more inclined to answer our survey.

Appendix Table 5. Respondents and response rate by field and country.

Country	Economics	Physics	Cardiology
Sweden	27.8	42.5	27.8
Norway	57.3	49.9	47.0
United Kingdom	11.4	12.8	(not invited)
Netherlands	20.0	24.2	12.8
Denmark	32.2	33.1	(not invited)

In paper 3, we analyzed the answers from researchers in Sweden, Norway, and the Netherlands, as there were no cardiologists in two of our countries (Denmark and the United Kingdom) due to difficulties in identifying these respondents, who often had affiliations with both hospitals and universities. In addition, we excluded respondents from research institutions, focusing only on university researchers. This sample included 1621 responses.

Appendix Table 6. Number of respondents by field and country in Paper 3.

Country	Cardiology	Economics	Physics	Total
Sweden	124	122	542	788
Norway	152	105	170	427
Netherlands	65	133	208	406
Total	341	360	920	1621

In paper 4 we studied answers from economists and physicists at universities in the five countries but only included respondents who answered that they had participated in recruitment processes. This sample included 848 responses.

Appendix Table 7. Number of respondents and response rate by field and country in Paper 4.

Country	Economics	Physics	Total
Sweden	57	242	299
Norway	60	82	142
United Kingdom	32	62	94
Netherlands	66	120	186
Denmark	44	83	127
Total	259	589	848

Questions included in paper 3

The best research in your field

[Q12] Think about the research you consider to be the best in your specific field/speciality.* Why do you consider this the best research? You may select more than one option.

**If you do research in multiple fields, select the one you can most easily answer for.*

- (1) Has answered/solved key questions/challenges in the field
- (2) Has changed the way research is done in the field (e.g. methodological breakthrough)
- (3) Has enabled researchers in the field to produce more reliable or precise research results
- (4) Has changed the key theoretical framework of the field
- (5) Was published in a journal with a high impact factor
- (6) Has attracted many citations
- (7) Has been a centre of discussion in the research field
- (8) Has drawn much attention in the larger society
- (9) Has benefited society (e.g. in terms of application in industry, new clinical practices, informed public policy)
- (10) Is what all students/prospective researchers need to read
- (11) Other, please specify: _____
- (12) Cannot say

Review for funding agencies

[Q17] Thinking of the last time you reviewed grant proposals, please indicate the kind of funding scheme/call you reviewed for:

- (1) Strategic/targeted research (applications to calls addressing thematic priorities)
- (2) Independent researcher-initiated research (applications to calls open to all research questions)
- (4) Do not remember/cannot answer
- (5) Other, please specify _____

[Q18] Please indicate the kind of grant you reviewed

- (1) Research project
- (2) Fellowship
- (3) Large grant/centre
- (5) Do not remember/cannot answer
- (6) Other, please specify _____

Thinking of what you found to be the best proposal, please indicate what was important for your assessment.

You may answer more generally if you find it difficult to remember one particular proposal.

	Not important	Somewhat important	Highly important	Do not remember/cannot answer
Project description: research question/problem selection	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Project description: methods/research plan	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Track record of the research team: number of publications/productivity	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Track record of the research team: citation impact of past publications	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Track record of the research team: important prior contributions in the relevant research field (assessed independently of citation scores and source of publication)	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Track record of the research team: experience with risk-taking research	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
The research environment: resources and facilities for performing the proposed research	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Communication/dissemination plan for scientific publications	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>

	Not important	Somewhat important	Highly important	Do not remember/cannot answer
Communication/dissemination plan addressing user groups outside academia	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>

Other, please specify: (1) (2) (3) (4)

Please enter a few keywords on the main positive characteristics of what you found to be the best proposal (e.g. what made the project description or the team the best):

Positions in your research group/department

[Q20] Thinking of the last time you assessed candidates for research/academic positions. Please indicate the kind of candidates you assessed:

- (1) Junior/early career
- (2) Senior/tenure
- (5) Other, please specify _____
- (4) Do not remember/cannot answer

Questions analyzed in both paper 3 and paper 4

[Q21] Thinking of the candidate you found to be the best, please indicate the aspects that were important for your assessments.

You may answer more generally if you find it difficult to remember one particular candidate.

	Not important	Somewhat important	Highly important	Do not remember/cannot answer
Matching field/expertise to the needs of the group/unit/project	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Research achievements: number of publications/productivity	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Research achievements: citation impact of past publications	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Research achievements: important prior research contributions (assessed independently of citation scores and source of publication)	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Teaching experience/achievements (including supervision of students)	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Experience/achievements from work outside science, e.g. professional/clinical practice, industry or public administration	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Experience in interacting with the public/users/industry	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Standing of the unit/group where the candidate is/has been working/trained	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
The potential for future achievements	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Communication and language skills	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Ability to compete for research grants	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Ensure diversity in the group/department (e.g. gender, ethnicity, age)	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
General impression from interview with candidate	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>
Other, please specify:	(1) <input type="checkbox"/>	(2) <input type="checkbox"/>	(3) <input type="checkbox"/>	(4) <input type="checkbox"/>

[Q22] Below are the aspects for which you selected "highly emphasised" in your previous answer. Please indicate which one of these was the most important aspect in your assessment.

- (1) Matching field/expertise to the needs of the group/unit/project
- (2) Research achievements: number of publications/productivity
- (3) Research achievements: citation impact of past publications
- (4) Research achievements: important prior research contributions (assessed independently of citation scores and source of publication)
- (5) Teaching experience/achievements (including supervision of students)
- (6) Experience/achievements from work outside science, e.g. professional/clinical practice, industry or public administration
- (7) Experience in interacting with the public/users/industry
- (8) Standing of the unit/group where the candidate is/has been working/trained
- (9) The potential for future achievements
- (10) Communication and language skills
- (11) Ability to compete for research grants
- (13) Ensure diversity in the group/department (e.g. gender, ethnicity, age)
- (14) General impression from interview with candidate
- (12) text entered/specified

Please enter a few keywords on the main positive characteristics of what you found the best candidate/the last candidate you recommended for a position (e.g. what made his/her past achievements or competences the best):
