SpringerBriefs in Political Science

Gunnar Sivertsen · Liv Langfeldt *Editors*

Challenges in Research Policy

Evidence-Based Policy Briefs with Recommendations

Springer

SpringerBriefs in Political Science

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic. Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

SpringerBriefs in Political Science showcase emerging theory, empirical research, and practical application in political science, policy studies, political economy, public administration, political philosophy, international relations, and related fields, from a global author community.
SpringerBriefs are characterized by fast, global electronic dissemination, standard publishing contracts, standardized manuscript preparation and formatting guidelines, and expedited production schedules.

Gunnar Sivertsen • Liv Langfeldt
Editors

Challenges in Research Policy

Evidence-Based Policy Briefs with Recommendations

Springer

Editors
Gunnar Sivertsen
Nordic Institute for Studies in Innovation,
Research and Education (NIFU)
Oslo, Norway

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

ISSN 2191-5466 ISSN 2191-5474 (electronic)
SpringerBriefs in Political Science
ISBN 978-3-031-69579-7 ISBN 978-3-031-69580-3 (eBook)
https://doi.org/10.1007/978-3-031-69580-3

Nordic Institute for Studies in Innovation, Research and Education This work was supported by the Nordic Institute for Studies in Innovation, Research and Education.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

About the Editors and Contributors

About the Editors

Gunnar Sivertsen is Research Professor at the Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway. Sivertsen's research contributes to science-based innovation in the development of research evaluation, research funding, and the use of indicators. He has advised the development of the research evaluation and funding systems in several countries.

Liv Langfeldt is Research Professor at the Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway. She has extensive experience in the study and evaluation of peer-review procedures. Langfeldt has been in charge of evaluations of funding schemes and their review process in several countries.

Contributors

Kaare Aagaard Aarhus University, Aarhus, Denmark

Dag W. Aksnes Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway

Siri Brorstad Borlaug Oslo Business School, Oslo Metropolitan University, Oslo, Norway

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

André Brasil Centre for Science and Technology Studies (CWTS) at Leiden University, Leiden, Netherlands

Christina Vogsted Drange Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway

Jens Jungblut Department of Political Science at University of Oslo, Oslo, Norway

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

Ingeborg Meijer Centre for Science and Technology Studies (CWTS) at Leiden University, Leiden, Netherlands

Ingvild Reymert Oslo Business School, Oslo Metropolitan University, Oslo, Norway

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

Alex Rushforth Centre for Science and Technology Studies (CWTS) at Leiden University, Leiden, Netherlands

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

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

Silje Marie Svartefoss TIK Centre for Technology, Innovation and Culture, at University of Oslo, Oslo, Norway

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

Thed van Leeuwen Centre for Science and Technology Studies (CWTS) at Leiden University, Leiden, Netherlands

Chapter 1
Introduction

Gunnar Sivertsen and Liv Langfeldt

Research has become a large and integrated part of the public sector with considerable investments, complex organizations, advanced infrastructures, and a high degree of collaboration and mobility across countries. Public research needs to be governed, organized, funded, assessed, and prioritized responsibly to ensure the best outcomes for society from well-organized scientific progress. This is challenging because research policy is dynamic by interacting with other policies.

Thirty years ago, research policy was mainly about the efficiency of the research system and its contribution to innovation and economic growth. Then came an era where excellence seemed to be a common denominator for disciplinary standards which could include all areas of research and build a bridge over to societal expectations. But excellence was also experienced as a strong instrument of prioritization: We will only fund the best, and the funded will get more. More recently, the focus has changed to ensuring responsibly assessed good research cultures that contribute to solving grand societal challenges while practising open research in interaction with society. It is already evident, however, that the idea of open research can be challenged by security policy.

While research policy may be influenced by changing values and concerns, each generation of actors need to understand and solve the core challenges of how to govern, organize, fund, assess, and prioritize research responsibly. This book is mainly about such stable challenges in research policy, but it also discusses temporary issues.

The book has two main parts. The first part (Chaps. 2–8) is about *research quality and evaluation*, while the second part (Chaps. 9–13) is about *research funding and governance*. Each chapter has the format of a user-friendly *policy brief* to support our aims to provide new actors in research policy with understanding of the challenges and to provide recommendations on how to approach them. All chapters

G. Sivertsen · L. Langfeldt (✉)
Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway
e-mail: gunnar.sivertsen@nifu.no; liv.langfeldt@nifu.no

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_1

are based on evidence from research, and references to further reading are given. The authors partly build on their own research and partly on an updated overview of the state of the art in research on the specific topic.

All authors have collaborated within the international *Centre for Research Quality and Policy Impact Studies (R-QUEST)*, which constituted an 8-year research commitment that was aimed at policy advice and funded by the Research Council of Norway (RCN grant number 256223). Another source of funding is the Norwegian Ministry of Education and Research, which provided the basis for Chap. 10 with a framework contract for NIFU to contribute to the knowledge base for a forthcoming white paper on the Norwegian research system.

The chapters are based on policy briefs written from 2016 to 2024 and so reflect different policy concerns and perspectives. The policy briefs for Chaps. 2, 4, and 9 were written before 2020 and address concerns for high-quality research, concentration of funding, and impact. Chapters 5 and 7 are based on more recent R-QUEST policy briefs (2021–22), while Chaps. 3, 6, 8, 10, 11, 12, and 13 are all from 2024. These most recent ones reflect current policy concerns such as interaction with society, open access, and responsibility and diversity in research assessments. All contributions have been edited to enhance readability and to serve the purpose of the book.

Part I
Notions of Research Quality and Evaluation of Research

Chapter 2
Identifying and Facilitating High-Quality Research

Liv Langfeldt, Kaare Aagaard, Siri Brorstad Borlaug, and Gunnar Sivertsen

Abstract Policies attempting to promote high-quality research are widespread. But what does it mean to promote high-quality research? And do these policies work? This chapter summarizes what we know and do not know about these issues.

2.1 The Politics of Research Quality

The ambition to promote high-quality research has been a prominent research policy feature for a long time, especially since the 1990s. The quest for quality will probably continue although recent trends in research assessment and funding focus on other dimensions as well, such as equity, diversity, and inclusion (EDI), societal challenges as expressed in UN's Sustainable Developments Goals, and more emphasis on the societal responsibility of research performing institutions, their research culture, and reform of research assessment in general. However, whenever there is competition for positions or resources, quality will most probably be a key concern for policymakers, funders, and peers and remain on top of the agenda—even when other dimensions also require attention.

At the level of countries, there is a widespread idea that world leading research groups performing ground-breaking research are needed to solve the challenges that societies confront. Governments have the obligation to ensure that public money on R&D is spent wisely. The most obvious choice is to prioritize the most successful

L. Langfeldt (✉) · G. Sivertsen
Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway
e-mail: liv.langfeldt@nifu.no; gunnar.sivertsen@nifu.no

K. Aagaard
Aarhus University, Aarhus, Denmark
e-mail: kaag@au.dk

S. B. Borlaug
Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway

Oslo Business School, Oslo Metropolitan University, Oslo, Norway
e-mail: siri.borlaug@oslomet.no

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_2

scientists and the most promising projects. There might also be a wish to maintain or improve the national scientific standing and status. *In sum, whatever the aim of research policy is, high-quality research has been presented as the solution and probably will continue to serve as an argument for prioritizing research expenditures.*

However, developing and implementing policies for research quality is by no means straightforward. While selecting the best, a country also needs to build up competences in new fields to solve societal challenges. Thus, there is a need for providing good general conditions for research to secure a broad knowledge base across a variety of fields and topics. It is not possible to predict all the competences and the knowledge needed in the future. Hence, diversity and excellence must be seen as complementary rather than contradictory considerations when allocating resources for research (Lamont and da Silva 2009).

Still, in general, *the public can best be convinced that a research policy is successful if the funding agencies and the authorities can document that they help to foster and attract world leading research groups.* The concept of high-quality research is appealing—and persuading—in terms of solving grand challenges and ensuring value for public resources spent on research. Not to be neglected is also the public interest in contributing to national competitiveness and pride, e.g. winning Nobel prizes, having highest-ranking universities, and achieving brain gain in general.

2.2 Different Aspects and Perceptions of Research Quality

Summarizing scholarly and empirical studies of research quality within science, we find three basic aspects of the concept (Gulbrandsen 2000; Gulbrandsen and Langfeldt 1997; Lamont 2009; Langfeldt et al. 2020; Polanyi 1962):

1. Plausibility/solidity, methodological soundness (and feasibility)
2. Originality/novelty
3. (a) Scientific and (b) societal value/significance

Each of these aspects may be specified and emphasized in different ways in different fields of research and in different evaluation contexts. Reviewing grant proposals is different from assessing candidates for professorships or reviewing manuscripts for publishing. On a more general level, they derive from the definition of research. To qualify as scholarly research, the work (1) should be well-founded in scientific methods, (2) provide new knowledge, and (3a) be relevant to the research community and/or (3b) society. Some of the common concepts of research quality combine two or more of these aspects, such as "frontier research" which is a combination of 2 and 3a in terms of generating valuable new knowledge at the frontier of science. And then, it also needs to be plausible and hold adequate standards of ethics, soundness, and integrity (1) to be valuable. This first aspect has gained more attention in recent years because of the increasing attention to scientific fraud based on new technologies and as a response to perverse incentives.

Even if clear and comprehensible at this basic conceptual level, "research quality" is contested and elusive. While there is general consensus that good research is solid, original, ethical, and significant, there is less consensus about what this means or how to identify good research. What is perceived as the most solid and significant contributions to a specific research field may vary between peers. Furthermore, numerous studies have pointed out biases in peer review, for instance that interdisciplinary and unconventional research is disfavoured (Ayoubi and Pezzoni 2021; Chubin and Hackett 1990; Lamont 2009; Langfeldt 2006; Laudel 2006; Luukkonen 2012). The outcome of peer review may even depend on the way the review is organized (Langfeldt 2001).

2.3 Identifying High Quality

Then, what do public authorities do to identify and facilitate high-quality research? And how can they document that they succeed with this? *Even with its many limitations and potential biases, peer review is often the best—and only—option when it comes to identifying high-quality research.* Peer review is thus widely used for allocating project grants, for evaluating the outcome of programmes and policy initiatives, and even for performance-based institutional funding in a few countries.

In some contexts, peer review is supplemented—or even replaced—by bibliometrics and other performance indicators. Such indicators are generally based on the (aggregated) outcome of peer review of papers submitted for publication, on the number of citations to published work, and/or on the outcome of review of grant applications. Aggregated to the organisational level, they form the foundation of indicator-based systems for performance-based funding (see Chap. 11) and are seen as indicators of policy success (e.g. by comparing countries or institutions, or the outcome of funding schemes). However, being based on the aggregated outcome of peer review, these bibliometric indicators also risk reproducing the biases in peer review (e.g. discriminating interdisciplinary and original research). Moreover, indicators based on citations primarily reflect scientific impact, which is only one of several aspects of research quality (see Chap. 6; Aksnes et al. 2019).

In addition, *quantitative indicators come with the risk of producing dysfunctional incentives*. If a researcher's future funding is based on her quantifiable output, she might easily give priority to quantity over quality in her research. There are indications that performance metrics in academia may influence the research activities, so that researchers "think with quantitative indicators" when they plan projects, and give less attention to, e.g., originality, long-term scientific contributions, or societal relevance (Müller and de Rijcke 2017). As stated in the Leiden Manifesto for research metrics: with metrics "We risk damaging the system with the very tools designed to improve it" when used by "organisations without knowledge of, or advice on, good practice and interpretation" (Hicks et al. 2015). The first principle of the Leiden Manifesto is thus that quantitative evaluations should support, not substitute, expert assessments, while the European Agreement on Reforming

Research Assessment (2022) advises criteria that recognize the diversity of research activities and practices (see Chap. 7).

In sum, metrics can inform and sometimes contradict, but seldom overcome the limitations, biases, and indecisiveness of peer review, and there are additional limitations and biases attached to them. In combination with expert advice/direct peer review, they may however still contribute to the identification of high-quality research and researchers (see Chap. 8). Metrics have important benefits as they demand far fewer resources than peer review, may challenge and inform peer review, and trigger thorough expert panel discussions. On the other hand, there is also the risk that metrics may misguide peer review or lead to less thorough panel discussions. Moreover, it should be kept in mind that the concept of research quality is multidimensional, that its operationalization is often contested, and that scholarly research is dynamic by nature. This implies that a fixed "agreement" on what is the most solid and significant research may be counterproductive in the long run—even if policy makers may perceive such a need. In the research community, diversity and open discussions are more important than consensus.

2.4 What Facilitates High Quality?

When there is limited consensus on how to define and identify high-quality research, how do responsible authorities and funders know how to promote it, and how do they know whether their policies for doing this are successful? There is limited knowledge about how governance arrangements affect research (Gläser and Laudel 2016). Key success factors may vary between fields of research (Laudel and Gläser 2014; Tirado and Nedeva 2023). Moreover, there may be different success factors behind the different aspects of quality and types of research (see Chap. 3). There is still a large body of literature trying to establish a connection between policy and research performance/quality in general. For example, many studies try to link differences between countries' performance on bibliometric indicators to differences in research policy. Others have studied researchers' perceptions of what promotes high quality.

In sum, the studies point to a high degree of complexity in the relation between research policy and research performance. The factors influencing performance are connected in multi-level systems with complex paths from changes in input factors at a macro-level to changes in individual and group level behaviour which eventually constitute the basis of the developments in national publication performance. In addition, relations may often be non-linear, meaning that both too much and too little of a certain factor may have negative effects on research performance. Moreover, *the high performing part of a research system can be rather independent of changes in general frame conditions* due to better access to external funding, a higher degree of autonomy, focus on beneficial publication behaviours, and benefits from existing cumulative advantages (Aagaard and Schneider 2016).

Studies that attempt to identify the factors which explain why certain countries regularly outperform others in terms of publications and citations have put much emphasis on the effects of changes in funding, as funding is one of the main channels by which authority is exercised over research (Edquist 2003; Whitley et al. 2010). National level *studies find no straightforward connection between financial incentives and the efficiency of university systems in terms of research performance* (Auranen and Nieminen 2010; see also Chap. 10). Moreover, turning to the impacts of specific instruments, studies indicate that research funding instruments have limited impact on research performance as measured in citation impact, but may impact productivity in terms of number of publications (Jacob and Lefgren 2011; Langfeldt et al. 2015; Sandström 2009) and increased career success (Bloch et al. 2014).

On the other hand, several studies appear to *agree on the importance of relative funding stability over longer time periods* (Heinze 2008; Hollingsworth 2008; Öquist and Benner 2012). Some argue that the combination of widespread autonomy and a competitive environment creates good performance (Aghion et al. 2010) and stimulates scientific innovation (Whitley 2003), while some find that competitive project funding can increase career uncertainty and anxiety and make it harder to establish the research group as a community (Franssen and de Rijcke 2019).

Another strand of studies has investigated organizational conditions for academic performance and creativity. The majority of these concern research groups and emphasize factors such as autonomy and flexibility in the interaction with colleagues; scholarly diversity; a balance between basic and applied research; small/moderate research group size; access to extramural skills and resources; and facilitating leadership and good collaboration with department and university management (Gulbrandsen 2000; Carayol and Matt 2006; Heinze et al. 2009;; Hollingsworth 2008; Pelz and Andrews 1966; Salter and Martin 2001). However, recently there has been an increased interest in the role of the research organizations (universities, university hospitals, public research institutes) and how they may contribute to high-quality research (see Chaps. 7 and 11). In sum, studies at researcher level indicate that *policy has an impact on the organization of research and researchers' practices, but not necessarily on their performance.* In national level studies, research policy is seen as a main foundation for research performance, but *it has not been possible to establish a causal link between research policy and research performance***.** We know and understand much about policy, research organizations, and scientific practices, but very little about the dependencies between these.

Policy Implications

- *Targeted policy:* As a basis for policy-making, there is a need for insight into the vital conditions for the various aspects of research quality (plausibility/integrity, originality, scientific and societal significance) and how they vary between research contexts. Vital conditions for solid laboratory science may be quite different from the conditions for solid research in the humanities, and very different from the conditions for originality or soci-

etal significance in these fields. Hence, policy instruments need to be targeted, which implies:

  - Defining the aspects of research quality to be facilitated (e.g. originality/ground-breaking science)
  - Identifying the conditions for high performance on these aspects of research quality in the relevant research fields

In developing such targeted policies, there is a need for close collaboration with the research community and in-depth expertise.

- *Open policy:* There are significant limitations when it comes to identifying future scientific success. Whereas scholarly diversity is seen as important for high quality in research, both peer review and science metrics tend to be conservative and may disfavour, e.g., interdisciplinary and less conventional research. To ensure a fertile and diverse research landscape, there is a need for policy measures providing general good conditions for research and to monitor adverse effects. In this, different forms of maps and visualisations (rather than one-dimensional rankings) may contribute to establish a general overview of research activities and support policy discussions.

References for Further Reading

Aagaard, K. and J. W. Schneider (2016). Research funding and national academic performance: Examination of a Danish success story. *Science and Public Policy*, 43(4): 518–531. http://spp.oxfordjournals.org/content/early/2015/10/01/scipol.scv058.abstract

Aghion, P., M. Dewatripont, C. Hoxby, A. Mas-Colell, and A. Sapir (2010). The governance and performance of universities: evidence from Europe and the US. *Economic Policy,* 25(61), 7-59.

Agreement on Reforming Research Assessment (2022). https://coara.eu/agreement/the-agreement-full-text

Aksnes, Dag W., Liv Langfeldt, and Paul Wouters (2019). Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *SAGE Open* 9(1): 1–17.

Auranen, O. and M. Nieminen (2010). University research funding and publication performance - an international comparison. *Research Policy,* 39, 822-834.

Ayoubi, C. M. and Pezzoni, F. Visentin (2021). Does It Pay to Do Novel Science? The Selectivity Patterns in Science Funding, *Science and Public Policy*, 48(5): 635–648

Borlaug, S.B. and S.M. Svartefoss (2024). Evaluating transdisciplinary research quality. In G. Sivertsen and L. Langfeldt (eds). *Challenges in Research Policy*. Springer. [Chapter 3]

Bloch, C., E. K. Graversen and H. S. Pedersen (2014). Competitive Research Grants and Their Impact on Career Performance. Minerva 52 (1): 77–96

Carayol, N. and M. Matt (2006). Individual and collective determinants of academic scientists' productivity. Information economics and policy, 18: 55–72.

Chubin, D. E. and E. J. Hackett (1990). *Peerless Science. Peer Review and U.S. Science Policy*. New York, State University of New York Press.

Edquist, O. (2003). Layered science and science policies. Minerva 41: pp. 207–221.

Franssen, T. and S. de Rijcke (2019). The rise of project funding and its effects on the social structure of academia. In F. Cannizzo and N. Osbaldiston (Eds.) *The social structures of global academia* (Chapter 9). London: Routledge.

Gläser, J., and Laudel, G. (2016). Governing science: How science policy shapes research content. *European Journal of sociology/Archives Européennes de sociologie*, *57*(1), 117-168.

Gulbrandsen, M. and L. Langfeldt (1997). *Hva er forskningskvalitet? En intervjustudie blant* norske forskere. Oslo: NIFU-rapport 9/97.

Gulbrandsen, J. M. (2000). *Research quality and organisational factors: An investigation of the relationship*. Trondheim, Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology.

Heinze, T. (2008). How to sponsor ground-breaking research: a comparison of funding schemes. *Science and Public Policy*, 35(5), 302–318.

Heinze, T., P. Shapira, J. D. Rogers and J. M. Senker. 2009. Organizational and Institutional Influences on Creativity in Scientific Research. *Research Policy* 38: 610-623.

Hicks D., P. Wouters, L. Waltman, S. de Rijcke, I. Rafols (2015). The Leiden Manifesto for research metrics. *Nature* **520**, 429–431. https://doi.org/10.1038/520429a

Hollingsworth, J. R. (2008) 'Scientific Discoveries: An Institutionalist and Path-Dependent Perspective'. In: Hannaway, C. [ed.]; Biomedicine in the Twentieth Century: Practices, Policies, and Politics. Bethesda. pp. 317-353.

Jacob, B. A. and L. Lefgren (2011). The impact of NIH postdoctoral training grants on scientific productivity, *Research Policy*, 40/6, 864–74.

Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Cambridge, MA: Harvard University Press.

Lamont, M. and G. M. da Silva (2009). Complementary rather than contradictory: diversity and excellence in peer review and admissions in American higher education. *21st Century Society: Journal of the Academy of Social Science* 4(1):1-15.

Langfeldt, L., M. Nedeva, S. Sörlin, D. Thomas (2020). Co-existing Notions of Research Quality: A Framework to Study Context-specific Understandings of Good Research. *Minerva* 58(1): 115–137.

Langfeldt, L (2006). The policy challenges of peer review: Managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation* 15(1):31-41.

Langfeldt, L. (2001). The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome. *Social Studies of Science* 31(6):820-841.

Langfeldt, L, C. Bloch, G. Sivertsen (2015). Options and limitations in measuring the impact of research grants – evidence from Denmark and Norway. *Research Evaluation* 24 (3): 256-270. http://rev.oxfordjournals.org/content/24/3/256.abstract

Laudel, G. (2006). Conclave in the Tower of Babel: how peers review interdisciplinary research proposals. *Research Evaluation,* 15(1):57-68.

Laudel, G.and J. Gläser (2014). Beyond breakthrough research: Epistemic properties of research and their consequences for research funding. Research Policy 43, 1204-1216.

Luukkonen, T. (2012). Conservatism and risk-taking in peer review: Emerging ERC practices. *Research Evaluation,* 21(1):48-60.

Müller, R. and S. de Rijcke (2017). Thinking with indicators. Exploring the Epistemic Impacts of Academic Performance Indicators in the Life Sciences. *Research Evaluation* 26(3):157–168.

Pelz, D.C. and F.M. Andrews (1966). *Scientists in Organizations. Productive Climates for Research and Development.* John Wiley and Sons, New York.

Polanyi, M. (1962). The republic of science: Its Political and Economic Theory. *Minerva* 1:54-73.

Salter, A.J. and B. R. Martin (2001). The economic benefits of publicly funded basic research: a critical review. *Research Policy* 30: 509-532.

Sandström, U. (2009). Research quality and diversity of funding: A model for relating research money to output of research. *Scientometrics* 79/2, 341–9.

Tirado, MM, Nedeva, M. & Thomas, D.A. (2023). Aggregate level research governance effects on particle physics: A comparative analysis. Research Evaluation. https://doi.org/10.1093/reseval/rvad025

Öquist G. and M. Benner (2012). Fostering breakthrough research: a comparative study. Stockholm: Kungliga Vitenskabsakademien. https://www.forskasverige.se/wp-content/uploads/Fostering-Breakthrough-Research.pdf

Whitley, R., J. Gläser and L. Engwall (Eds.) (2010). *Reconfiguring Knowledge Production: Changing Authority Relations in the Sciences and Their Consequences for Intellectual Innovation.* Oxford: Oxford University 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), 1015-1029.

Chapter 3
Evaluating Transdisciplinary Research Quality

Siri B. Borlaug and Silje Marie Svartefoss

Abstract Transdisciplinary research (TDR) addresses societal issues and involves non-academic partners. To meet today's societal challenges TDR is often seen as part of the solution. But TDR projects differ considerably from regular and disciplinary projects and therefore require other conceptions of quality and evaluative approaches.

3.1 What Is Transdisciplinarity?

Transdisciplinarity is a slippery concept with several definitions. One of the most common is that TDR is research which crosses disciplinary boundaries and involves non-academic partners. The research is context specific and often problem-driven (Klein 2008; Pohl et al. 2011). To give an example: commissioned research is often transdisciplinary as the research addresses the need of the commissioner; the commissioner is involved in setting the agenda and defining the research question. Still, for commissioned research to be truly transdisciplinary active involvement of the non-academic partners is required throughout the research process. TDR thus differs from so-called collaborative projects in which the external partners are not obliged to participate in all stages of the research project.

Further, TDR differs from disciplinary, multidisciplinary, and interdisciplinary research as it involves work to overcome social, cognitive, and organisational barriers between academic and non-academic partners (Belcher et al. 2016). In

S. B. Borlaug (✉)
Oslo Business School, Oslo Metropolitan University, Oslo, Norway

Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway
e-mail: siri.borlaug@oslomet.no

S. M. Svartefoss
Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway

TIK Centre for Technology, Innovation and Culture, University of Oslo, Oslo, Norway
e-mail: s.m.svartefoss@tik.uio.no

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_3

comparison, interdisciplinary research is defined as research that transcends disciplinary fields (Huutoniemi et al. 2010; Klein 2008).

Thus, TDR challenges common conceptions of quality and evaluation approaches and practices. Given the increasing emphasis on the importance of transdisciplinarity for solving societal challenges, and hence the increased requirements for involvement of non-academic partners in research project applications and grants, it is important for policymakers and funders to be aware of that TDR requires other conceptions and approaches than disciplinary and interdisciplinary research projects.

3.2 Research Quality in a Disciplinary Context

Current tendencies in research policy show a movement from quantitative and summative evaluations towards more formative evaluations of research and researchers (Sivertsen and Rushforth 2024). In the midst of this, notions of quality have, however, remained underdeveloped. For a long time, the emphasis in research policy has been on internationalisation and scientific impact, or "excellence", characterised by a focus on developing and supporting research at the international frontier. Within this, research quality has generally been perceived and operationalised as scientific impact, and bibliometric indicators have been used as the main means to measure research quality. There are many reasons for this, but access to quantifiable data is certainly a major one. The emphasis on scientific impact poses however challenges for TDR as this may not be the main aim of the research.

Furthermore, in the design of peer review of research project proposals, quality notions tend to be ill-specified or very general, asking reviewers to assess the "originality", "rigour", or "value". With the increased emphasis on TDR, a broader understanding of research quality is needed for ensuing proper evaluation processes. Especially given that former discussions of research quality have often been limited to differences in conceptions of quality between fields and within areas, like the humanities (Hug et al. 2013; Lamont 2009; Mårtensson et al. 2016), or different sites. For example, Langfeldt et al. (2020) argue that there are several co-existing notions of research quality in different sites such as knowledge communities, research organisations, funding agencies, and policy, and they identify three core attributes of quality: originality, plausibility, and value. The first refers to novelty and innovativeness—key attributes for research to become a legitimate contribution to the stock of knowledge. Plausibility or reliability refers to sound methods, rigor, integrity, and research ethics, and value refers to both scientific and societal value/usefulness.

While external usefulness is seen as an attribute of research quality, the authors do not really pay attention to the role of stakeholders outside science (except for the policy sphere) and their influence on research quality. In this perspective, relevance or usefulness outside of science is still a contested criterium of research quality. One consequence of this is the classical two-dimensional rhetorical and political divide

and perceived tension between "quality and relevance". In transdisciplinary research, however, relevance is embedded in the research process and a core criterium. We will expand on this in Sect. 3.4.

3.3 Evaluation in a Disciplinary Context

Current approaches to evaluation of research projects tend to further contribute to a strengthening of the divide between quality and relevance. Project proposals are usually reviewed by a panel of academic peers, and studies of panel dynamics have found that grant peer review seems to disfavour inter- and transdisciplinary proposals (Ayoubi et al. 2021; Langfeldt 2006). Moreover, traditional evaluation procedures like midterm review of a large project or centres, and after the project is concluded, are usually not well suited for transdisciplinary research. For instance, societal impact evaluations focus on the relevance of research after the ending of the project, creating an artificial separation between research and the impact phase, and do not do justice to the nature of the research process (Franssen 2022), which in itself is a transdisciplinary effort.

Furthermore, evaluations may also play a significant role in setting direction for what type of outcome of the projects that is valued (Borlaug 2016). An emphasis on scientific publications as a valued outcome may thus represent a hindrance for developing genuinely transdisciplinary projects.

In short, the increased emphasis on transdisciplinary research and its importance for sustainable research and role in solving societal challenges call for a broader operationalisation of research quality beyond the scientific conceptions, and an inclusion of relevance throughout the different evaluative phases of the research process.

3.4 Transdisciplinarity and Quality

Four key principles are seen as essential elements of TDR quality (Belcher et al. 2016, p. 8): relevance, credibility, legitimacy, and effectiveness. Each principle has a set of criteria, i.e., conditions that need to be met in order to achieve a principle.

- *Relevance* is "…the importance, significance, and usefulness of the research project's objectiveness, process and findings to the problem context and to society". TDR quality involves addressing societally relevant problems and produces useful knowledge and includes all phases of the research process from problem formulation through the applicability of the research.
- *Credibility* refers to the robustness of the research findings and the adequacy of data and methods. The inclusion of external actors helps to achieve relevance and

legitimacy but also heightens requirements related to credibility such as transparency, reflection, and reflexivity.

- *Legitimacy* is whether the research process is perceived as fair and ethical by the end users. On the one hand, this includes that researchers reflect and account for their own positions and interest and on the other hand to make the process transparent to stakeholders external to the research. A delineation of the inclusion and engagement of societal actors along the whole process is therefore important for the legitimacy of the research.
- *Effectiveness* is here understood as research that contributes to a positive change, may it be social, economic, and/or environmental. This principle is a bit tricky; it can be indicated or assessed at the proposal stage and during the research process but assessing effectiveness ex post still remains a challenge. Learning and societal capacity building are central goals of TDR and therefore may effectively relate to changes in knowledge, attitudes, skills, and/or practices, not only products and the like.

Notably, these principles differ considerably from the ones seen as constituting research quality in disciplinary research projects. In TDR the relevance aspect is up front, and although credibility and legitimacy are reminiscent of what Langfeldt et al. (2020) labelled plausibility, these aspects of quality are more emphasised in TDR than when evaluating disciplinary research.

The lack of generally accepted quality standards for TDR is one of the reasons why the proliferation of TDR has progressed moderately, some argue (Jahn and Keil 2015). There is as such a need for building up a practice of quality assurance and evaluation of TDR.

3.5 Transdisciplinarity and Evaluation

There are several approaches and frameworks for how to evaluate TDR research, and many of them are rather field or context specific. For the purpose of this chapter, we highlight principles that evaluations of TDR should incorporate and outline how some frameworks for evaluation specifically suggest that TDR quality should be evaluated.

Principles that Evaluations of TDR Should Incorporate

The literature points to four general principles. First, TDR evaluations need to adapt to account for the fact that TDR includes external actors with different perceptions and expectations about the results of TDR. As a general rule stakeholders should be included in the evaluations of TDR projects (Belcher et al. 2019c; Brennan and

Rondón-Sulbarán 2019; Klein 2006, 2008; Mitchell et al. 2015; Wickson and Carew 2014). This also includes representatives of non-human stakeholders (Franssen 2022).

Second, TDR should be evaluated in relation to the various stakeholders it was intended to be useful for (Belcher et al. 2019c; de Oliveira et al. 2019; Hansson and Polk 2018; Holzer et al. 2018; Kaufmann and Kasztler 2009; Roux et al. 2010). As an example, Kaufmann and Kasztler (2009) argue that it is not possible to define the quality of research output by only considering the output in itself. It needs to be defined in relation to the reception within different communities, both scientific and non-scientific.

Third, evaluations should take into account that the context of TDR varies; that is, there are differences in institutional conditions and type of partner/actor (Belcher et al. 2016; Belcher et al. 2019c; Belcher et al. 2020; Di Iacovo et al. 2016; Hansson and Polk 2018). This challenges the applicability of generic evaluation frameworks and calls for field and context-specific approaches.

Fourth, even though many argue that the impact of TDR should also be a part of a TDR evaluation (Belcher et al. 2016; Belcher et al. 2019c; Belcher et al. 2020; Hansson and Polk 2018; Janinovic et al. 2020; Roux et al. 2010), it is very challenging to do this in practice. To evaluate impact, a substantial amount of time must have passed from the end of the project to the start of the evaluation (Roux et al. 2010). Additionally, it is often quite challenging for external actors to connect a change in practice to a specific piece of knowledge; a change in practices is typically influenced by several sources (Belcher et al. 2020).

Frameworks for Evaluation of TDR and Quality

There are few studies which provide frameworks for how evaluations can be conducted.

In general, there is an emphasis on the initial phase of a TDR project. A common issue is the underlining of the need to define the ends and outcomes of the project early in order to structure different perceptions and expectations of the involved stakeholders (Belcher et al. 2019c; Brennan and Rondón-Sulbarán 2019; Mitchell et al. 2015; Wickson and Carew 2014). As engagement of all stakeholders is key to TDR quality, it is important that stakeholder engagement is a part of the assessment of TDR projects, both in the proposal phase and afterward (Franssen 2022).

One promising approach to evaluation builds on the theory of change (ToC) which specifies how change is expected to occur in a given context (Belcher et al. 2019b; Belcher et al. 2020). A ToC entails formulating testable hypotheses that outline how and why change is expected to occur as a result of the TDR project. These should relate to the project's long-term goals and may involve changes in knowledge, attitudes, skills, relationships, and behaviour. A ToC can be used for both planning and monitoring in addition to evaluation (Belcher et al. 2019a). The approach primarily understands TDR quality as researchers' ability to optimise the

research design and output to increase the potential for outcomes and impact. Hence, the lack of outcomes or impact of TDR does not imply low quality.

This chapter highlights that TDR requires other conceptions of quality than those often used in guidelines of agencies funding research. Furthermore, TDR requires emphasis on the initial phases of a research project and the evaluation of other outcomes than scientific impact. This implies that TDR may need to set aside more resources for evaluation and that evaluation may play a significant role in the outcome of the projects.

Policy Implications

- Funding of TDR needs other quality conceptions and evaluation approaches than disciplinary and even interdisciplinary research.
- There are four key principles for TDR quality: relevance, credibility, legitimacy, and efficiency. Although an important aspect of disciplinary research, the involvement of non-academic partners requires heightened consciousness on these aspects.
- It is important to invest considerable work and resources in the early phase of a TDR project, which is in the research design and the implementation of the projects, in order to achieve wanted outcome.
- When evaluating TDR, one needs to (1) take into account that project participants have different perceptions and expectations about the results of the project, (2) evaluate the project in relation to the various stakeholders it is intended to be useful for, (3) take into account that context, fields, and areas of the project vary, and (4) impact is inherent in TDR.
- A framework for evaluating TDR is theory of change (ToC). This emphasises the importance of continuously evaluating change in relation to a broad set of outcomes like knowledge, attitudes, skills, relationships, and behaviours.

References for Further Reading

Ayoubi, C., Pezzoni, M., & Visentin, F. (2021). Does It Pay to Do Novel Science? The Selectivity Patterns in Science Funding. Science and Public Policy, 48(5), 635–648. https://doi.org/10.1093/scipol/scab031

Belcher, B. M., Claus, R., Davel, R., Jones, S., & Ramirez, L. (2019a). Research Theory of Change: A Practical Tool for Planning and Evaluating Change-oriented Research. https://researcheffectiveness.ca/wp-content/uploads/sites/7/2019/08/Theory-ofChange-Toolkit.pdf

Belcher, B. M., Claus, R., Davel, R., & Ramirez, L. F. (2019b). Linking transdisciplinary research characteristics and quality to effectiveness: A comparative analysis of five research-for-development projects. Environmental Science & Policy, 101, 192–203. https://doi.org/10.1016/j.envsci.2019.08.013

Belcher, B. M., Davel, R., & Claus, R. (2020). A refined method for theory-based evaluation of the societal impacts of research. MethodsX, 7, 100788. https://doi.org/10.1016/j.mex.2020.100788

Belcher, B. M., Ramirez, L. F., Davel, R., & Claus, R. (2019c). A response to Hansson and Polk (2018) "Assessing the impact of transdisciplinary research: The usefulness of relevance, credibility, and legitimacy for understanding the link between process and impact". Research Evaluation, 28(2), 196–201. https://doi.org/10.1093/reseval/rvy037

Belcher, B. M., Rasmussen, K. E., Kemshaw, M. R., & Zornes, D. A. (2016). Defining and assessing research quality in a transdisciplinary context. Research Evaluation, 25(1), 1–17. https://doi.org/10.1093/reseval/rvv025

Borlaug, S. B. (2016). Moral hazard and adverse selection in research funding: Centres of excellence in Norway and Sweden. Science and Public Policy, 43(3), 352–362. https://doi.org/10.1093/scipol/scv048

Brennan, M., & Rondón-Sulbarán, J. (2019). Transdisciplinary research: Exploring impact, knowledge and quality in the early stages of a sustainable development project. World Development, 122, 481–491. https://doi.org/10.1016/j.worlddev.2019.06.001

de Oliveira, T. M., Amaral, L., & Pacheco, R. C. dos S. (2019). Multi/inter/transdisciplinary assessment: A systemic framework proposal to evaluate graduate courses and research teams. Research Evaluation, 28(1), 23–36. https://doi.org/10.1093/reseval/rvy013

Di Iacovo, F., Moruzzo, R., Rossignoli, C. M., & Scarpellini, P. (2016). Measuring the effects of transdisciplinary research: The case of a social farming project. Futures, 75, 24–35. https://doi.org/10.1016/j.futures.2015.10.009

Franssen, T. (2022). Enriching research quality: A proposition for stakeholder heterogeneity. Research Evaluation, 31(3), 311–320. https://doi.org/10.1093/reseval/rvac012

Hansson, S., & Polk, M. (2018). Assessing the impact of transdisciplinary research: The usefulness of relevance, credibility, and legitimacy for understanding the link between process and impact. Research Evaluation, 27(2), 132–144. https://doi.org/10.1093/reseval/rvy004

Holzer, J. M., Carmon, N., & Orenstein, D. E. (2018). A methodology for evaluating transdisciplinary research on coupled socio-ecological systems. Ecological Indicators, 85, 808–819. https://doi.org/10.1016/j.ecolind.2017.10.074

Hug, S. E., Ochsner, M., & Daniel, H.-D. (2013). Criteria for assessing research quality in the humanities: A Delphi study among scholars of English literature, German literature and art history. Research Evaluation, 22(5), 369–383. https://doi.org/10.1093/reseval/rvt008

Huutoniemi, K., Klein, J. T., Bruun, H., & Hukkinen, J. (2010). Analyzing interdisciplinarity: Typology and indicators. Research Policy, 39(1), 79–88. https://doi.org/10.1016/j.respol.2009.09.011

Jahn, T., & Keil, F. (2015). An actor-specific guideline for quality assurance in transdisciplinary research. Futures, 65, 195–208. https://doi.org/10.1016/j.futures.2014.10.015

Janinovic, J., Pekovic, S., Vučković, D., Đoković, R., & Peic, M. (2020). Innovative Strategies for Creating and Assessing Research Quality and Societal Impact in Social Sciences and Humanities. Interdisciplinary Description of Complex Systems, 18, 449–458. https://doi.org/10.7906/indecs.18.4.5

Kaufmann, A., & Kasztler, A. (2009). Differences in publication and dissemination practices between disciplinary and transdisciplinary science and the consequences for research evaluation. Science and Public Policy, 36(3), 215–227. https://doi.org/10.3152/030234209X427121

Klein, J. T. (2006). Afterword: The emergent literature on interdisciplinary and transdisciplinary research evaluation. Research Evaluation, 15(1), 75–80. https://doi.org/10.3152/147154406781776011

Klein, J. T. (2008). Evaluation of Interdisciplinary and Transdisciplinary Research. American Journal of Preventive Medicine, 35(2), S116–S123. https://doi.org/10.1016/j.amepre.2008.05.010

Lamont, M. (2009). How professors think: Inside the curious world of academic judgment. Harvard University Press.

Langfeldt, L. (2006). The policy challenges of peer review: Managing bias, conflict of interests and interdisciplinary assessments. Research Evaluation, 15(1), 31–41. https://doi.org/10.3152/147154406781776039

Langfeldt, L., Nedeva, M., Sörlin, S., & Thomas, D. A. (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

Mitchell, C., Cordell, D., & Fam, D. (2015). Beginning at the end: The outcome spaces framework to guide purposive transdisciplinary research. Futures, 65, 86–96. https://doi.org/10.1016/j.futures.2014.10.007

Mårtensson, P., Fors, U., Wallin, S.-B., Zander, U., & Nilsson, G. H. (2016). Evaluating research: A multidisciplinary approach to assessing research practice and quality. Research Policy, 45(3), 593–603. https://doi.org/10.1016/j.respol.2015.11.009

Pohl, C., Perrig-Chiello, P., Butz, B., Hadom, G. H., Joye, D., Lawrence, R., Nentwich, M., Paulsen, T., Rossini, M., Truffer, B., Wasti-Walter, D., Wiesmann, U., & Zinsstag, J. (2011). Questions to evaluate inter- and transdisciplinary research proposals [Working paper]. td-net for Transdisciplinary Research.

Roux, D. J., Stirzaker, R. J., Breen, C. M., Lefroy, E. C., & Cresswell, H. P. (2010). Framework for participative reflection on the accomplishment of transdisciplinary research programs. Environmental Science & Policy, 13(8), 733–741. https://doi.org/10.1016/j.envsci.2010.08.002

Sivertsen, G., & Rushforth, A. (2024). The ongoing reform of research assessment In G. Sivertsen and L. Langfeldt (eds). *Challenges in Research Policy*. Springer. [Chapter 7]

Wickson, F., & Carew, A. L. (2014). Quality criteria and indicators for responsible research and innovation: Learning from transdisciplinarity. Journal of Responsible Innovation, 1(3), 254–273. https://doi.org/10.1080/23299460.2014.963004

Chapter 4
Evaluating and Improving the Societal Impact of Research

Gunnar Sivertsen and Ingeborg Meijer

Abstract Societal impact of research does not occur primarily as unexpected, extraordinary incidents of particularly useful breakthroughs in science. Is it more often a result of normal everyday interactions between organisations that need to create, exchange, and make use of new knowledge to further their goals. This chapter discusses how to assess and improve the cocreation and use of research in normal research–society relations.

4.1 What Is Normal Societal Impact of Research?

Societal impact has gained a central focus in research policy and evaluation. Research is increasingly expected to meet societal challenges and to interact responsibly with society. National and international research funding organisations are asking for evidence or indicators of societal impact. Several frameworks for the understanding and evaluation of societal impact have been proposed and piloted (Bornmann 2013).

We (Sivertsen and Meijer 2019) define *normal societal impact* as the more-or-less active, productive, and responsible interactions between (units of) research organisations and other organisations according to their purposes and aims in society. Within the research organisations, such interactions will often occur informally at the individual researcher or research group level, but they may also follow formalised agreements or well-established traditions for collaboration. With a similar definition of societal impact, Spaapen and van Drooge (2011) "understand productive interactions as exchanges between researchers and stakeholders in which

G. Sivertsen (✉)
Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway
e-mail: gunnar.sivertsen@nifu.no

I. Meijer
Centre for Science and Technology Studies (CWTS) at Leiden University, Leiden, Netherlands
e-mail: i.meijer@cwts.leidenuniv.nl

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_4

knowledge is produced and valued that is both scientifically robust and socially relevant". While inspired by the concept of productive interactions, our definition of *normal impact* also includes cases in which the expected interaction is missing, impaired, or inadequate, or in which the outcome is neither scientifically robust nor socially relevant.

We will show that evaluation and policy designs can be improved and made more relevant and effective by using a distinction between normal and extraordinary impact and by separating between organisational and individual-level activities and responsibilities in science–society relations.

In contrast to normal impact, we define *extraordinary impact* as more rare incidences where traditional and typical or new and untypical interactions have unexpected widespread implications for society. In this definition, we include extraordinary cases of negative impact ("grimpact", Derrick et al. 2018), since such cases can be understood and evaluated—not as accidents—but as violations of the expected normal and responsible impact. A current method for evaluating the societal impact of research on the basis of evidence-based *case studies* tends to select individual incidents of particularly interesting or impressive impact. These incidents may be extraordinary in the sense that they have unusually wide implications or demonstrate impact in new relations where impact normally does not occur, e.g., in the relation between the humanities and the pharmaceutical industry.

Contrary to such extraordinary impact—which by definition is rare and often based upon serendipity—evaluation of normal impact implies a focus on the quality of everyday normal interactions between research and society in areas of research and sectors of society where such interaction can be expected.

4.2 The Linear Model for Understanding Societal Impact

In 2014, the Research Excellence Framework for the evaluation and funding of universities in the United Kingdom was the first to introduce a broad ex post assessment of societal impact of research (Derrick and Samuel 2017). It soon became the most studied and discussed approach to assess societal impact so far in the literature (Pedersen et al. 2018), and it has been adopted by other institutions and countries. The REF methodology, which was also used in the 2021 exercise and will be used again in 2028, requires evidence of societal impact related to specific achievements in research. There is a template for the written *case reports* (REF2014 2012) which among other things demands the identification and documentation of:

- The research that underpinned the impact: "This section should outline the key research insights or findings that underpinned the impact, and provide details of what research was undertaken, when, and by whom."
- The resulting impact: "A clear explanation of the process or means through which the research led to, underpinned or made a contribution to the impact (for example, how it was disseminated, how it came to influence users or beneficiaries, or how it came to be exploited, taken up or applied)."

The typical analysis of case studies based on the REF methodology has been to identify pathways, beneficiaries, and effects of research in the reported cases, with a clear stance on excellence, not only in science but also in societal impact. This model for collecting and evaluating reported cases of societal impact is implicitly based on an understanding of societal impact that reminds of the so-called *linear model* of innovation (Godin 2006) or communication (Shannon and Weaver 1949). It thereby has a basic problem with being at odds with most empirical studies of the science–society interactions in our time and what more theoretically has been called Mode 2 in the interactive dynamics between science and contemporary societies (Gibbons et al. 1994).

Moreover, the REF requirements to demonstrate evidence of societal impact are exposed to some general problems with linking research activities to societal impacts. These are problems with e.g.:

- Causality: relationships between research and innovation inputs, activities, outputs, and impacts are often unclear or non-linear.
- Attribution: it is difficult to separate the impact of research and innovation from other inputs and activities.
- Internationality: research and innovation activities, and value chains, are global and normally not identifiable in specific relations.
- Time scale: impacts in science–society relations are normally realised over a very long time and only extraordinarily in short time.

The REF is in the end about institutional funding. Inevitably, the REF methodology for evaluating societal impact is mostly focused on one side of the interaction. The case studies methodology also makes the universities report primarily examples of extraordinary impact, mostly at the individual level. This procedure has many valuable outcomes. It increases awareness of the societal responsibilities and provides strong stories to tell in the media. But the procedure does not result in an evaluation to learn from.

4.3 Alternative Frameworks and Methods

Our definition above of *normal impact* includes both sides of the interaction and is based on an *interactive model* for understanding societal impact. Evaluating normal impact implies asking—in specific and typical relations—how the interaction is functioning on a daily basis on both sides, according to organisational purposes and aims. Moreover, the problems with causality, attribution, internationality, and time scale are less important for the analysis. Other evidence about daily operations and their management and infrastructure will be in focus.

The Swedish Higher Education Authority is currently (2023–24) running a formative and mutual learning assessment on how each of the country's 37 higher education institutions are providing "the conditions for interactive cocreation of knowledge with society for mutual benefit". The Swedish word for "societal impact"

is "samverkan", which means both "collaboration" and "cocreation" and is without the connotation of the English word "impact", which implies that one thing influences another. The Swedish assessment exercise has its background in a recent change of the higher education law in 2021 where the former wording of the obligation to *serve society with knowledge* was replaced by *samverkan* with society for *mutual benefit.* This modernisation of Swedish law is clearly based on the interactive model for understanding societal impact.

Also more in line with what we mentioned above as the Mode 2 theory of the interactive dynamics between science and contemporary societies are several alternative frameworks for the understanding of the societal impact of research, such as the Payback framework (Levitt et al. 2010; Klautzer et al. 2011), the SIAMPI/ERiC model (Spaapen and van Drooge 2011; Molas-Gallart and Tang 2011; Olmos-Peñuela et al. 2014), the Flows of knowledge framework (Meagher et al. 2008), the Research Contribution Framework (Morton 2015), Contribution Mapping (Kok and Schuit 2012), and the IMPACT-EV (Flecha et al. 2014). Overviews of such frameworks are found in Greenhalgh et al. (2016), Pedersen et al. (2018), and Giménez-Toledo et al. (2023).

4.4 Normal Interactions with Society Differ Across Fields of Research

The missions of general universities towards society are usually expressed in very general terms. Less vaguely expressed are the aims and purposes of research organisations with a more specialised profile (e.g. agricultural universities or public health research institutes). Evaluations of normal impact will need this kind of specificity, as societal relations differ by fields and subfields of research.

This was clearly demonstrated by two recent evaluations of the humanities and social sciences (SSH) in Norway. Both included evidence-based case studies and evaluations of societal impact according to the REF methodology. A few of the cases from the *humanities* demonstrated extraordinary contributions to information technology, bioethics, peace processes, emergency communication, and genetic counselling. The commissioner of the evaluations, the Research Council of Norway, chose to highlight these extraordinary cases when reporting from the exercise. However, the SSH research more typically contributed to societal development, policy design, public administration, international affairs, societal integration and understanding of different languages and cultures, education at all levels, cultural life, media and information, and history, the "memory of society". The case studies demonstrated that research in the SSH is integrated in, and not operating at a distance from, certain domains in society where the disciplines may have specific purposes and play specific roles in specific societal and cultural contexts. Musicology usually contributes to musical life while research in international relations normally contributes to diplomacy and foreign policy.

These purposes and roles may often be more specific than seen in a general typology or description of pathways, beneficiaries, and effects. Examples of such generalisations may be "improving health and well-being" or "commercialisation and exploitation". At the same time, the specific aims of the research–society interactions may be more general than the individual case report can account for. Hence, a more specific typology of normal societal relations in each field of research is needed.

Law studies, for example, are concentrated in the universities' Faculty of Law in most countries. Their typical interaction with society is different from other faculties and at the same time more specific than a university's general societal responsibility: It serves the legal system of a country by educating professionals and responding to societal needs in the legal system. Moreover, studies in, e.g., EU Law (the research is international in focus and applications) or Criminal Law (the research is national in focus and directly concerned with the civil society) will have different relations to society. Such specific relations need to be understood before they are evaluated. Extraordinary cases of particularly impressing impact will not be sufficient for such an understanding.

Just as a Faculty of Law is part of a country's legal system, a Medical Faculty is part of a country's healthcare system. In this perspective, an extraordinary "grimpact" case may lead to an understanding of the basis for normal impact and its "societal contract". An extraordinary example of scientific fraud and misconduct leading to the death of patients is the Macchiarini case at the Karolinska Institute in 2016 (Nature 2016) which after a while turned out not to be only the responsibility of an individual surgeon and researcher. The Karolinska Institute, after years of disregarding various reports' concern about clinical and scientific misconduct against their scientist, eventually took the responsibility and followed up, acknowledging that it was a case of violation of the "societal contract" between the Swedish medical research organisation and Swedish society. The extraordinary "grimpact" of the Macchiarini case can be contrasted with a positive example of normal impact in the health sciences. The so-called Health Technology Assessment (HTA) is a well-organised methodology to responsibly avoid harmful treatments in the normal relations between research and healthcare practices (Raftery et al. 2016).

4.5 Involving Stakeholders and Improving Relations

Whether based on a linear or interactive model of understanding, the frameworks for assessing societal impact that we have mentioned above all have their major focus on evaluating the research performing side of the interaction with society. This is understandable since they most often have been developed for the needs of authorities that govern and fund research. However, if the purpose of an evaluation is formative (not only assuring value for money but improvement by learning from advice), and societal impact is studied as an interaction, both sides of the interaction should be able to learn from the evaluation.

In the literature, one approach to understand and evaluate ongoing interactions is named "realist evaluation". The evaluation should include in-depth case studies, focus on formative "real-time" evaluation, and take the "messy, unpredictable, and evolving interaction" into account (Raftery et al. 2016). Realist evaluation considers the mechanism through which the impact is made and suggests that research creates output only insofar as they introduce appropriate ideas and opportunities (mechanisms) in the appropriate settings (context) (Pawson and Tilley 1997). Realist evaluation "elaborates how mechanisms could work in a given context and asks the people who could know about it to provide evidence" (Stame 2004). In order to understand the context-mechanism-output, realist evaluation requires the contribution of the "people who know" (Stame 2004). The stakeholders must therefore be a part of the evaluation process.

A similar mode of thinking was launched a decade ago as the basis for the official policy for Responsible Research and Innovation (RRI) in the Horizon 2020 programme of the European Union (Schomberg 2013; Stilgoe et al. 2013). The policy implied "that societal actors (researchers, citizens, policy makers, business, third sector organisations, etc.) work together during the whole research and innovation process in order to better align both the process and its outcomes with the values, needs and expectations of society". RRI was essentially sharing responsibility, and it depends on groups and organisations rather than on individuals.

The RRI goals might be conflicting with the research evaluation criteria and methods connected to a more general research excellence policy. The selection processes based on international peer review might be disqualifying societal interaction. This type of conflict between international research evaluation regimes and the interaction with local needs has been demonstrated in several studies (e.g. Piñeiro and Hicks 2015; Bianco et al. 2016, and Chavarro et al. 2017) and is one of the motivations behind the *Helsinki Initiative on Multilingualism in Scholarly Communication* (helsinki-initiative.org) that was launched in 2019.

Policy Implications

- *Focus on normal impact rather than extraordinary impact:* Societal impact of research is normal and part of society. Normal impact is about daily activities and how well they are organised, not about individual incidents of particularly interesting or impressive impact.
- *Focus on relations and interactions:* Societal impact evaluation needs to consider both sides in the relations between research and society. The main purpose of the evaluation should be the improvement of the relations, rather than the assessment or funding of one side of the relation. Typologies of impact (e.g. cultural and heritage preservation) need to be supplemented by an identification of the relevant interactors or sectors in society, resulting in a typology of interacting organisations (e.g. museums).

- *Apply an organisational-level perspective:* In general, normal societal impact with possible positive effects can be seen as an organisational-level responsibility, not just as the responsibility of each individual researcher. An organisational perspective may also better serve the implementation and follow-up of societal impact evaluation. Evidence-based case studies imply a linear model of communication and interaction that creates well-known problems with attribution, time frames, etc. An organisational-level evaluation may instead focus on how well the systematic interaction is taken care of (in strategies, infrastructures, management, incentives, etc.).
- *Allow for diversity in incentive and reward schemes:* Normal impact may benefit from clear objectives with incentive and reward schemes that stimulate a wider diversity of tasks and skills within a research group or unit.

References for Further Reading

Bianco, M., Gras, N., Sutz, J. (2016). Academic Evaluation: Universal Instrument? Tool for Development? *Minerva*, 54 (4), 399–421.

Bornmann, L. (2013.) What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Association for Information Science and Technology*, 64(2): 217-233.

Chavarro, D., Tang, P., Ràfols, I. (2017). Why researchers publish in non-mainstream journals: training, knowledge bridging, and gap filling. *Research Policy*, 46 (9), 1666–1680.

Derrick, G.E., & Samuel, G. (2017). The future of societal impact assessment using peer review: Pre-evaluation training, consensus building and inter-reviewer reliability. *Palgrave Communications*, 3, 17040

Derrick, G.E., Faria, R., Benneworth, P., Budtz-Petersen, D., Sivertsen, G. (2018). Towards characterising negative impact: Introducing Grimpact. In *Proceedings of the 23rd International Conference on Science and Technology Indicators*. CWTS/Leiden University. https://hdl.handle.net/1887/64521

Flecha, R., Soler, M., Oliver, E., Puigvert, L., Sordé, T., Schubert, A., Donovan, C. (2014). *Impact evaluation of FP6 (last call) and FP7 SSH research projects. Report 3.* IMPACT-EV. Retrieved from http://impact-ev.eu/wp-content/uploads/2015/08/D3.2-Report-3.-Impact-evaluation-of-FP6-last-call-and-FP7-SSH-research-projects.pdf

Gibbons et al. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, London: Sage.

Giménez-Toledo, E., Olmos-Peñuela, J., Castro-Martínez, E., Perruchas, F. (2023). The forms of societal interaction in the social sciences, humanities and arts: Below the tip of the iceberg. *Research Evaluation*, rvad016, https://doi.org/10.1093/reseval/rvad016.

Godin, B. (2006). The Linear Model of Innovation: The Historical Construction of an Analytical Framework. *Science, Technology & Human Values*, 31(6), 639-667.

Greenhalgh, T., Raftery, J., Hanney, S., & Glover, M. (2016). Research impact: A narrative review. *BMC Medicine*, 14(1).

Klautzer, L., Hanney, S., Nason, E., Rubin, J., Grant, J., & Wooding, S. (2011). Assessing policy and practice impacts of social science research: the application of the Payback Framework to assess the Future of Work programme. *Research Evaluation*, 20(3), 201–209.

Kok, M. O., & Schuit, A. J. (2012). Contribution mapping: a method for mapping the contribution of research to enhance its impact. *Health Research Policy and Systems*, 10(1), 21.

Levitt, R., Celia, C., & Diepeveen, S. (2010). *Assessing the Impact of Arts and Humanities Research at the University of Cambridge*. Technical Report. RAND Corporation. Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED510286

Meagher, L., Lyall, C., & Nutley, S. (2008). Flows of knowledge, expertise and influence: a method for assessing policy and practice impacts from social science research. *Research Evaluation*, 17(3), 163–173.

Molas-Gallart, J., & Tang, P. (2011). Tracing "productive interactions" to identify social impacts: an example from the social sciences. *Research Evaluation*, 20(3), 219–226.

Morton, S. (2015). Progressing research impact assessment: A "contributions" approach. *Research Evaluation*, 24(4), 405–419.

Nature (2016). Macchiarini scandal is a valuable lesson for the Karolinska Institute. Editorial. *Nature*, 537(7619), 137-137. https://doi.org/10.1038/537137a

Olmos-Peñuela, J., Molas-Gallart, J., & Castro-Martínez, E. (2014). Informal collaborations between social sciences and humanities researchers and non-academic partners. *Science and Public Policy*, 41(4), 493–506.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.

Pedersen, D.B., Grønvad, J., Hvidtfeldt, R. (2018). Methods for mapping the impact of social sciences and humanities—a literature review. *Research Evaluation*, 29(1), 4–21.

Piñeiro, C.L., Hicks, D. (2015). Reception of Spanish sociology by domestic and foreign audiences differs and has consequences for evaluation. *Research evaluation*, 24(1), 78–89.

Raftery, J., Hanney, S., Greenhalgh, T., Glover, M., & Blatch-Jones, A. (2016). Models and applications for measuring the impact of health research: Update of a systematic review for the Health Technology Assessment programme. *Health Technology Assessment*, 20(76), 1-254.

REF2014 (2012). Assessment framework and guidance on submissions. Retrieved from http://www.ref.ac.uk/pubs/2011-02/

Schomberg, R. V. (2013). A Vision of Responsible Research and Innovation. *Responsible Innovation*, 51-74.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, Illinois: University of Illinois Press.

Sivertsen, G., & Meijer, I. (2019). Normal versus extraordinary societal impact: how to understand, evaluate, and improve research activities in their relations to society? *Research Evaluation*, 29(1), 66–70.

Spaapen, J., & van Drooge, L. (2011). Introducing "productive interactions" in social impact assessment. *Research Evaluation*, 20(3), 211–218.

Stame, N. (2004). Theory-Based Evaluation and Types of Complexity. *Evaluation*, 10(1), 58-76.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568-1580.

Chapter 5
The Many Dilemmas of Grant Peer Review

Liv Langfeldt

Abstract Peer review is the prime basis for allocating research grants. The systems and principles of grant review were formed in a time with reasonable high success rates and available expert reviewers. Today, in a situation with lower success rates and reviewer fatigue, grant peer review is often heavily criticised. How should distrust, reviewer fatigue, and low success rates be dealt with? This chapter summarises the aims and dilemmas of grant peer review and some advice on how to handle them.

5.1 Why Grant Peer Review?

Both public and private funding agencies use researcher expertise (peers) for evaluating research proposals, and peers' assessments are normally their key basis for allocating research funding. Peer review has two main functions in this: quality assurance and quality enhancement. *Quality assurance* is about ensuring that funded research holds good scientific standard and is in line with programme objectives and of value for science and/or society. The *quality enhancement* aspect, on the other hand, is foremost based on the benefits of competition. Competition between applicants is expected to improve the proposed and funded research. Especially if the competition is high and the funding scheme prestigious, it may attract more qualified applicants, better prepared projects, and reviewers able to improve the projects. The competitive dynamics created by such funding schemes may even have positive effects on the research community as such. While quality assurance requires reviewers competent to filter out inadequate projects, quality enhancement requires highly competent and trusted reviewers, with expertise in the fields of the individual proposals, as well as someone with the ability to compare proposals.

Notably, funding agencies normally try to achieve both quality assurance and competition/quality enhancement. Still, the two functions do not support the same kind of objectives:

L. Langfeldt (✉)
Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway
e-mail: liv.langfeldt@nifu.no

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_5

- *Quality assurance* aims to ensure that funds are spent wisely on promising research projects. The reviewers are gatekeepers who ensure scientific standards and discard research ideas, methods, and perspectives that are not considered adequate or sufficiently interesting and relevant to the call for proposals. Unwanted effects of such gatekeeping may be conservatism in the sense of curbing new and unconventional research ideas, methods, etc.
- *Quality enhancement through competition* aims to make researchers perform better and improve research. Researchers need to excel to obtain a grant and may need to learn specific skills in writing and presenting research proposals. Review procedures are set up to incentivise researchers to develop excellent projects and reviewers to thoroughly assess them. Unwanted effects of this may be that the selection process demands disproportionally more time and resources, possibly concentrating research resources in some environments that are particularly competitive (Mathew effect/cumulative advantages) and reducing pluralism in the funding portfolio. Some research environments and topics may not fit the notions of excellence underlying the selection.

Core principles of grant peer review—such as competence, impartiality, and efficiency (ESF 2011; Science Europe 2020)—are more demanding to ensure when the review process is based more on competition than gatekeeping. The next section outlines the challenges and dilemmas.

5.2 Challenges and Dilemmas

A main dilemma in grant review is ensuring goal attainment in terms of selecting the most adequate proposals and obtaining the objectives in the calls for proposals while minimising the resources for the allocation process (and hence maximising the resources for grants), i.e. combining effectiveness and cost efficiency. Below we discuss the many factors and dilemmas that may impede goal attainment and increase process costs.

Uncertainty and Constructed Agreements The dynamics and logic of science include trial and error, uncertainty, and limited agreement. This easily conflicts with a research funding policy based on research quality as a defined and measurable characteristic to be rated on pre-set criteria and ranked to select the best projects. Grant review is prospective; it assesses a plan for research to be performed. It aims to predict success, and the success factors are uncertain. Reviewers often disagree about research proposals (Cole et al. 1981). They may have different notions of research quality and emphasise different aspects and qualities of the proposals (Langfeldt et al. 2020). Moreover, there are different quality notions in different fields of research, complicating any comparison of proposals between fields, as well as the selection of competence for assessing multidisciplinary proposals (Lamont 2009; Langfeldt 2006). In brief, grant peer review constructs conclusions on something that is genuinely uncertain. While peers may agree on a group of top propos-

als, consensus for differentiating within this top group is much harder. Hence, the problem of uncertainty increases when success rates get as low as 10–20% (Fang et al. 2016; Cole 1992, p. 83; Bornmann et al. 2008, p. 9).

Competence and Conflicts of Interest A main challenge when organising peer review is to match experts to proposals, i.e. defining who is a peer and who has no conflict of interest. And the main dilemma is that the more expertise a reviewer has in the field of the proposal, the higher chance there is for a conflict of interest (Li 2017). In small, specialised international fields with high interaction and close connections it may be impossible to find peers who have no links to the proposed research or the applicants (Chubin and Hackett 1990, p. 194).

Distrust, Reviewer Fatigue, and Low Success Rates Who is assigned to assess a grant proposal is a result not only of who the funding agencies identify and select for the work, but also who has the time for it, and can be motivated. Experts are generally busy and may be reluctant to take on the (often numerous) review tasks they are offered—from a variety of funding agencies and journals—on top of their regular research and teaching assignments. It is not uncommon that research councils have to go far down the list of relevant reviewers before they get a positive reply. According to a study from 2019, funding agency staff may spend 6 h or more to find reviewers for each proposal (Publons 2019, p. 22). Peer review presupposes that reviewers have the competence to be the watchdogs and gatekeepers of science—i.e. that they have competence in the field of the proposal, and preferably at the same level or higher than the applicants they assess. Hence, when we struggle to get competent reviewers to contribute, peer review will not work according to intentions. Combined with low success rates, difficulties in attracting reviewers add to the challenges of providing a thorough and fair review of all proposals. Lower success rates imply much work needed to review a large number of proposals to identify a few proposals to be funded, i.e. increased proposal and review resources per funded project. At the same time, it incentivises applicants to invest ever more resources into preparing (more) competitive proposals. Hence, more resources are put into preparing and reviewing proposals, rather than performing research. *Along with low success rates and demanding review processes comes distrust in the review.* Funding agencies are faced with applicants who argue that their proposals have not been properly reviewed, that the reviewers were not competent, and that the review reports were flawed. Even if a majority of researchers agree that "grant review is the best method of allocating research funding", a substantial proportion does not agree that it is fair and unbiased and treats junior researchers objectively (Publons 2019, p. 20). There is a danger that these factors—low success rates, reviewer fatigue, and distrust in the review process—reinforce each other: For example, that low success rates generate more proposals, make it gets harder to separate the best proposals/select the few winners, increase the burden on reviewers, and generate more distrust.

Biases and Cumulative Advantages Peer review is criticised both for not identifying mistakes and fraud and for being too conservative—to curb innovative and

ground-breaking research (Lee 2015; Lane et al. 2022; Luukkonen 2012). It is furthermore criticised for bias against young scholars and women when basing assessments on applicant's track record (Guthrie et al. 2019). More generally, grant peer review may add to cumulative advantages in research: those with the most resources to write proposals and best track record on grants and publications are best positioned to win. Hence, pluralism may be sacrificed on the altar of grant review. Notably, empirical studies on biases in the review of grant proposals are not conclusive—biases vary by context and further research on the topic is recommended (Arensbergen et al. 2014; Guthrie et al. 2019).

Randomness at Multiple Levels The outcome of grant review is contextual. It depends on who assesses what in what way. And contextual elements leave a lot of room for randomness in the outcome. Your chances in a review process may depend on:

- *The proposals:* The characteristics of the other applicants/proposals, and how many are competing for the grants.
- *The reviewers:* The agendas and motivations of the reviewers, what time they are willing and able to spend on the review job, and which reviewers are assigned to assess and rate your proposal (i.e. "luck of the reviewer draw", Cole et al. 1981). The reviewers' different frames of reference to understand the proposals, e.g. what specific research topics and research environments they are familiar with, and what time they use to expand their frame of reference. If there is a panel meeting for rating and ranking the proposals, the reviewers' scholarly standing and negotiation skills may also impact the outcome.
- *How the review and selection process is organised:* A given proposal may have very different prospects when it competes against proposals in other fields, compared to a process with a separate budget line for each field. In the former situation, chances may depend on the presence and negotiation abilities of field representatives on a review panel. Likewise, with a separate budget line for interdisciplinary proposals or young scholars, these proposals may have better chances for funding than in a process where they compete against disciplinary proposals and senior scholars. Moreover, rating scales, criteria, and budget restrictions may have substantial effects on the outcome. For example, openly defined criteria give more leeway for adapting assessment to different fields and ensuring scholarly pluralism. Assessing scientific and societal value separately may give different results than assessing this jointly. Rough rating scales, heterogeneous panels, open processes, and high success rates give more leeway for innovative/risky projects (Langfeldt 2001).

Generally, there is more room for randomness when success rates are low. When only a few among many projects that may appear equally important and promising are to be selected, the so-called luck of the reviewer draw may play a more prominent role in the process. Under such circumstances, success appears both more unlikely and more unpredictable.

5.3 How to Improve Grant Review

Grant review systems and principles were built up in a time with success rates around 30–50% and fewer demands on expert reviewers. Hence, less time was spent on writing rejected proposals and more applicants were satisfied. In an expanded research system, with high competition, very low success rates, and reviewer fatigue, measures need to be taken to ensure reviewer competence, transparency, fairness, and impartiality:

Increase Competence and Transparency To ensure reviewer competence, one needs to (1) attract and motivate expertise for proposals within specific fields and for interdisciplinary proposals and to (2) enable the selected reviewers to do a good job and to enhance their review competences. To achieve the first, *motivation*, more involvement of the academic society in identifying reviewers and applicant-nominated reviewers may be needed. At the same time, one must take into consideration potential positive bias of applicant-nominated reviewers (Severin et al. 2020). Additionally, in some contexts it may be helpful to include local/domestic reviewers who know the research environments and the funding instruments, and not only detached foreign scholars who may be less willing to devote time in the review. Notably, studies indicate that researchers see grant review assignments as part of their scholarly duty and a service to their field and the research community, while an important reason for declining assignments is that the proposal is outside their field of expertise (Publons 2019). The second element, *reviewer learning*, may be facilitated through discussions in review panels (rather than only individual review), training provided by the funder (Sattler et al. 2015), interviews with applicants, and follow-up of the reviewers providing them with information about the result of the selection process and the outcome of the projects. Furthermore, both motivation and reviewer learning may be promoted by involving the reviewers in developing the review criteria and processes.

Increase Fairness and Impartiality In a situation with high competition and high rejection rates, and proposals based on different fields of research competing against each other, potential field biases need to be monitored. Quality notions vary between research fields and may cause biases in multidisciplinary panels: When some fields have clearer criteria for scientific success and/or higher visibility of successful groups, these fields may more easily succeed in multidisciplinary panels. Adequate measures may be extra efforts on *matching reviewer expertise* to proposals (same field and scholarly perspectives) and *monitoring success rates* for different fields. Moreover, to promote interdisciplinary research, separate budget lines may be needed, while allowing proposals with divergent assessments to be reassessed and reconsidered may reduce biases against unconventional research. A partial *randomisation* (e.g. a lottery among all proposals rated high by reviewers) may also help reduce biases, in addition to saving resources spent on the selection of process (Bendiscioli et al. 2022; Roumbanis 2019).

Policy Implications

- *Low success rates and reviewer fatigue put a pressure on trust and competence in grant peer review.* There is no sole best practice of grant review. Proposal requirements and review procedures need to be adapted to the size and aims of the funding scheme and the number of proposals. Some schemes may promote unconventional research and diversity in the funding portfolio with a partially randomised selection procedure; others may obtain this by selecting and training reviewers dedicated for the aim. *When success rates decrease, the selection processes will need adaption to reduce the burden of the application and review process and new measures to ensure trust. When a high proportion of reviewer invitations are refused, one should consider involving the research community more in identifying and motivating competent reviewers.*
- *Distrust needs to be met by transparency.* Transparency is a key characteristic of good grant review. This demands public information about the selection procedures, review panels and criteria, adequate feedback to applicants, and the possibility for applicants to indicate competent (and incompetent) experts for their proposals and when feasible allow applicants to respond to reviews (rebuttals). Moreover, involvement of and transparency for the reviewers in the review policy and process may give more motivated reviewers and more adequate reviews.
- *Competence and impartiality need to be matched.* The main task in organising grant review is to recruit competent, dedicated, and impartial reviewers; i.e. reviewers should be close enough for a thorough and dedicated review (preferably in the same field as the proposal), but still distant enough to be impartial. Moreover, they should be able to review and compare multiple proposals and have no ties to any of the research environments involved in the proposals. This demands good overview of review expertise and clear regulations and monitoring of conflicts of interest and biases of applicant-nominated reviewers. It may also demand some compromises, e.g. conflicts of interest regulations that are not too strict to allow competent and dedicated peer review.

References for Further Reading

Arensbergen, P. v.; I. v. d. Weijden; P. v. d. Besselaar (2014). The selection of talent as a group process. A literature review on the social dynamics of decision making in grant panels. *Research Evaluation* 23(4): 298–311.

Bendiscioli, S.; Firpo, T.; Bravo-Biosca, A.; Czibor, E.; Garfinkel, M.; Stafford, T.; et al. (2022). *The Experimental Research Funder's Handbook*. Research on Research Institute. https://doi.org/10.6084/m9.figshare.19459328.v2

Bornmann, L.; G. Wallon; A. Ledin (2008). Does the Committee Peer Review Select the Best Applicants for Funding? *PLoS ONE 3(10): e3480*.

Chubin, D. E.; E. J. Hackett (1990). *Peerless Science. Peer Review and U.S. Science Policy.* New York, State University of New York Press.

Cole, S. (1992). *Making Science.* Cambridge, Harvard University Press.

Cole, S.; J. R. *Cole*; G. A. Simon (1981). Chance and Consensus in Peer Review. *Science* 214: 881-886.

ESF (2011). European Peer Review Guide. Integrating Policies and Practices into Coherent Procedures. Strasbourg: European Science Foundation (www.esf.org)

Fang F. C.; A. Bowen; A. Casadevall (2016). NIH peer review percentile scores are poorly predictive of grant productivity. *eLife*. https://doi.org/10.7554/eLife.13323

Guthrie, S.; D. R. Rincon; G McInroy; B. Ioppolo; S. Gunashekar (2019). Measuring bias, burden and conservatism in research funding processes. *F1000Research* 8:851. https://doi.org/10.12688/f1000research.19156.1

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

Lane, J. N., Teplitskiy, M., Gray, G., Ranu, H., Menietti, M., Guinan, E., & Lakhani, K. R. (2022). Conservatism Gets Funded? A Field Experiment on the Role of Negative Information in Novel Project Evaluation. *Manage Sci*, 68(6):4478-4495.

Langfeldt, L. (2006). The policy challenges of peer review: Managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation 15(1):31-41.*

Langfeldt, L. (2001). The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome. *Social Studies of Science, vol 31(6):820-841.*

Langfeldt, L.; M. Nedeva; S. Sörlin; D. Thomas (2020). Co-existing Notions of Research Quality: A Framework to Study Context-specific Understandings of Good Research. *Minerva* 58(1): 115–137.

Lee, C. (2015). Commensuration Bias in Peer Review. *Philosophy of Science*, 82(5): 1272–1283

Li, D. (2017). Expertise versus Bias in Evaluation: Evidence from the NIH. *American Economic Journal: Applied Economics* 9(2): 60–92.

Luukkonen, T. (2012). Conservatism and risk-taking in peer review: Emerging ERC practices. *Research Evaluation* 21(1):48-60.

Publons (2019). Grant review in Focus. Global state of peer review series. https://publons.com/static/Grant-Review-in-Focus-web.pdf

Roumbanis, L. (2019). Peer Review or Lottery? A Critical Analysis of Two Different Forms of Decision-Making Mechanisms for Allocation of Research Grants. *Science Technology & Human Values* 44(6): 994–1019.

Sattler, DN, McKnight PE, Naney L, Mathis R (2015). Grant Peer Review: Improving Inter-Rater Reliability with Training. *PLoS ONE* 10(6): e0130450.

Science Europe (2020). Recommendations on research assessment processes. Brussels: Science Europe. https://www.scienceeurope.org/media/3twjxim0/se-position-statement-research-assessment-processes.pdf

Severin A.; J. Martins, R. Heyard; F. Delavy; A. Jorstad. M. Egger (2020). Gender and other potential biases in peer review: cross-sectional analysis of 38 250 external peer review reports. *BMJ Open* 10:e035058. https://doi.org/10.1136/bmjopen-2019-035058

Chapter 6
How Citations Relate to Research Quality

Dag W. Aksnes and Liv Langfeldt

Abstract Citations are often used as performance metrics in research policy and within the academic community. Usually, citations are assumed to reflect the impact of the research or its quality. What is the justification for these assumptions and how do citations relate to research quality? These and similar issues have been addressed through several decades of scientometric research. This chapter provides an overview of some of the main issues at stake.

6.1 Metrics Use

Traditionally, peer review has been the "gold standard" for research assessment. Increasingly metrics are being applied as an alternative, by its own or in combination with peer review. Examples include the use of citation indicators in the evaluation of the performance of research groups, departments, and institutions, in the evaluation of the track record of grant applicants, in the allocation of research funding, or in the hiring of academic personnel. Citation measures are also core indicators in several university rankings, such as the Leiden ranking and Academic Ranking of World Universities (ARWU).

This raises the question of the reliability and validity of citations as performance indicators. In which contexts and for which purposes are they suitable? These are questions which have been debated over the past few decades.

In the most radical version, it has been argued that assessment of research based on citations and other bibliometric measures is superior compared to the traditional

This chapter is a shortened and adapted version of the article "Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories" written by D.W. Aksnes, L. Langfeldt, and P. Wouters. Published in *Sage Open*, 9(1) 2019. https://doi.org/10.1177/2158244019829575

D. W. Aksnes (✉) · L. Langfeldt
Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway
e-mail: dag.w.aksnes@nifu.no; liv.langfeldt@nifu.no

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_6

peer-review method (Abramo and D'Angelo 2011). Nevertheless, the application of bibliometric indicators for assessing scientific performance has always been controversial. For a long time, the use of journal impact factors in research evaluation contexts has been heavily criticised (Hicks et al. 2015; Seglen 1989), recently reinforced by the Declaration on Research Assessment (DORA 2012) and the Coalition for Advancing Research Assessment (CoARRA 2022).

Moreover, the application of citations indicators has also been criticised more generally, with respect to their validity as performance measures, as well as their potentially negative impact upon the research system (Osterloh and Frey 2015; Weingart 2004).

Broadly speaking, while extensive discussions appeared during the 1970s and 1980s on what citations actually "measure" and how citations relate to scientific quality, this issue seems to have received less attention in recent decades. Nowadays, it is often taken for granted that citations in some way measure scientific impact, one of the constituents of the concept of scientific quality. More attention has been paid to methodological issues such as appropriate methods for normalising absolute citation counts (Waltman et al. 2011), in addition to the development and examination of new citation-based indicators such as the h-index (Bornmann and Daniel 2007; Waltman 2016). Although the latter development has contributed to important progress in the field, the limitations of citations discussed in the 1970s and 1980s did not disappear. In the scientific paper, the references have various purposes. Authors are not including references merely because of their scientific quality. The selection of references is determined by various factors, one being their relevance for the research topic being addressed. These limitations cannot be overcome by the construction of technically more sophisticated or reliable indicators.

6.2 The Origin

The development of bibliometrics as a field is strongly linked to the creation of the Science Citation Index (SCI) by Eugene Garfield in 1961. Originally, this bibliographic database was mainly constructed for information retrieval purposes to aid researchers in identifying relevant articles in the huge research literature archives (Welljams-Dorof 1997). As a supplemental property it enabled scientific literature to be analysed quantitatively.

In the database, all the references of the indexed articles are registered. Based on this, each article can be ascribed a citation count showing how many times it has been cited by later papers registered in the database. Citation metrics can then be calculated for aggregated levels, for example representing research units, departments, or scientific fields. Later several other databases with citation metrics have been launched, such as Scopus and Google Scholar.

The question of what citations "measure" has for a long time been an important question in bibliometrics. Two of the pioneers within citation studies, the Cole brothers, often referred to citations as a measure of quality, although a slightly more

cautious definition was given in the introduction of their book on social stratification in science: "The number of citations is taken to represent the relative scientific significance or 'quality' of papers" (Cole and Cole 1973). Even today, citation indicators are sometimes presented as measures of scientific quality (see e.g. Abramo and D'Angelo 2011; Durieux and Gevenois 2010).

6.3 Scrutinising Validity

Empirical studies have revealed a multitude of factors involved in the citation process. The references have different functions in the scientific article; only a small proportion of the relevant literature is cited and the authors have a multitude of motives for including particular studies as references. To what extent this affects the use of citations as performance indicators is still a matter of debate.

As a validity issue, this has been approached by comparing citation indicators with the outcome of peer review. In these studies, assessments by peers have been typically considered a kind of standard to which citation indicators can be validated. The basic assumption is that there should be a correlation if citations legitimately can be used as indicators of scientific performance.

Overall, most of the comparative studies seem to have found a moderately positive correspondence but the correlations identified have been far from perfect and have varied among the studies. This means that there is so far little empirical support for claiming that citations metrics reflect the same aspects of research quality or impact as peer-review assessments. However, the extent to which the correlation is seen as sufficient depends on the context and goals of the evaluation.

Moreover, the relation between citations and peer assessments is complex and will arise differently depending on the field analysed, the database used, the time frame and indicators applied, and so forth. In addition, research quality is a multidimensional concept, where plausibility/soundness, originality, scientific value, and societal value are commonly perceived key characteristics. Below we will look further into these dimensions.

6.4 Solidity and Plausibility

The first dimension of the quality concept regards the plausibility, soundness, and solidity of the research. How citations relate to or reflect these aspects of the quality concept is complex to assess as many different dimensions need to be considered. Even when solidity and related academic virtues are aspects which are considered by peers when manuscripts are submitted to journals for publications, there are large differences when it comes to the solidity and plausibility of published studies.

The literature contains numerous publications of which the solidity is poor, the results unreliable, or even involving misconduct or scientific fraud (Fanelli 2009).

The latter issue has also been investigated empirically, showing that some publications which have been retracted due to fabrication and falsification of results are very highly cited, some with several hundreds of citations (Fang et al. 2012). Moreover, a disproportionally high number of the articles retracted due to fraud were published in prestigious high impact journals. Although articles retracted due to fraud represent a very small percentage of the overall scientific literature, the problem may be increasing (Fang et al. 2012). The journal referees have apparently considered these papers as sufficiently solid to be published. More generally, there are also indications that methodological soundness and plausibility are not sufficiently emphasised in the review of manuscripts for publication (Lee 2015). Thus, the referee system does not fully ensure the quality dimension related to solidity and plausibility and there are no indications that high citation counts reflect solidity.

The issue may be considered from another angle: that of the reader and potential citer. One might think that in cases where the solidity or plausibility is assessed as poor, the work will not be considered as worth citing (i.e. will be neglected) and in cases where more than one study shows similar results, an author may choose to cite the study she perceives as the most solid. As a consequence, solidity/plausibility—as perceived at the time of citing—may to a certain extent be reflected in citation patterns. There is however little knowledge about the extent to which this actually is the case and studies of citation behaviour have identified a multitude of factors that are not per se associated with the solidity of the studies. Therefore, it seems unlikely that citations can be seen as valid indicators of the solidity of the publications.

6.5 Originality and Novelty

The second dimension, originality and novelty, derives from the fundamental demand that research should produce new knowledge. It seems reasonable to assume that studies with high originality or novelty will be much cited. For example, it has been argued that potential breakthrough discoveries in science can be identified on the basis of citation patterns (Winnink et al. 2016). Moreover, Nobel laurates, who presumably have contributed to research of extraordinary high originality and novelty, tend to be more highly cited than the average scientists (Gingras and Wallace 2010; Wagner et al. 2015), and many have published so-called citation classics. Based on such observations, Garfield previously explored the possibility for using citation statistics to predict future winners (Garfield and Welljamsdorof 1992). At the same time, high citation counts do not necessarily imply breakthrough or Nobel class research. The extremely highly cited Lowry et al.'s paper on protein measurement (Lowry et al. 1951) is an interesting case in this respect. As a consequence of referencing norms, the article has probably been cited almost every time the method has been used. But according to Lowry himself: "It just happened to be a trifle better or easier or more sensitive than other methods, and of course nearly everyone measures proteins these days" (quoted in Garfield 1979).

Example of papers which typically would be considered to have low originality and novelty would be the so-called replication studies. Although such studies are important for the validating of research and for testing and demonstrating the generalisability of existing findings, they tend to be seen as "bricklaying" exercises, rather than as major contributions to the field (Everett and Earp 2015). If the results of studies only corroborate those of previous studies, they have low novelty and are probably less likely to be cited. Many journals appear to be reluctant to publish replications because they would have a negative influence on the citation rate, the impact factor, of the journal (Martin and Clarke 2017). However, attention to the lack of replicable results in biomedical, clinical, and psychological studies (Ioannidis 2005) may lead to a higher social status of replications studies.

The above considerations show that there is no simple relationship between originality or novelty and citations. Studies with high originality may include both major scientific advances and minor contributions. In the latter case, articles may not be cited because their research question is a "dead end" which means that it does not function as a positive basis for further work—despite being novel or original in approach. This brings us to the next dimension of the research quality, scientific value.

6.6 Scientific Value

Scientific value and significance are dimensions of the quality concept to which some citations may most directly relate. This is commonly argued as follows. When a scientist refers to a paper, it has been useful or relevant in some way for the present research or for the writing of the publication. Thus, frequently cited articles may be assumed to have been more useful than publications which are hardly cited or not at all and possibly be more useful and thus important in their own right. This means that the number of citations may be considered as a measure of the article's usefulness, impact, or influence on other research. The same reasoning can be used for aggregated levels of articles. This is the typical way of justifying the use of citations as performance indicator.

In 1983, Martin and Irvine described the conceptual difference between quality and impact in this way: "'Quality' is a property of the publication and the research described in it. It describes how well the research has been done, whether it is free from obvious 'error' […] how original the conclusions are, and so on." The impact of a publication, on the other hand, is defined as the "actual influence on surrounding research activities at a given time". In the view of Martin and Irvine, it is the impact of a publication that most closely is related to the concept of scientific progress—a publication causing a great impact represents a major contribution to knowledge at the time it is published. Using these definitions, it is also evident that impact would be a more adequate interpretation of citations than quality. As an example, even a "mistaken" publication can have a large impact by stimulating further research. Similarly, a publication by a recognised scientist may be more visible and

therefore have more impact, earning more citations, even if its quality (in terms of originality and solidity) is no greater than those by lesser known researchers (Martin 1996).

Impact is the most commonly used concept for what citations reflect, although other concepts such as influence, importance, significance, and utility are also occasionally used (Moed 2005). However, the use of impact as the most appropriate concept has usually been justified by theoretical considerations, and there are few attempts to address the issue empirically, or relate it to previous findings on citation behaviour.

6.7 Societal Value and Relevance

Societal relevance is often considered to be something which is much harder to measure than scientific relevance or impact (Martin 2011). There seems to be a widespread assumption that this issue cannot be adequately assessed through standard citation indicators, and increasing attention has been devoted to developing methodologies for assessing and measuring societal relevance and impact (Bornmann 2013).

A general reason why societal relevance is difficult to assess through citation counts is that the literature indexed in Web of Science and Scopus consists mostly of academic and scholarly publications. While citations may reflect intra-scientific use, use and applications that take place along other dimensions are far less likely to be captured by citation counts in such journals. For example, Hanney et al. (2006) showed that some diabetes papers which were assessed as having had an important impact on clinical practice did not receive many citations. Similarly, research of mainly national or local interest may often be poorly cited by the literature published in international academic journals.

Nevertheless, it is clear that scientific contributions with great societal relevance may also be highly cited. For example, Edward C. Prescott and Finn E. Kydland received the Nobel Memorial Prize in Economics in 2004 for two papers which profoundly influenced the practice of economic policy in general and monetary policy in particular. These papers are also very highly cited in the academic literature. Similarly, in 1994, the Scandinavian Simvastatin Survival Study (4S) provided the first unequivocal evidence that lowering LDL cholesterol via statin treatment reduces cardiovascular events and overall mortality (Pedersen et al. 1994). This paper is now cited more than 9500 times in the Web of Science database. Simvastatin was developed by Merck & Co and came into medical use in 1992 and has had a major impact on human health (Li 2009). Prior to losing its patent protection, simvastatin was Merck's largest-selling drug and second-largest-selling cholesterol lowering drug in the world. Despite these and numerous similar examples, it is not possible to identify societal relevance from citation counts per se and uncited or little cited publication may have contributed to results of great societal relevance.

Policy Implications

- Citations reflect—with important limitations—aspects related to scientific impact and relevance, but there is *no evidence that citations reflect other key dimensions of research quality*.
- *Given the limitations of citation-based indicators, they should not replace peer reviews.* These indicators alone cannot provide a sufficiently nuanced or comprehensive measure of research quality when used in isolation.
- *Peer assessments also have their limitations and shortcomings.* Therefore, bibliometric analyses should be used to complement peer evaluations, and combining these methods can enhance the reliability of evaluations.
- *Analyses covering entire fields of research, nations, or institutions are generally more robust* and less susceptible to the typical concerns and caveats typically associated with citation data.
- *Sensible use of indicators is important.* Citation indicators should not be misused or applied in contexts where they lack justification or validity. For instance, such indicators have larger limitations in the social sciences and humanities, than in the natural sciences and medicine.

References for Further Reading

Abramo, G., & D'Angelo, C. A. (2011). Evaluating research: from informed peer review to bibliometrics. *Scientometrics, 87*(3), 499-514. https://doi.org/10.1007/s11192-011-0352-7

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *SAGE Open*, 9(1), 1-17. https://doi.org/10.1177/2158244019829575

Bornmann, L. (2013). What is societal impact of research and how can it be assessed? a literature survey. *Journal of the American Society for Information Science and Technology, 64*(2), 217-233. https://doi.org/10.1002/asi.22803

Bornmann, L., & Daniel, H. D. (2007). What do we know about the h index? *Journal of the American Society for Information Science and Technology, 58*(9), 1381-1385. https://doi.org/10.1002/asi.20609.

Coara (2022). Agreement on Reforming Research Assessment. https://coara.eu/agreement/the-agreement-full-text

Cole, J. R., & Cole, S. (1973). *Social Stratification in Science*. Chicago: The University of Chicago Press.

DORA (2012). San Francisco Declaration on Research Assessment. https://sfdora.org/read/

Durieux, V., & Gevenois, P. A. (2010). Bibliometric Indicators: Quality Measurements of Scientific Publication. *Radiology, 255*(2), 342-351. https://doi.org/10.1148/radiol.09090626

Everett, J. A. C., & Earp, B. D. (2015). A tragedy of the (academic) commons: interpreting the replication crisis in psychology as a social dilemma for early-career researchers. *Frontiers in Psychology, 6*, 4. https://doi.org/10.3389/fpsyg.2015.01152

Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. *Plos One, 4*(5). https://doi.org/10.1371/journal.pone.0005738

Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America, 109*(42), 17028-17033. https://doi.org/10.1073/pnas.1212247109

Garfield, E. (1979). *Citation Indexing—Its Theory and Application in Science, Technology and Humanities*. New York: John Wiley & Sons.

Garfield, E., & Welljamsdorof, A. (1992). Of Nobel Class—A Citation Perspective on High impact Research Authors. *Theoretical Medicine, 13*(2), 117-135. https://doi.org/10.1007/bf02163625

Gingras, Y., & Wallace, M. L. (2010). Why it has become more difficult to predict Nobel Prize winners: a bibliometric analysis of nominees and winners of the chemistry and physics prizes (1901-2007). *Scientometrics, 82*(2), 401-412. https://doi.org/10.1007/s11192-009-0035-9

Hanney, S. R., Home, P. D., Frame, I., Grant, J., Green, P., & Buxton, M. J. (2006). Identifying the impact of diabetes research. *Diabetic Medicine, 23*(2), 176-184. https://doi.org/10.1111/j.1464-5491.2005.01753.x

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature, 520*(7548), 429-431.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *Plos Medicine, 2*(8), 696-701. https://doi.org/10.1371/journal.pmed.0020124

Lee, C. J. (2015). Commensuration Bias in Peer Review. *Philosophy of Science, 82*(5), 1272-1283. https://doi.org/10.1086/683652

Li, J. J. (2009). *Triumph of the Heart: The Story of Statins*. New York: Oxford University Press.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randal, R. J. (1951). Protein Measurement with the Folin Phenol Reagent. *Journal of Biological Chemistry, 193*, 265-275.

Martin, B. R. (1996). The use of multiple indicators in the assessment of basic research. *Scientometrics, 36*(3), 343-362.

Martin, B. R. (2011). The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster? *Research Evaluation, 20*(3), 247-254. https://doi.org/10.3152/095820211x13118583635693

Martin, B. R., & Irvine, J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy, 12*, 61-90.

Martin, G. N., & Clarke, R. M. (2017). Are Psychology Journals Anti-replication? A Snapshot of Editorial Practices. *Frontiers in Psychology, 8*(523), 1-6.

Moed, H. F. (2005). *Citation Analysis in Research Evaluation*. Dordrecht: Springer.

Osterloh, M., & Frey, B. S. (2015). Ranking Games. *Evaluation Review, 39*(1), 102-129. https://doi.org/10.1177/0193841x14524957

Pedersen, T. R., Kjekshus, J., Berg, K., Haghfelt, T., Faergeman, O., Thorgeirsson, G., . . . Grundstrom, I. (1994). Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet, 344*(8934), 1383-1389.

Seglen, P. O. (1989). From bad to worse: Evaluation by Journal Impact. *Trends in Biochemical Sciences, 14*, 326-327.

Wagner, C. S., Horlings, E., Whetsell, T. A., Mattsson, P., & Nordqvist, K. (2015). Do Nobel Laureates Create Prize-Winning Networks? An Analysis of Collaborative Research in Physiology or Medicine. *Plos One, 10*(7). https://doi.org/10.1371/journal.pone.0134164

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics, 10*(2), 365-391. https://doi.org/10.1016/j.joi.2016.02.007

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics, 5*(1), 37-47. https://doi.org/10.1016/j.joi.2010.08.001

Weingart, P. (2004). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics, 62*(1), 117-131.

Welljams-Dorof, A. (1997). Quantitative Citation Data as Indicators in Science Evaluations: A Primer on Their Appropriate Use. In M. S. Frankel & J. Cave (Eds.), *Evaluating Science*

and Scientists. An East-West Dialogue on Research Evaluation in Post-Communist Europe. Budapest: Central European University Press.

Winnink, J. J., Tijssen, R. J. W., & van Raan, A. F. J. (2016). Theory-changing breakthroughs in science: The impact of research teamwork on scientific discoveries. *Journal of the Association for Information Science and Technology, 67*(5), 1210-1223. https://doi.org/10.1002/asi.23505

Chapter 7
The Ongoing Reform of Research Assessment

Gunnar Sivertsen and Alex Rushforth

Abstract Since being released in July 2022, an Agreement on Reforming Research Assessment has been signed by more than 700 research performing and funding organisations within and outside of Europe. It is intended to guide a reform and mutual learning process within a coalition of its signatories, CoARA. This chapter analyses the agreement critically and provides recommendations for further development.

7.1 Three Contexts for Implementation

The Agreement on Reforming Research Assessment (hereinafter: ARRA) addresses three contexts of evaluation and provides the clearest guidelines for implementation in the first two of them:

1. Individual researchers as they apply for positions, promotions, or internal resources
2. Individual research proposals in applications for external funding
3. Research performing organisations and units

ARRA was developed in collaboration between the European University Association, Science Europe, and the European Commission. Their involvement may strengthen the potential for implementation of the reform in two of the three contexts: While the members of the EUA are directly responsible for assessments in the first context, members of Science Europe are directly responsible for assessments in the second context.

G. Sivertsen (✉)
Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway
e-mail: gunnar.sivertsen@nifu.no

A. Rushforth
Centre for Science and Technology Studies (CWTS) at Leiden University,
Leiden, Netherlands
e-mail: a.d.rushforth@cwts.leidenuniv.nl

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_7

The situation is different for the European Commission and its possible role in the third context. The assessment and funding of public research organisations is a responsibility within widely differing national research systems, as shown in overviews initiated by the Commission (e.g. Jonkers and Zacharewicz 2016) and through mutual learning processes facilitated by the Commission (e.g. Debackere et al. 2018). These experiences and documents demonstrate that mutual learning is more feasible than shared European guidelines for organisational assessment and funding. These conditions for change are neglected by ARRA, and the above-mentioned publications are not among the initiatives and literature listed in the European Commission's scoping report *Towards a reform of the research assessment system* (2021), which laid the basis for ARRA and mainly focuses on agreed reforms in contexts 1 and 2 above.

We reason that ARRA will be easier to promote (though of course not without challenges) in the contexts for assessment of individual researchers and individual research proposals. See Sect. 7.7.

7.2 The Core Commitments and Their Agendas

By signing ARRA, the organisations are effectively committing to ensure that their research assessments will:

- Recognise and reward the plurality of contributions researchers make to academic life (not just publishing and bringing in grant money)
- Respect epistemic differences between research fields
- Reward new (or newly emphasised) quality dimensions such as open science (broadly defined), research integrity, and societal relevance

These commitments reflect two influential agendas in recent years. One of them is the agenda of the official European policy of *Open Research* as adapted to career assessment and development in research organisations. It is inspired by documents such as *Evaluation of research careers fully acknowledging Open Science practices* (European Commission, 2017), *Research assessment in the transition to Open Science* (European University Association, 2019), and national initiatives in the Netherlands, Finland, and Norway (Pölönen and Mustajoki 2022).

The other influential agenda is expressed in the aim "to enable a move away from inappropriate uses of metrics". ARRA follows this agenda of responsible metrics by referencing the *DORA declaration* (2012), the *Leiden Manifesto* (Hicks et al. 2015), *the Metric Tide report* in the United Kingdom (Wilsdon et al. 2015), the Global Research Council's *Statement on Principles on Peer Review/Merit Review* (2018), and the *Hong Kong Principles* for assessing researchers (Moher et al. 2020). With the organisation of CoARA, ARRA might be more successful than the former initiatives in mainstreaming reform in the contexts of individual proposal and researcher assessments.

The two agendas were already becoming connected before ARRA. A recent study of *responsible metrics* as a professional reform movement (Rushforth and Hammarfelt 2023) shows that it evolved from mainly focusing on indicator use to a broader discussion of research assessment criteria and practices and the wider academic culture they help underpin. These movements promote a re-legitimation of evaluative bibliometrics, whereby they can still play a role if used appropriately. ARRA's relationship to bibliometrics is however ambivalent: while endorsing key texts and mantras from the responsible metrics movement, elsewhere the ARRA text is hostile toward bibliometrics, as we show in the next section.

7.3 Neglecting a Field of Research

According to ARRA, "responsible use of quantitative indicators can support assessment where meaningful and relevant". Positive examples of such *indicators* are given, but none are publication-based. The term then changes from *indicators* to *metrics*, as in "journal- and publication-based metrics", with only negative examples. The third of four "core commitments" in ARRA reads:

> Abandon inappropriate uses in research assessment of journal- and publication-based metrics, in particular inappropriate uses of Journal Impact Factor (JIF) and h-index.

ARRA warns that the use of such metrics "may negatively affect the quality and impact of research" and "result in a 'publish or perish' culture that falls short of recognising diverse approaches and could come at the expense of quality".

The term "journal- and publication-based metrics" comes close to a common definition of *bibliometrics*, a term never used in the document, which is a field of research where much professional work is invested in developing appropriate indicators for research assessment. Notably, the three recurring negative examples of metrics in ARRA (JIF, AIS, H-index) are not among indicators professionally developed and tested by the field. ARRA's sometimes hostile tone towards "journal- and publication-based metrics" risks tarring all forms of bibliometrics with the same brush as these discredited examples. To promote responsible development and use of bibliometric indicators, closer relations are needed between ARRA and the field of research that the *Leiden Manifesto* (Hicks et al. 2015) originated from: The annual STI conference series organised by the European Network of Indicator Designers. Restoring the relation to this field of research would be in line with ARRA's call for development, monitoring, and renewal of research assessment criteria, tools, and processes, to be informed by state-of-the-art "research on research" evidence (Commitment 10).

7.4 The Reliance on Peer Review

The Leiden Manifesto defends the use of bibliometric information for research assessment in its first principle:

> Quantitative evaluation should support qualitative, expert assessment. Quantitative metrics can challenge bias tendencies in peer review and facilitate deliberation.

This statement recognises the human aspects and possible limitations of peer review. Much more optimistically, peer review is described in ARRA as "the most robust method known for assessing quality", and possible problems are easily solved: "To address the biases and imperfections to which any method is prone, the research community reassesses and improves peer review practices regularly."

However, current problems with fatigue and distrust in peer review, as they might be experienced among the members of CoARA, are not adequately addressed in the ARRA. Though the establishment of a working group on recognition of peer review is encouraging, it is difficult to envisage how realisation of CoARA's vision would not entail more time and resources needing to be invested in assessments.

Also useful would be a review of the scientific literature of studies of under what conditions and with what possible constraints peer review works well in research assessment. Parts of this literature deal with the increasing problems with reviewer fatigue and distrust and present ideas about how they might be tackled. An overview pertaining to reviews in external funding contexts is given by Liv Langfeldt in Chap. 5, while Ingvild Reymert in Chap. 11 discusses the combination of bibliometrics and peer review in academic recruitment.

There is a tendency in ARRA to play quantitative and qualitative off against one another. We suggest that the challenge is to find the best configuration of both in each assessment context.

7.5 Documentation for Narratives

ARRA not only calls for qualitative rather than quantitative assessment. Another important hallmark is to broaden the basis for research assessment "beyond journal publications". A long list is provided of possible experiences, qualifications, and outputs to be recognised in a holistic assessment.

Earlier investigations of the options for including a wide range of qualifications and outputs in qualitative assessment have arrived at *narratives* provided by the applicant as the solution. ARRA will move in the same direction. The earlier studies observed *documentation* of the narratives as a challenge. ARRA will need to make the same observation and try to solve it.

ACUMEN (Academic Careers Understood through Measurement and Norms), a large EU-funded project in 2011–14, addressed the problem with combining multiple qualitative and quantitative evidence sources for a broad assessment of the

qualifications and outputs of individual researchers. The ideas behind the project were that narratives could not stand alone without documentation and that the sources of evidence needed to be further developed and standardised to avoid large workloads in application and assessment processes. The team developed an "ACUMEN Portfolio" for the purpose and looked for relevant data sources. They found data from social media (as used in *altmetrics*) too limited in scope. They also investigated institutional research information systems as possible data sources without being able to implement the idea. A more recent project funded by Universities Norway, *NOR-CAM*, arrived at a similar possible solution, a flexible and interactive CV drawing on data from the *Norwegian Current Research Information System (CRISTIN)*, but so far without being able to implement it.

Our view is that CoARA could have an important mission in solving the problem with data sources for the documentation of broader qualifications and outputs.

7.6 Publications as Documentation

ARRA gives outputs "beyond journal publications" much more attention than publications. There is a risk, therefore, that ARRA will disregard current developments in the scientific publishing system that may provide available information about research practices in a much broader sense than we are used to.

Publications are peer-reviewed and open to public discussion. The idea behind them is to make the research process behind the results transparent, open to criticism, and available for further use. This idea is often not followed in practice, but it can be reinforced. The developments in digital publishing and towards Open Science allow for this.

ARRA lists items that should be assessed in addition to publications: data, software, models, methods, theories, algorithms, protocols, and exhibitions. All of them are now publishable within a publication, in an appendix, or in linked documents. In fact, all *Indicators of responsible research practices* published with the Hong Kong Principles for assessing researchers (Moher et al. 2020) may now be represented in a scientific publication or by indicators derived from it.

ARRA also says: "Value a range of other contributions to responsible research and scholarly activity, such as peer review for grants and publications, mentoring, outreach, and knowledge exchange". Again, data sources and indicators for such activities are being developed within the scientific publishing system. Examples are those mentioned in the Annex of ARRA: Open science badges; Publons, ORCID, open peer review; CRediT; reporting guidelines (e.g. EQUATOR Network); and metrics (Altmetrics, PlumX).

ARRA will need to clarify the value of scientific publications as documentation for research assessment. They demonstrate experience, achievements, and qualifications from performed research, and they may document many aspects of the research practices that are missing in other information sources.

7.7 Differentiation of Assessment Contexts

ARRA rightly calls for differentiation between different aims and contexts of research assessment. The impression is nevertheless that the same main principles and commitments will be applied in all contexts.

We find the guidelines in ARRA fully adequate for the assessment of persons as they apply for positions, promotions, or internal resources. Within research organisations, for their broad missions to be fulfilled, there needs to be a broad portfolio of qualities considered when evaluating candidates. This requires flexibility, judgement, and discretion: in other words, qualitative peer review, informed—where appropriate—by quantitative indicators.

The same ARRA guidelines may need adjustment and *concentration on research qualifications and their documentation* to be adequate and practical for the assessment of *project proposals* in contexts of research funding. The responsibility of funding organisations is to carefully select the most promising and innovative projects. They may also prioritise research themes independently of immediate institutional recruitment needs and in response to long-term societal needs. Some parts of a CV will be more relevant than other parts. Publications will be significant as documentation of experiences and achievements in performing research.

As indicated in the first section, ARRA tries to address *organisational research assessment* independently of the national systems in which such assessments are at work. Peer review and documentation serve other purposes in these contexts than they do in individual level assessments. Statistics (a term never used in ARRA) can be much more adequate here than in individual level assessments. In general, the ARRA guidelines need to be further developed to be practical in organisational research assessment. Here, again, scientometricians can play an important role, for instance, in helping evaluators to select appropriate advanced, valid indicators, with clear conceptual foundation (Waltman 2018). The core logic of ARRA, that expert peer review ought to be at the heart of organisational assessments, is complicated by the fact that, at this level of assessment, peers typically are not operating in the same (or related) field as the research group or organisation. Instead, evaluators are often being asked to compare between complex organisations, which even individually encompass multiple, different forms of highly specialised research, of which individual peers will have little to no grasp. This is not to say that advanced bibliometrics hold a perfect solution, but, rather, that ongoing debate and awareness about the tensions and complementarities between peer review and bibliometric information is needed when it comes to organisational level assessments. Scientometric debates over several decades can help orient CoARA's followers on the particular challenges peer review and quantitative indicators pose at the organisational level.

As discussed in Chaps. 4 and 11, there is already an overload of summative organisational evaluations in the research sectors. Most of them are related to performance-based funding. Summative organisational evaluations look back at past performance, check whether goals or expectations have been reached, and serve decisions and/or resource allocation. Past performances are usually summed

up from the individual to the organisational level. Formative evaluations, on the other hand, serve strategic development. They do not ask how individual researchers performed; they ask how the organisation could improve in supporting good research. ARRA is mostly focused on reforming the assessment of individual performances. A reform of organisational research assessment in the direction of formative evaluations could also be helpful for achieving the aims of ARRA.

Policy Implications

- There is a need to develop a more constructive approach to bibliometric indicators.
- We suggest collaboration with researchers in the fields of research evaluation and indicator development.
- The increasing problems with reviewer fatigue and distrust need to be considered.
- CoARA could have an important mission in solving the problem with data sources for the documentation of broader qualifications and outputs.
- There is a need to clarify the value of scientific publications as documentation for research assessment. Current developments in the scientific publishing system may provide broader information about research practices.
- ARRA is adequate for the assessment of persons but needs adjustment for the assessment of research proposals. It is so far less adequate for research assessment at the organisational level.

References for Further Reading

Debackere, K., Arnold, E., Sivertsen, G., Spaapen, J., Sturn, D. (2018). *Mutual Learning Exercise: Performance-based funding of university research*: Horizon 2020 Policy Support Facility, summary report. European Commission.

European Commission. (2017). *Evaluation of research careers fully acknowledging Open Science practices – Rewards, incentives and/or recognition for researchers practicing Open Science*.

European University Association. (2019). *Research assessment in the transition to Open Science*.

Hicks, D., Wouters, P., Waltman, L. et al. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature,* 520, 429–431.

Jonkers, K., and Zacharewicz, T. (2016). *Research Performance Based Funding Systems: A Comparative Assessment.* EUR 27837. Luxembourg: Publications Office of the European Union.

Moher, D., et al. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biol,* 18(7): e3000737.

Pölönen, J., and Mustajoki, H. (2022). European recommendations on responsible research assessment. Presented at EASST Conference, Madrid, July 6-9, 2022. Session on Responsible Research Assessment and STS, organized by A. Rushforth, M. Sienkiewicz, J. Zuijderwijk, S. de Rijcke.

Rushforth, A. and Hammarfelt, B. (2023). The rise of responsible metrics as a professional reform movement: A collective action frames account. *Quantitative Science Studies*, 4(4), pp.879-897.

Waltman, L. (2018). 'Responsible metrics: One size doesn't fit all', Presentation at the 23rd International Conference on Science and Technology Indicators (STI 2018). Leiden, the Netherlands, September 13, 2018. https://www.slideshare.net/LudoWaltman/responsible-metrics-one-size-doesnt-fit-all

Wilsdon, J. et al. (2015) *The Metric Tide: The Independent Review of the Role of Metrics in Research Assessment and Management*. https://doi.org/10.13140/RG.2.1.4929.1363.

Chapter 8
The Use of Metrics in Academic Recruitment

Ingvild Reymert

Abstract Concerns about the use of metrics in assessments of individual researchers have been raised by influential initiatives such as the DORA declaration (2012), the Leiden Manifesto (2016), and CoARA's Agreement on reform of research assessment. Even though metrics are applied in the evaluation of candidates for academic positions, there is evidence that metrics primarily serve as supplementary screening tools for panels reviewing applications for academic positions and not as replacements for peer reviews which still serves as the core evaluation practice in academic recruitment.

8.1 Claims that Metrics Have Replaced Traditional Peer Reviews

Academic recruitment is one of universities' most important processes: selecting their most crucial resource, the talented scholars who enable them to fulfil their two primary goals: research and teaching.

Historically, candidates for academic positions have undergone thorough evaluation by tenured professors based on the candidates' research contributions (Herschberg et al. 2018; Musselin 2010; Van den Brink et al. 2010). However, recent studies indicate an increasing reliance on metrics in candidate evaluations, with recruiters demanding candidates' h-index and favouring extensive publication records (Stephan et al. 2017; Van den Brink and Benschop 2011).

These concerns have spurred initiatives like the Leiden Manifesto Hicks et al. 2015, the San Francisco Declaration on Research Assessment (DORA) (n.d.), and the Coalition for the Agreement on Research Assessment (CoARA) n.d., all advocating against the utilisation of metrics for evaluating individual researchers. As of

I. Reymert (✉)
Oslo Business School, Oslo Metropolitan University, Oslo, Norway

Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway
e-mail: ingvild.reymert@oslomet.no

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_8

20 March 2024, 715 organisations have signed the Agreement on Reforming Research Assessment, advocating for broader recognition of researchers' contributions beyond traditional metrics (as outlined in Chap. 7).

Despite claims of widespread use, empirical evidence on the extent of use of metrics in academic recruitment is scarce, partly due to the necessity of confidentiality in recruitment procedures. However, a study analysing confidential recruitment reports from academic recruitment processes does not support these claims (Reymert 2020). In this study, confidential report from the assessment of individual candidates in four fields between 2000 and 2017 at a Norwegian university was analysed.

This study unveils that metrics primarily served as screening tools during the initial stages of recruitment, to narrow and decrease the large pool of applicants to a more manageable group for more thorough evaluation by expert committees. The further decision process involved traditional qualitative peer review and constituted the most important part of the recruitment process.

The findings from the analysis of the recruitment documents were later confirmed by interviews with people involved in the recruitment of professors as well as a cross-country European survey with questions about evaluative criteria in recruitment processes (Langfeldt et al. 2020; Reymert 2021; Reymert et al. 2021; Reymert 2022).

These studies hence indicate that concerns about the use of metrics in recruitment may be exaggerated and that more empirical evidence from and understanding of these procedures may reveal more responsible practices of metrics use than anticipated and critiques by the global initiatives.

8.2 Bibliometrics as Screening Tools

Academic recruitment in Norway is regulated as sequential decisions processes involving a selection committee aimed at screening eligible candidates based on their CVs and research records, an expert committee consisting of scholars conducting a more thorough evaluation of the candidates, and an interview committee.

Figure 8.1 shows the most important criteria used by these three different committees. While metrics were deemed the most important criterion by more than half of the selection committees, metrics held such significance in only one expert committee and in none of the interview committees.

8.3 Disciplinary Differences

The use of metrics in recruitment also varies significantly across disciplines, reflecting divergent views on research quality where each field conducted their own approaches of assessing candidates. These differences are depicted in Fig. 8.2,

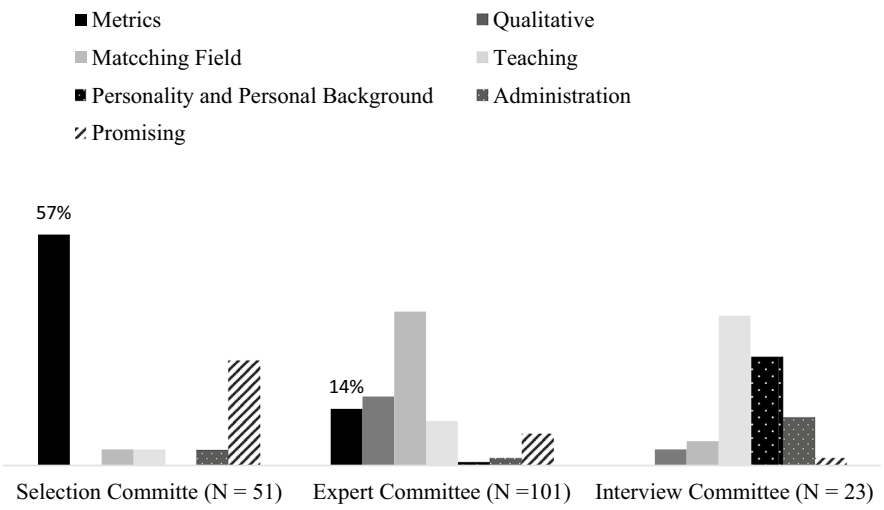


Fig. 8.1 Most important assessment criteria (in percentages) by committee type in four academic disciplines at the University of Oslo from 2000 to 2017 (Source: Reymert 2020/CC BY 4.0)

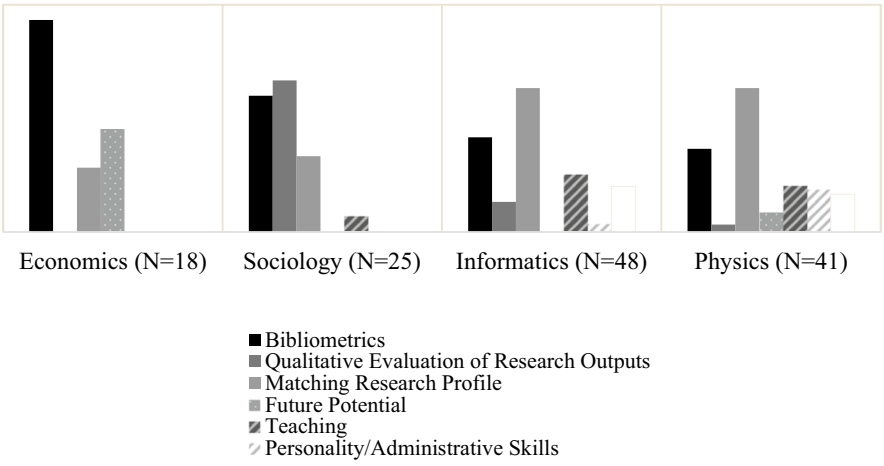


Fig. 8.2 Most important assessment criteria (in percentages) among the expert committees in four academic disciplines at the University of Oslo from 2000 to 2017 (Source: Reymert 2020/CC BY 4.0)

illustrating the most significant criteria in expert committee reports across four distinct disciplines, with "N" denoting the number of identified primary criteria in each discipline's reports.

Even though the reliance of metrics in general is moderate, there was a pronounced reliance on metrics in candidate evaluations in economics. In this field, the study discovered a significant increase in the reliance on metrics from 2000 to 2017,

contrary to other fields where the use of metrics remained relatively stable. Only in economics, candidate evaluations notably prioritised scientific output in international journals, with expert committees consistently deeming metrics the most crucial assessment criterion over the study period. Moreover, in economics, there was a shift in expert committees' reports towards shorter summaries of CVs and metrics, indicating that in this field of research, metrics not only supplement but to some extent replace more quantitative assessment criteria. Hence, claims of increased use of metrics are valid when it comes to economics. The strong reliance on metrics in this field is also observed in other studies (Hylmö 2018).

8.4 Moderate Country Differences

The study of the confidential recruitment processes only covered one country, Norway; however, a comparative survey-based study of assessment criteria in five European countries (Norway, Sweden, Denmark, the Netherlands, and the United Kingdom) instead aimed to unveil potential disparities among the countries.

This study also showed that publication records were just one among several criteria identified by researchers involved in recruitment as crucial. Only about half of the respondents regarded publication records as highly important criteria when evaluating researchers for a position, whereas factors such as the candidate's future potential, the candidate's alignment with the field, the overall impression, and the candidates' research contributions were deemed as highly important by a larger number of researchers (Reymert et al. 2021). These differences are shown in Fig. 8.3 displaying what researchers who had been involved in recruitment identified as highly important criteria when evaluating candidates for a position.

These studies also indicated only moderate differences among countries. The disciplinary disparities were far more pronounced, indicating that the use of metrics in the assessment of individual researchers is embedded in disciplinary cultures (Reymert et al. 2021). Another implication is that the results from the study of the procedures at Norwegian universities may indicate how recruitment is practised in other countries as well.

8.5 Policy Implications

The initiatives to reform research assessment need more than mere concerns about the current use of metrics and agreed-upon principles on how to change them. These initiatives must also be evidence-based, capable of discerning nuances, and willing to learn from good practices wherever they may be found. These initiatives must acknowledge that there is a significant distinction between utilising metrics as an initial screening tool and employing them as a replacement to qualitative judgements.

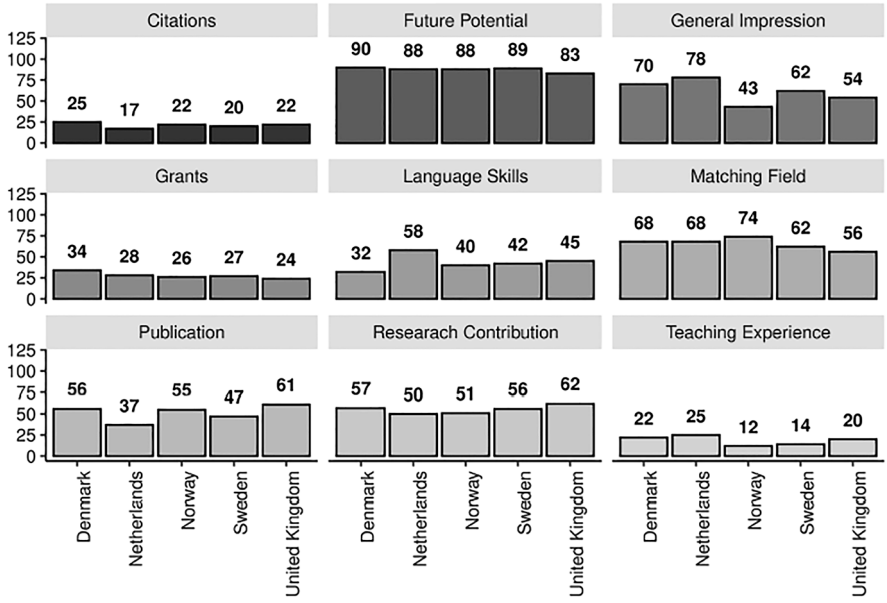


Fig. 8.3 Highly important evaluative criteria when evaluating candidates for a position by country (percent). From a survey to researchers that have participated in recruitment processes (Source: Reymert et al. 2021/CC BY 4.0)

As universities and policymakers endeavour to reform recruitment processes, they must recognise that these processes often unfold sequentially, involving multiple and distinct evaluation stages that assess different competencies using diverse methods and criteria. The incorporation of metrics into recruitment processes does not necessarily imply that they serve as the primary criteria or have supplanted other means of candidate evaluations.

Finding fair and efficient screening methods can be challenging in the global academic job market, with applicants from many different countries with highly diverse backgrounds. Metrics may offer a pragmatic solution to screen numerous candidates if used responsibly, with adequate bibliometric expertise, and with an understanding of field-specific differences and the limitations of datasets and indicators. Given the escalating numbers of applicants, traditional evaluation processes may become overwhelmed, making metrics a feasible option for screening candidates, or at the very least, a more viable alternative compared to other approaches. Suggesting a more thorough evaluation of multiple candidates may also prove to be difficult in a time of peer review fatigue.

Responsible use of metrics in recruitment, as one among many criteria, may also counterbalance the inherent subjective nature of individual candidate assessment, potentially mitigating gender biases and inbreeding.

However, while the study suggests a subtler use of metrics in recruitment, its precise effects remain unobserved. Questions regarding the extent to which

metrics-driven selection committees identify qualified candidates remain unanswered. Even moderate use of metrics may inadvertently deter researchers from pursuing innovative ideas, as they prioritise maintaining requisite publication records for future recruitment prospects.

These nuanced implications underscore the need for further research and that the use of metrics will always need careful consideration when applied in academic recruitment.

Policy Implications

- There is a significant distinction between utilising metrics as an initial screening tool and employing them as a replacement to qualitative judgements.
- If used responsibly, with adequate bibliometric expertise and understanding of field-specific differences and its limitations, metrics may offer a pragmatic solution to screen high numbers of candidates.
- Even moderate use of metrics may inadvertently deter researchers from pursuing innovative ideas, as they prioritise maintaining requisite publication records for future recruitment prospects.

References for Further Reading

The Declaration on Research Assessment (DORA). (n.d.). https://sfdora.org/.

Coalition for Advancing Research Assessment (CoARA). (n.d.). https://coara.eu/.

Herschberg, C., Benschop, Y., & 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(807-825). https://doi.org/10.1007/s10734-018-0237-2

Hicks, D., Wouters, L., Waltman, S. d. R., & Rafols, I. (2015). The Leiden Manifesto for research metrics. Nature, 22(April). https://doi.org/10.1038/520429a

Hylmö, A. (2018). Disciplined Reasoning: Styles of reasoning and the mainstream-heterodoxy divide in Swedish economics [Lund University]. Lund, Sweden.

Langfeldt, L., Reymert, I., & Aksnes, D. W. (2020). The role of metrics in peer assessments. Research Evaluation, 30(1), 112-126. https://doi.org/10.1093/reseval/rvaa032

Musselin, C. (2010). The Market for Academics [Les marché des universitaires]. Routledge.

Reymert, I. (2020). Bibliometrics in academic recruitment: A screening tool rather than a game changer. Minerva.

Reymert, I. (2021). Controlling the Future of Academe. University of Oslo, Faculty of Social Sciences. PhD thesis.

Reymert, I. (2022). Handling Multiple Institutional Logics in Professorial Recruitment. Higher Education Policy. https://doi.org/10.1057/s41307-022-00294-w

Reymert, I., Jungblut, J., & Borlaug, S. (2021). Are Evaluative Cultures National or Global? A cross-national study on evaluative cultures in academic recruitment processes in Europe. Higher Education Journal. https://doi.org/10.1007/s10734-020-00659-3

Stephan, P., Veugelers, R., & Wang, J. (2017). Reviewers are blinkered by bibliometrics. Nature, 544, 411-412. https://doi.org/10.1038/544411a

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., Benschop, Y., & Jansen, W. (2010). Transparency in Academic Recruitment: A Problematic Tool for Gender Equality? 31(11), 1459-1483. https://doi.org/10.1177/0170840610380812

Part II
Research Funding and Governance

Chapter 9
Quality Criteria and Concentration of Research Funding

Kaare Aagaard

Abstract Across many countries concentration of research funding is becoming more pronounced affecting both diversity and topic selection. What is driving these developments? And what are the arguments for and against increased concentration? We address these questions in this chapter and highlight how differing notions of research quality can be both a central driver of concentration and a possible remedy for potential negative effects.

9.1 Trends Towards Funding Concentration

Allocation of research funding is an influential element in governing contemporary science, affecting the scope, content, direction, and impact of public research (e.g. Sörlin 2007; Gläser and Velarde 2018). More pervasive competition, increased performance orientation, stronger emphasis on excellence, and higher reliance on project funding are seen as essential to optimise returns on public investments in science in many countries. These and related developments are likely to affect the balance between concentration and dispersal of the available funding. A central question is therefore: *Do large shares of funding allocated to a small number of scientists yield most value for money? Or is scientific progress and support for societal needs better served by allocating fewer resources across more numerous teams and more diverse research topics?*

These are vital questions given recent research indicates growing funding concentration: Bloch and Sørensen (2015) report a trend towards funding concentration at both individual and group level across a range of countries. Katz and Matter (2019) find funding inequalities in the US National Institutes of Health have increased considerably between 1985 and 2015, with a small segment of investigators and institutes accumulating an increasing proportion of funds. Two Canadian studies (Larivière et al. 2010; Mongeon et al. 2016) find the same trends across a

K. Aagaard (✉)
Aarhus University, Aarhus, Denmark
e-mail: kaag@au.dk

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_9

broad range of fields, and Ma et al. (2015) show similar patterns for UK engineering and physical sciences. However, evidence is still scattered and concentration trends may play out differently across countries, fields, and specialties. Nonetheless, *a thorough examination of concentration, how it develops, and its potential consequences seems both necessary and timely.*

9.2 Concentration of Danish Research Funding

To examine the full degree of concentration within a specific national system, we recently collected funding information for almost 20,000 grants allocated by 15 of the largest public, private, and non-profit Danish research funding foundations during 2004–2016 (Aagaard et al. 2019b; Madsen and Aagaard 2020). Here close to 53 billion DKK (7 billion Euro) was allocated to nearly 7500 PIs (only main grant holders were counted). Our analysis shows that among the grantees alone, the top 20 percent accounted for 75 percent of the allocated funding. Even with a conservative estimate of the full Danish population of public researchers (above PhD level), *the 20 percent of most successful grantees received almost 90 percent of allocated funding* (Fig. 9.1).

A similar picture is observed when recipients are ranked in groups based on their sum of received funding. Figure 9.2 shows that the top 100, measured on grant

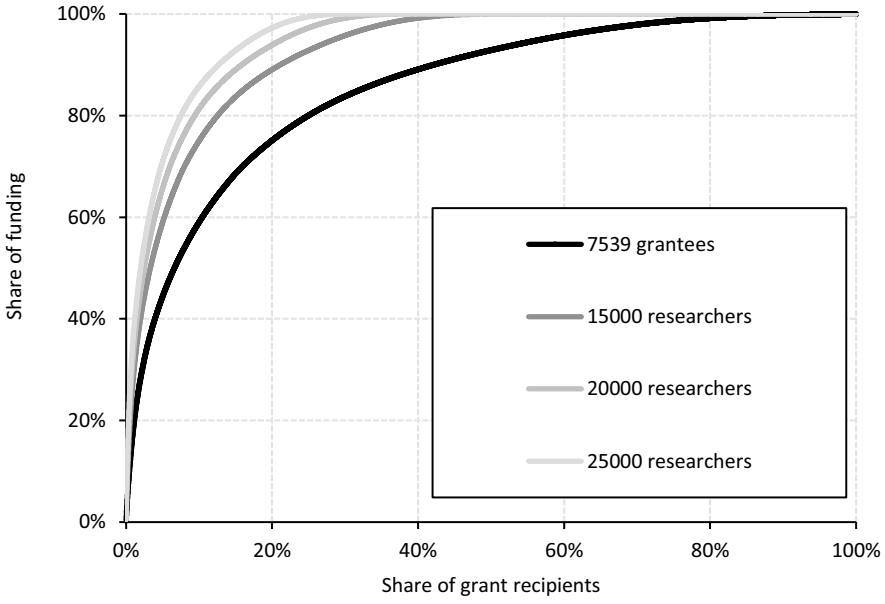


Fig. 9.1 Cumulative distribution of Danish research funding allocated across 15 foundations, 2004–2016. Black line shows grant recipients only. Grey lines show different estimates of the population of public researchers during the period

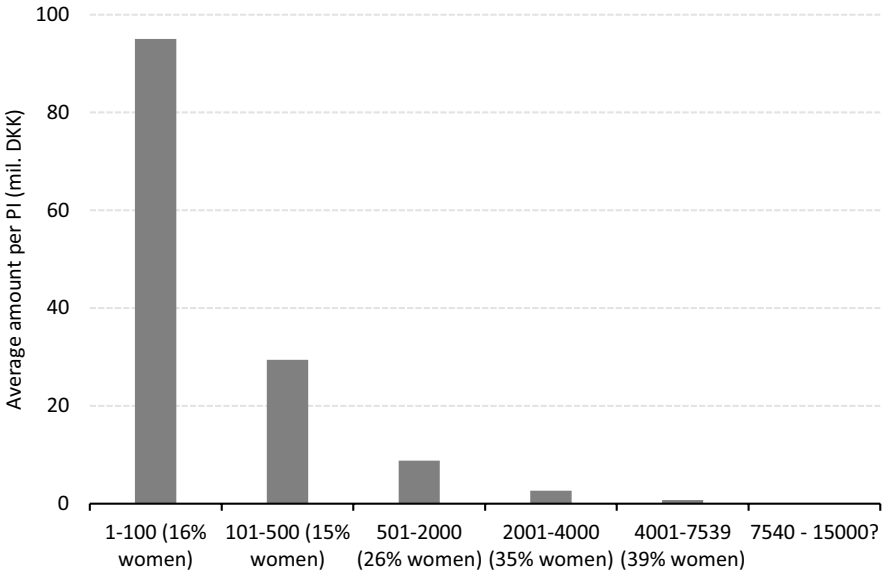


Fig. 9.2 Ranked groups based on total funding received during 2004–2016

success, received an average amount of just below 100 million DKK, while the 101–500 group received an average sum per person slightly below 30 million DKK. Beyond the first 2000 grantees the amounts secured per person become very limited or non-existent. We also see women only accounting for 15–16% of all grantees in the top two groups. The proportion of women increases for the following groups but only as average total grant amount decreases. *In other words, concentration also seems to amplify gender biases.*

However, concentration of funding not only has a gender bias; it also influences the selection of research topics. To examine this, we conducted a more detailed case study on funding of disease-specific research. Here, we correlated funding patterns with so-called DALY (disability adjusted life years) measures developed by the WHO (Madsen and Aagaard 2019). Obviously, these measures cannot alone determine research priorities. But they arguably provide some indication of societal needs, which should be taken into account. Figure 9.3 shows very weak correlations between investment levels and the societal burden of specific diseases. Some disease-specific topics, especially diabetes and breast cancer, are substantially overfunded relative to their DALYs; other diseases garner scant funding relative to their societal burden. These patterns indicate that concentration towards certain disease research topics is not driven primarily by societal needs, i.e. by the proportional social burden related to those diseases.

This disparity between apparent needs and concentrated investments corresponds to evidence we have previously seen elsewhere (e.g. Evans et al. 2014; Jones and Wilsdon 2018; Ràfols and Yegros 2018). Perhaps more surprising is the similarity of topic priorities across funders. In the Danish context, it could for instance be

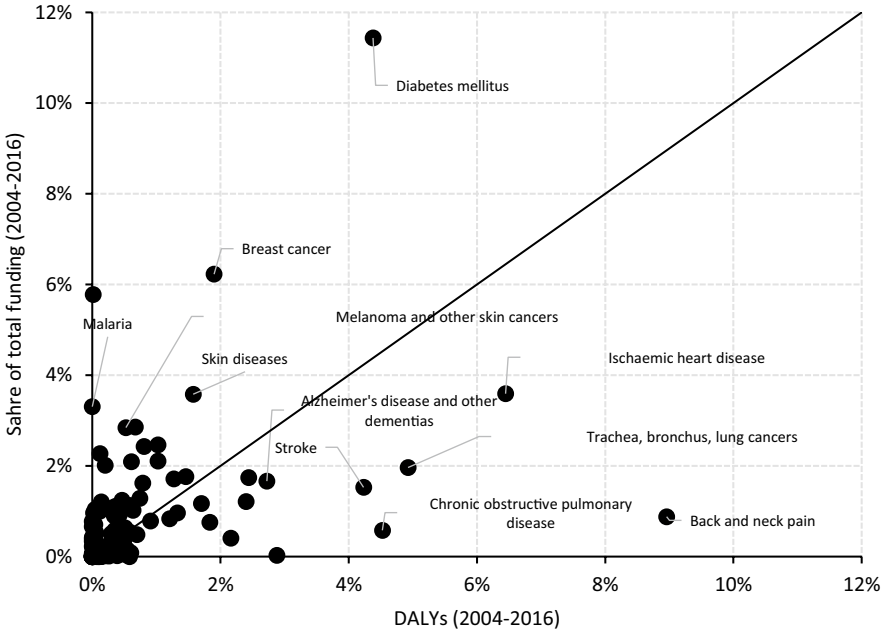


Fig. 9.3 Correlation between funding of disease-specific research and DALYs

assumed that the observed patterns for diabetes and breast cancer are mainly driven by some of the influential non-public funders with particular interests, e.g. the pharmaceutical foundations and the Danish Cancer Society, respectively. These foundations do indeed play a significant role in funding the most well-funded diseases. However, our analysis shows that the majority of funding actually still comes from the public foundations. Hence, *multiple foundation types mirror each other's priorities rather than perform different or complementary roles within the funding landscape*. Recall that these overlapping priorities do not strongly correspond to the burdens of societal needs, so they appear to be driven by other factors.

9.3 Drivers of Concentration

As shown, concentration of Danish research funding is quite pronounced with apparent consequences for the research population as a whole, for gender equality, and for topic selection. This concentration is especially surprising since the Danish system historically has been considered highly egalitarian. This gives us reason to presume we may find similar or greater concentration in other national contexts with comparable or lower egalitarian features. An important question then is how this high degree of concentration might be explained. *There is however no single simple explanation, but most likely rather a number of interacting causes*

reinforcing each other. Firstly, the institutional structure of science is itself biased towards concentration—even in cases where external pressures are absent (Merton 1968). However, recent policy changes in funding and assessment of science likely have amplified this inherent bias. On the one hand, we see conscious and deliberate research policy choices, e.g. larger grants, support for critical mass, and initiatives to create "world leading" environments. But funding concentration may on the other hand also be the result of less obvious, less deliberate factors. Two seem particularly important. The first concerns dominant research quality criteria. When different funding agencies operate with relatively uniform criteria based on narrow notions of excellence (Stilgoe 2014) (typically judged by elite peers supported by metrics like h-indexes and journal impact factors) *priorities are likely to be mirrored even across fairly different funders*. Hence, when a majority of funders aim to pick and fund the "best" researchers based on these similar quality criteria, the result will be increased concentration. This tendency is likely further amplified by a second factor: a lack of oversight of allocation decisions made elsewhere in the system. Lack of coordination and transparency within and across grant bodies may in other words result in higher concentration than any single funder aims. Even if each single grant decision in isolation is sound, systemic effects may be undesirable when the majority of the funders select using identical parameters with many funders inadvertently ending up funding the same researchers and the same narrow topics. Hence, *a combination of strong competition, large grants, low success rates, many competing funding organisations selecting using similar one-dimensional excellence criteria, and lack of coordination is likely to foster undesirable levels of concentration*. And yet exactly these features and dynamics appear widespread and rising in many national funding systems around the world.

In Denmark these characteristics have defined the funding landscape's development for the past 15 years. The share of project funding has increased from less than a third of total research funding to nearly half. *Grant sizes have grown, success rates have dropped, and a drive for excellence has intensified across both public and private funders*. Private foundations often have specific topic interests and aim to establish and support highly visible, impactful research groups and topics. These thereby gain an upper hand within the broader competition for public funding. *This trend is further amplified when the most successful grant recipients subsequently also get rewarded with additional institutional funding* via *performance-based internal funding allocation criteria*. And so the cycle continues and perpetuates even further concentration.

9.4 Pros and Cons

Given all this, we might ask then what levels of concentration might actually strengthen the academic and societal impacts of the science system as a whole? Here we conducted a literature review focusing on scholarly arguments for and against increased concentration of funding (Aagaard et al. 2019a).

Some arguments clearly favour at least some degree of concentration. First, we find a classical meritocratic argument that scientists with greatest potential to produce (potentially) path-breaking research should be rewarded according to their abilities. Economies of scale, critical mass, and access to expensive instrumentation are also marshalled here as arguments for concentration. Funding concentration is furthermore argued to give increased flexibility to researchers, allowing them to take risks and pursue their research process with long time horizons. Other arguments highlight spillovers (the "trickle-down" argument), recruitment, and collaboration effects. These all seem rather strong arguments and yet there are indications—as we return to below—that many of these apparent benefits might also be achieved with more moderate degrees of concentration without the potential systemically counterproductive effects of overly high concentration.

In support of dispersal rather than concentration we find arguments that supporting many lines of inquiry spreads risk and increases chances of breakthroughs by allowing for a broader variety of perspectives, interpretations, heuristics, and predictions. Likewise, chances of serendipity also increase with a multitude of competing approaches. Dispersal at the same time likely secures better alignment with broad societal needs, whereas concentration based on narrow excellence notions focuses scientists' attention inwards rather than on problems of the "outside" world. Dispersal is furthermore perceived to foster resilience in constantly changing research systems, where concentration on the other hand can lead to stagnation and reduced systemic adaptability. Another argument is to avoid large self-perpetuating research units that reduce the capacity of the system to respond flexibly. Concentration is also argued to turn group leaders into "science managers" with little time for research and mentoring and with overly strong incentives and pressure to apply for and obtain ever more resources than can be productively spent. Dispersal alternatively is argued as supporting a broader knowledge pool, creating absorptive capacity across systems as a whole and underpinning research-based teaching across all disciplines. In doing so, it may also secure a strong future growth layer of early and mid-career researchers and keep a broader group of researchers and students active in research. Finally, dispersal is argued as preferable over concentration as it reduces trends towards hyper-competition and may mitigate a peer-review system that is perceived as unreliable, subject to a number of biases and often unable to identify the most promising projects.

9.5 Balancing Dispersal and Concentration

Numerous empirical studies have shown that, on average, there is declining marginal return on each Euro invested in research above a certain threshold. This threshold varies across disciplinary and national boundaries. However, it is not—generally—very high. And these studies only examine concentration from a metrics point of view. Adding the concerns highlighted above the case for increased dispersal may seem even stronger. However, reducing ideal or optimal funding to a simple

question of evidence for or against concentration would oversimplify a complex, multifaceted problem. The "proper" balance between funding concentration and dispersal of research funding is more a matter of degree: *both too little and too much concentration appears inefficient in both economic and epistemic terms*. Similarly, studies also indicate that a healthy research system ecology includes both large and small groups. However, the literature we have reviewed still presents a fairly strong case against high concentration. There are clear indications that most countries and fields need less, not more of it. Policymakers obviously worry about spreading out available funding too thinly, and while some selectivity certainly is justified due to differences in talent and originality across populations of researchers and due to differing expected impacts of various research topics, most systems currently have seemingly moved too far towards high concentration. There is therefore a need to consider how to calibrate these systems better to secure more healthy balances between concentration and dispersal. A number of suggestions are found in the literature:

First, better oversight is needed within and across funding organisations to ensure allocation decisions are more based on broad portfolio perspectives and less on assessments of individual applications in isolation. Secondly, *experiments are needed with funding mechanisms seeking to counter the concentration bias* associated with large parts of current allocation systems (Vaesen and Katzav 2017). A radical proposal here even suggests using a modified lottery model for grant applicants who pass an initial quality screening (Fang and Casadevall 2016). Others suggest experimentation with new funding instruments to promote risky research and diversity, for instance by fully blinding the review process. But *most importantly, there is a clear need to start operating with a broader understanding of research quality*. Here we must acknowledge more explicitly that "excellence" is multifaceted and multidimensional. Allocation mechanisms must be better equipped to capture and reward this inherent variety of academic and societal dimensions.

Real changes will require political will and courage from both public and private funders. Danish experiences suggest that private foundations actually are beginning to take a different view of concentration and dispersal. So far, the implemented changes are only affecting the margins of the system, but more may be coming.

Policy Implications

The policy implications of the examined patterns and drivers of concentration are important. They question the rationale behind funding trends and may point towards more efficient ways to allocate research resources.

- A strong excellence orientation is likely to create self-reinforcing mechanisms rewarding already successful researchers and assign even more funding to research topics that are already very well-supported.
- While this may be justified at the level of each funded project, it may still be undesirable from a systemic perspective.

- A highly excellence-oriented system may in addition further reinforce overly rigid disciplinary boundaries and detach research from broader societal needs.
- Funders operating with broader notions of research quality and experimenting with alternative funding mechanisms may on the other hand make it possible to support more diverse, flexible, and resilient research systems.
- Such systems may both increase chances of scientific breakthroughs and promise better alignment with pressing societal needs and expectations.
- Hence, policies leading to better balances between dispersal and concentration are not only preferable from a scientific perspective, but also likely to be more aligned with values of a democratic society and with the political system that provides the resources in the first place.

References for Further Reading

Aagaard, K., Kladakis, A. & Nielsen, M. W. (2019a). Concentration or dispersal of research funding? *Quantitative Science Studies*, 1-33.

Aagaard, K., Schneider, J. W., & Andersen, J. P. (2019b). Koncentration eller spredning? Fordeling af konkurrenceudsatte forskningsmidler på tværs af danske bevillingsorganer. Aarhus.

Bloch, C. & M. P. Sørensen (2015). The Size of Research Funding: Trends and Implications. *Science and Public Policy* 42(1):30–43.

Evans, J. A., Shim, J. M. & Ioannidis, J. P. A. (2014). Attention to local health burden and the global disparity of health research. *PLoS ONE,* 9(4), 1–9. https://doi.org/10.1371/journal.pone.0090147

Fang, F.C., Casadevall, A., (2016). Research Funding: the Case for a Modified Lottery. *MBio* 7, e00422-16.

Gläser, J. & Velarde, K. S. (2018). Changing funding arrangements and the production of scientific knowledge: Introduction to Special Issue; *Minerva* 56 (1):1-10.

Jones, R. & Wilsdon, J. (2018). *The Biomedical Bubble: Why UK research and innovation needs a greater diversity of priorities, politics, places and people*. London: Nesta.

Katz, Y. & U. Matter (2019). Metrics of Inequality: The Concentration of Resources in the U.S. Biomedical Elite. *Science as Culture*. https://doi.org/10.1080/09505431.2019.1694882

Larivière, V., B. Macaluso, É. Archambault & Y. Gingras (2010). Which Scientific Elites? On the Concentration of Research Funds, Publications and Citations. *Research Evaluation* 19(1):45–53.

Ma, A., R.J. Mondragón & V. Latora (2015). Anatomy of Funded Research in Science. *Proceedings of the National Academy of Sciences of the United States of America* 112(48):14760–65.

Madsen, E.B. & Aagaard, K. (2020). Concentration of Danish research funding on individual researchers and research topics: Patterns and potential drivers. *Quantitative Science Studies* 1(3):1159–1181.

Madsen, E. B. & Aagaard, K. (2019). *Fordeling af forskningsbevillinger i Danmark: Fordelingen af udvalgte konkurrenceudsatte forskningsmidler på hovedområder og discipliner, 2004-2016.* Commissioned report for The Danish Research and Innovation Policy Council (DFiR). Aarhus.

Merton, R.K. (1968). The Matthew Effect in Science. *Science* 159(3810), 56–63.

Mongeon, P., C. Brodeur, C. Beaudry & V. Larivière (2016). Concentration of Research Funding Leads to Decreasing Marginal Returns. *Research Evaluation 25(4):rvw007.*

Ràfols, I. & Yegros, A. (2018). Is research responding to health needs? *Social Observatory of the La Caixa Foundation*, 1–13.
Stilgoe, J. (2014). Against excellence. *The Guardian*, 19.
Sörlin, S. (2007). Funding diversity: performance-based funding regimes as drivers of differentiation in higher education systems. *Higher Education Policy*, 20(4), 413-440.
Vaesen, K., Katzav, J. (2017). How much would each researcher receive if competitive government research funding were distributed equally among researchers? *PLoS One* 12, e0183967.

Chapter 10
Balancing Basic and External Research Funding: A Comparative Analysis

Espen Solberg and Christina Vogsted Drange

Abstract Research funding allocation varies significantly among countries. Is there an optimal balance between direct allocation (basic funding) and competitive funding schemes? In this chapter, we explore this question using scholarly literature and comparative data on R&D funding.

10.1 Main Forms of Public R&D Funding

Most national research systems rely heavily on public funding. Hence, public allocations to research and development (R&D) constitute a main tool in research policy. While most attention is paid to the level of funding, the way public funds are allocated is also of great importance. A key question in this context is whether public funds should be distributed directly to research performing organisations or provided as external funding, often exposed to open competition for research projects. Balancing these two funding modes[1] constitutes a major policy question.

Available R&D statistics show considerable variations in this balance, between countries, sectors, and over time. These variations are not only the result of policy concerns, but are strongly dependent on the funding systems and traditions at place in each system. Nevertheless, our review of recent policy documents indicates that different systems might share a common "pain threshold" if basic funding drops below a certain level.

[1] In order to align with the terms most commonly used in the scholarly literature and relevant policy documents, we use the terms "basic funding" and "external funding", while we comply with the distinction between institutional and project funding suggested by Lepori et al. (2023) and van Steen (2012).

E. Solberg (✉) · C. V. Drange
Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway
e-mail: espen.solberg@nifu.no; christina.drange@nifu.no

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_10

10.2 No Apparent Golden Formula

Within the rich scholarly literature on public research funding, several studies have focused specifically on the relationship between external funding and basic funding. See Schwach et al. (2024) for an updated overview.

A key finding in this literature is that the share of external funding has increased considerably, particularly during the 1990s and the first part of the 2000s. This is often linked to the emergence of New Public Management and its ensuing focus on competitive public finding in general. Others underline the increased importance of accountability and "value for money" driven by constraints on public expenditure (Geuna and Martin 2003). Over the past decade there are signs that the balance between basic and external funding is stabilising, while public funding is generally stagnating or declining (Zacharewicz et al. 2023).

Many studies also observe an increasing and more widespread use of performance-based components within the basic funding allocations. This is apparent across countries, and particularly from the late 1990s (Hicks 2012; Sivertsen and Aagaard 2024; Jongbloed et al. 2022). Such performance-based research funding systems appear in three major types with possible combinations: evaluation-based, indicator-based, and performance agreements (Sivertsen 2023).

The literature is, however, more mixed when it comes to how the balance between forms of funding affects the productivity, quality, or other factors of research performance: Some studies suggest that institutions with high external funding have higher productivity and better research results (e.g. Aghion et al. 2010), while others come to the opposite conclusion (e.g. Sandström and Van den Besselaar 2018). And several studies find no clear correlations between research results and various forms of funding (Auranen and Nieminen 2010; Zacharewicz et al. 2023).

Many studies have also sought to identify the pros and cons of the two forms of funding. The most important arguments in favour of basic funding are, firstly, that it provides research institutions with stable and predictable funding so that they can plan for the long term. Also, direct basic funding saves time and resources that would otherwise be used to prepare and assess project proposals. A third argument is that basic funding is more appropriate for supporting unconventional, groundbreaking, and risky research projects, as it is often held that competition based on peer review will tend to select safe applications using recognised methods. Basic funding can also make it easier to link research activity to the institution's own goals and priorities, for example to their teaching profile or regional knowledge needs. A further argument is that basic funding may counter the risk that some disciplines or research groups win a disproportionate large share of funds, the so-called Matthew effects.

A main argument *in favour of* external funding is that open competition for funding makes it easier to identify and support the best and most promising research, as project proposals are subject to external quality assessments. There is also a widespread perception that competition for funding gives researchers an incentive to raise their professional level in order to succeed in the competition. Furthermore, external funding provides funding authorities with a better and more legitimate

basis for steering funding towards research responding to national concerns, for instance thematic and technological priorities. External funding can also promote efficient use of resources, as externally funded projects often require clear budgets and milestones.

Many of the arguments above can be used both positively and negatively, depending on policy needs and priorities. For example, a concentration of R&D resources may be desirable in some contexts and undesirable in others. A main conclusion from the literature is therefore that there is no "silver bullet" (Stampfer 2019). The balance between basic funding and external funding is thus not only a question of "what works", but just as much a question of what governments will achieve with their funding. A dynamic, evidence-based, and continuously learning approach to the question of a balance may be needed.

10.3 Considerable Variations Across Countries

In the absence of a common gold standard, international comparisons can provide an indication of what constitutes a reasonable level.[2]

Figure 10.1 displays country differences over time in the share of so-called general university funds" (GUF). According to the OECD, this indicator reflects the share of public funds allocated directly to higher education institutions.[3]

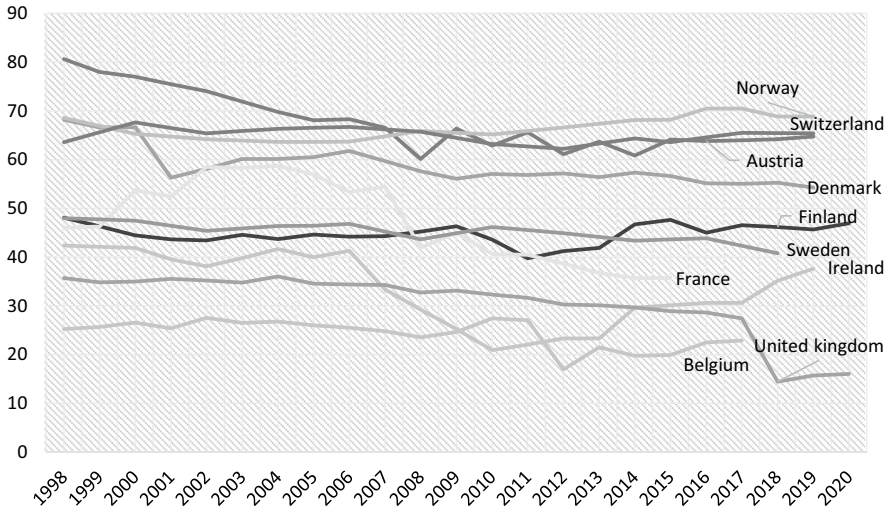
The figure displays a general decrease in the share of basic funding of R&D in the higher education sector in most countries. This confirms the general trend observed in the literature. We also see that Switzerland, Austria, and Norway stand out with a relatively high share of basic funding. For these three countries, recent data show a basic funding ratio of close to 70%, while in many comparable countries the share is around or below 50%. Norway also stands out as the only country with an increasing share of basic funding. The figures must be read with the precaution that "general university funds" can be defined and measured somewhat differently in each country. However, a cross-check with R&D statistics in the Nordic countries shows that OECD figures largely comply with national statistics.

An alternative data source for international comparisons is the European Tertiary Education Register (ETER), which inter alia provides comparable data on the share of "core funding" for European universities.[4] This largely corresponds to the term "basic funding" used in this chapter, although the ETER figures apply to the entire activity of the higher education sector, including the teaching activities.

[2] The main emphasis here is placed on the university sector, as there are few comparable data for the funding profile of research institutes and other R&D performing organisations.

[3] See OECD Frascati Manual chapter 12 Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development | en | OECD.

[4] ETER defines core funding as "funding available for the operations of the whole institution, which is not earmarked to specific activities and whose internal allocation can be decided freely by the institution itself".



Source: OECD/MSTI indicators 2023.

Fig. 10.1 R&D financed by general university funds (GUF) as a share of total higher education R&D expenditure. Selected OECD countries 1998–2021. Source: OECD/MSTI indicators 2023

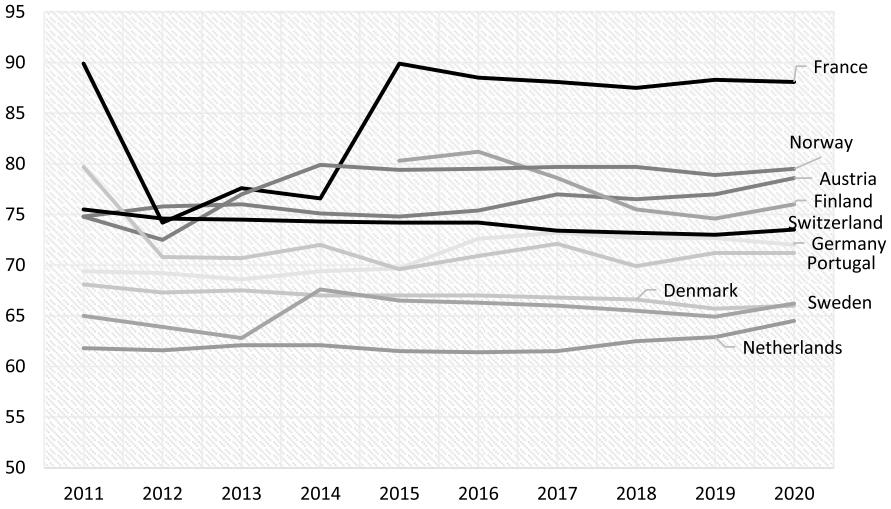
The ETER data reflect much of the same pattern as shown by the OECD figures. However, the share of basic funding is generally higher, primarily due to the fact that the majority of basic funding is used to finance higher education, which in many countries is regarded as a public responsibility and funded accordingly.

The share of core funding also shows a more stable pattern over time, which reflects that a main purpose of basic funding is to provide institutions with a predictable financial foundation. Again, we see that Norway and Austria appear with relatively high shares of core funding. Austria has experienced a particularly strong growth in this share, but the country now has a stated ambition to increase the share of external funding (Austrian Federal Government 2020).

The exceptionally high level of core funding in France contrasts with the same country's relatively low level of general university funds shown in Fig. 10.1. The main explanation for this discrepancy is that basic funding for French higher education institutions mainly covers education, while R&D activity is primarily financed by external funds and through basic funding to large public research organisations (PROS), of which the Centre National de Recherche Scientifique (CNRS) plays a major role. The United Kingdom and Ireland (not included in the figure) are examples of countries where universities' basic funding accounts for less than half of the total budget, mainly because tuition fees constitute a major source of funding.

A third source to shed light on the same phenomenon is Eurostat's regular data based on the distinction between institutional funding versus project funding.[5] In

[5] See Eurostat's metadata for further explanation: Government budget allocations for R&D (GBARD) (gba) (europa.eu).



Source: European Tertiary Education Register (ETER). NIFU's calculations.

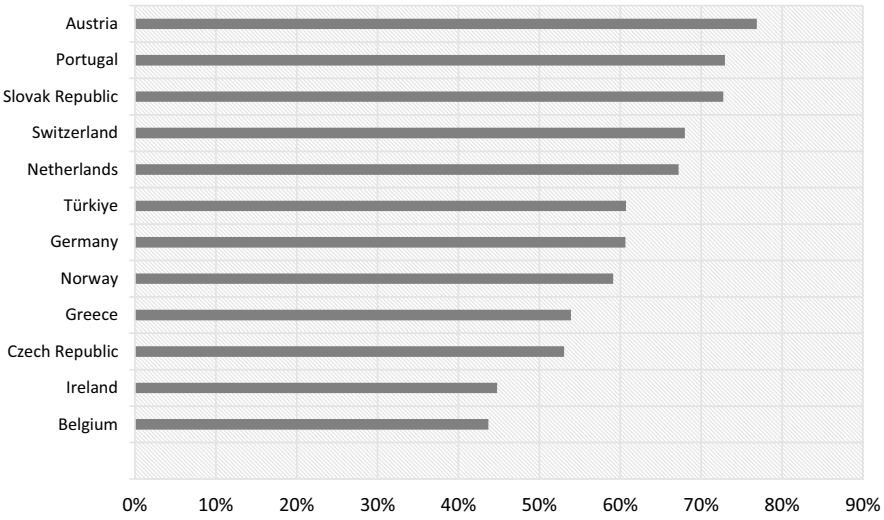
Fig. 10.2 Total core funding of the higher education sector. Share of the sector's total revenue. Selected European countries 2011–2020. Source: European Tertiary Education Register (ETER). NIFU's calculations

contrast to the data used in Figs. 10.1 and 10.2 , Eurostat's data include *all public R&D allocations* and are not limited to the higher education sector. The figure below shows the share of direct institutional allocations in 2021, which is the last year with updated statistics (Fig. 10.3).

We see that Switzerland and Austria still appear with rather high shares of institutional funding, but for other countries, the picture is changed. Portugal and the Netherlands now appear with relatively high shares of institutional funding, while Norway's share is considerably more modest. The drop in Norway's total share of institutional funding is largely due to the fact that the share of basic funding for independent research institutes is relatively low compared to that of similar institutes in other countries (OECD 2017; Solberg et al. 2018). The example shows the importance of observing that comparisons of the general public R&D funding provide a different picture from comparisons limited to only the higher education sector.

10.4 A Need for Better Comparisons of Research Institutes

Comparative data for the funding of independent research institutes are, however, poorly developed. The OECD has recently carried out a pilot study to map and categorise the extent of so-called R&D specialist institutions (Galindo-Rueda and van Beuzekom 2023), but so far without yielding official definitions and data.



Note: France has previously reported shares of institutional funding amounting to 90 percent, but has not reported data since 2013, and is therefore excluded from the chart
Source: Eurostat

Fig. 10.3 Public R&D allocations allocated as institutional funding. Share of total R&D allocations in 2021. Note: France has previously reported shares of institutional funding amounting to 90%, but has not reported data since 2013, and is therefore excluded from the chart. Source: Eurostat

Comparative studies of research institutes must therefore rely on ad hoc surveys and unofficial data.

One example is the joint OECD/EARTO study of 132 Research and Technology Organisations (RTOs) in Europe and beyond (Larrue and Strauka 2022). The study observed a general increase in the R&D activity of the research institutes covered. Measured in total revenues and number of employees, the relevant institutes have more than doubled between 2007 and 2019, both driven by the establishment of new institutes and increased activity in existing institutes. Secondly, the study found that basic funding is the largest source of income for many institutes, accounting for nearly 40% of the total in 2019. This share is however strongly influenced by some large institutes with particularly high basic funding. Among small and medium-sized institutes, the share of basic funding was found to lie between 20% and 30%. At the same time, survey data from the same study show that high basic funding often includes elements of steering from the funding authorities. Many European institutes also use part of their basic funding to cover co-financing of EU projects and funding from national research councils.

10.5 Policy Statements and Trends in Four Nordic Countries

To supplement the figures above, we have reviewed recent policy documents and reports in four of the Nordic countries, with particular focus on policy trends and considerations regarding the balance between basic funding and external funding:

Denmark's research system has undergone major overhaul in the last two decades and now consists of eight universities receiving the majority of publicly funded research. According to a recent review by the Danish Council of Research and Innovation Policy (DFiR), 75% of Danish university researchers do not have access to research funding from internal institutional funding sources. In contrast, Danish universities have considerable access to external funding, both from public and, not least, large private funds and foundations. Still, the survey shows that 23% of the researchers are also without access to such external funding. Hence, the Danish university system experiences major differences in the access to R&D funding among individual researchers and research groups. The institutions also report pressures on basic budgets as external funds require considerable co-financing from the institutions. In addition, concerns are raised that basic funding for research must be used to cover teaching activities. Dfir therefore concludes that "the balance is tipped", meaning that Danish universities' dependence on external funding has become too high. On this background, Dfir recommends increased basic funding and a clarification of how indirect costs of external financing are to be covered. The latter was largely met through a new agreement between the Danish authorities and the private research funding organisations, where the latter agree to cover a larger share of the real expenses associated with externally financed research projects.[6]

Current policies for R&D funding in *Finland* are expressed in a long-term plan published by a parliamentary working group appointed by the government (Finnish Government 2023). The plan calls for an ambitious and broad increase in research funding. The balance between external funding and basic funding is not discussed explicitly, but the plan calls for increased public basic funding of universities, with an emphasis on the education of candidates and researcher recruitment. Following a major funding reform in the period 2014–17 a substantial share of direct basic funding to research institutes was reallocated to external funding and subjected to open competition through a new strategic research council within the Academy of Finland. The recent long-term plan states that external funding will be further increased, both through increased funding from the EU and growth in allocations through the Academy of Finland. The latter is backed up by a recent evaluation by the Academy, which found that the grant rate for free project applications was "alarmingly low" (Finnish Ministry of Education and Culture 2022).

In *Norway*, public R&D expenditure has experienced a strong and steady increase for the last two decades. During this period, direct allocations to universities and university colleges have increased more than the funds distributed by the Research Council of Norway (RCN), which is the most important source of external R&D

[6] https://dkuni.dk/aftale-om-en-faelles-model-for-finansiering-af-forskningsprojekter/

funding. Hence, as shown in Fig. 10.1, Norway has been "swimming against the stream" both in terms of the volume and the development in basic funding of the education sector. While the current national long-term plan for research and higher education does not address the funding balance explicitly (Ministry of higher education and research 2022), the Norwegian Parliament expressed in 2023 an ambition to "increase the share of basic funding" in the years to come (Norwegian Parliament, Innst. 170 S (2022–2023). At the same time, a considerable adjustment of the higher education funding system has been adopted, removing performance-based indicators related to research activity and leaving the steering of R&D activity to performance agreements and the institutions' own priorities (Meld. St. 14 (2022–2023)). In these processes, the functions and framework conditions of research institutes and other research performing organisations have not been assessed or subject to reforms.

In *Sweden*, basic funding, the so-called basanslagen, accounts for just under half of the research expenditure at Swedish universities and university colleges. External funding has increased in importance over the last 10 years (see Fig. 10.1). A government-appointed committee has recently published a comprehensive assessment of Sweden's system for research funding (SOU 2023, p. 59). As part of the review, the report addresses the balance between external funding and basic funding, e.g. based on a survey of various actors in the Swedish research system. The survey reveals different views on this balance. Many actors argue that basic funding should increase, but not at the expense of external funding. The large universities express concerns that the institutions have become too dependent on external funding and that too much of the institutions' basic funding must be used for co-financing externally funded projects. The Commission largely shares this concern. They suggest that increased basic funding may be a solution to the problems identified, but propose no concrete model or funding formula (SOU 2023, p. 59). The report will be followed up in the Swedish research proposition, which is expected to be presented towards the end of 2024.

10.6 Conclusions and Policy Implications

The scholarly literature provides no clear indications of an optimal balance between external funding and basic funding. International comparisons also indicate large country differences. This means that changes in the general balance between basic and external funding should take into consideration the different purposes, features, and incentives that constitute their funding systems. For instance, recent developments in Sweden and Denmark show that increased dependence on external funding from private foundations may require measures that reduce the demands for co-funding of such resources. On the other hand, high shares of basic funding combined with few elements of performance-based institutional funding should raise questions about the need to increase the share of external funding, as for instance in Norway.

Furthermore, basic and external funding are two broad funding categories where each funding stream may contain a large variety of incentives and mechanisms. The balance between basic and external funding is therefore also a question of which types of funding arrangements that are at play within each category. For instance, the more external funding instruments are dominated by generic excellence-oriented mechanisms, the more basic funding mechanisms need to include elements of thematic orientation.

Policies should also account for *the entire* research system when designing their funding mechanisms and profile. As demonstrated in this chapter, research institutes and other research performing organisations have different funding profiles and framework conditions than higher education institutions, while they often compete for the same sources of external funding. Funding authorities should therefore to a larger extent assess how the funding systems impact all R&D performing sectors and design policies accordingly.

For this to happen there is a need to pursue ongoing efforts to develop more systematic and comparative data on R&D activity and R&D funding for research institutes and other actors outside the established categories and measurement systems. Better and more systematic data are also needed to capture R&D funding from private foundations. These funding streams are still subject to ad hoc data collections, but their importance is likely to increase as financial constraints limit the room for increased public funding in many countries.

Policy Implications

- Since an optimal balance between basic and external funding cannot be derived either from the scholarly literature or from international comparisons, governments should consider their funding mix in light of what works and what is needed in each system.
- As different research organisations collaborate and compete for many of the same funding sources, changes in the funding mix should not be purely based on its implications for universities but consider the implications for all actors in the system, including applied research institutes and hospitals.
- Likewise, policies for research funding should not focus solely on the implications of public measures, but account for the criteria and incentives that are also embedded in private and international funding mechanisms, and thus avoid overlaps, gaps, and other unintended effects in the entire funding landscape.
- In order to do so, governments need to develop better and more harmonised tools to monitor and compare the functioning of entire funding systems. In particular, there is a need to develop indicators that better capture the role of research institutes as well as the variety of funding mechanisms on the meso-level of funding systems, where research councils and funding agencies operate.

References for Further Reading

Aghion, P., Dewatripont, P., Hoxby, C, Mas-Colell, A. og Sapir, A. (2010) The governance and performance of universities: evidence from Europe and the US. Economic Policy, 25(61), 7–59.

Auranen, O., & Nieminen, M. (2010). University research funding and publication performance—An international comparison. Research policy, 39(6), 822–834.

Austrian Federal Government (2020) RTI Strategy 2030 Strategy for Research, Technology and Innovation of the Austrian Federal Government. Vienna, 2020.

Finnish Government (2023) Multi-annual plan for the use of research and development funding. Final report of the Parliamentary RDI Working Group 2022. Publications of the Finnish Government 2023:56.

Finnish Ministry of Education and Culture (2022) Evaluation of the Academy of Finland. Publications of the Ministry of Education and Culture, Finland 2022:7

Galindo-Rueda, F. and van Beuzekom, B. (2023) The contribution of R&D specialist institutions to R&D performance: Findings from the NESTI 2022 pilot data collection. OECD Science, Technology and Industry Working Papers, 2023/04, OECD Publishing, Paris.

Geuna, A., Martin, B.R. University Research Evaluation and Funding: An International Comparison. Minerva 41, 277–304 (2003).

Hicks, D. 2012. Performance-based university research funding systems. Research Policy 41(2) 251–252.

Jongbloed, B., C. MC Grath, H. de Boer, and A. d. Gayardon (2022). Final Report of the Study on the state and effectiveness of national funding systems of higher education to support the European Universities Initiative (Volume 1) European Commission, Directorate-General for Education, Youth, Sport, Culture.

Larrue, P. and Strauka, O. (2022) The contribution of RTOs to socio-economic recovery, resilience and transitions. OECD Science, Technology and Industry Policy Papers, 129/22 OECD Publishing, Paris.

Lepori, B., Jongbloed, B., & Hicks, D. (2023) Introduction to the handbook of public funding of research: understanding vertical and horizontal complexities. In Handbook of public funding of research (pp. 1-19). Edward Elgar Publishing.

Government Offices of Sweden (2023) SOU 2023:59. New agency structure for funding research and innovation.

Norwegian ministry of higher education and research (2022) Meld. St. 5 (2022–2023) Long-term plan for research and higher education 2023–2032. Norwegian government 2022.

OECD (2017) Reviews of innovation policy: Norway 2017. OECD publishing. Paris, 2017.

Sandström, U. and Van den Besselaar, P. (2018) Funding, evaluation, and the performance of national research systems. Journal of Informetrics,12, 365-384.

Schwach, V., Borlaug, S.B., Wollscheid, S. and Fossum L.W. (2024) Research and funding. An overview of research literature. NIFU Working Paper 2024:2.

Steen, J. (2012) Modes of Public Funding of Research and Development: Towards Internationally Comparable Indicators", OECD Science, Technology and Industry Working Papers, No. 2012/04, OECD Publishing, Paris, https://doi.org/10.1787/5k98ssns1gzs-en.

Sivertsen, G. and K. Aagaard (2024). Designing performance-based research funding system. In G. Sivertsen and L. Langfeldt (eds). *Challenges in Research Policy*. Springer. [Chapter 11]

Sivertsen, G. (2023). Performance-based research funding and its impacts on research organizations. In: Lepori B., Jongbloed B., and Hicks D. (Eds). Handbook of Public Research Funding, Edward Elgar Publishing, pp. 90-107.

Solberg, E., Ramberg, I. and Sandven, T. (2018) The role of research institutes in national R&D systems. A comparative overview. NIFU Working Paper 2018:9

Stampfer, M. (2019). Pros and Cons of Distribution Streams of Governmental Funding for Research. Formas, Stockholm

Zacharewicz, T.; N. P. Pavón, L. A. P. Martos & B. Lepori (2023). Do funding modes matter? A multilevel analysis of funding allocation mechanism on university research performance? Research Evaluation 32: 545–556.

Chapter 11
Designing Performance-Based Research Funding Systems

Gunnar Sivertsen and Kaare Aagaard

Abstract Performance-based research funding systems (PBFS) allocate direct institutional funding to universities and other public research organisations based on an assessment of their research. This chapter discusses their possible effects and how they could be designed to support academic value creation and good research cultures.

11.1 Three Main Types

Performance-based research funding systems (PBFS) for research performing organisations add an element of competition to direct institutional funding which comes in addition to the contest for indirect external project funding awarded by research councils and other funding organisations (Hicks, 2012). The systems appear in three main types: evaluation-based funding, indicator-based funding, and funding contingent on performance agreements. Combinations are possible (Sivertsen, 2023). Examples:

- Evaluation-based systems: Czech Republic, France, Italy, Latvia, New Zealand, Portugal, and the United Kingdom
- Indicator-based systems: Belgium, Croatia, Denmark, Estonia, Finland, Germany (at the level of federal states), Norway, Poland, Slovakia, and Sweden
- Systems based on performance agreements: Austria, Denmark, Estonia, Finland, Netherlands, and Switzerland

G. Sivertsen (✉)
Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway
e-mail: gunnar.sivertsen@nifu.no

K. Aagaard
Aarhus University, Aarhus, Denmark
e-mail: kaag@au.dk

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_11

Based on experiences with the designs, implementations, developments, and discussions of PBFS in 26 countries, Sivertsen (2023) explains why the systems differ within each of the three main types:

> Although some systems may seem similar across countries, they are never the same and they are modified all the time. PBFS differ because they are anchored in the local traditions and mechanisms of state budgeting and embedded in the local negotiations about priorities and developments in the higher education sector. They are dynamic because they are continuously contested and thereby often adjusted. Countries also mutually learn from each other and inspire changes in their PBFS. The systems are conservative as well. Once implemented, they become games with rules and specific terminologies and infrastructures that are difficult to change. Also, they need to be predictable because they influence budgets and the spending of tax revenues on the funding side. There is need to ensure some stability of budgets at the institutions.

Discussions of the pros and cons of different types and designs of PBFS often take the perspective of best practice in *research evaluation*. However, for the purpose of *funding* and from the perspective of *methods*, evaluation-based PBFS is only one of the three main types mentioned above. Evaluation as a method for PBFS was first introduced in the United Kingdom in 1986. The purpose was clearly *funding allocation:* to avoid cutbacks in the places that least deserved it at a time when Prime Minister Margaret Thatcher pursued to minimise public spending (Martin and Whitley, 2010). Well-established methods for peer review in *competitive grant allocation* were borrowed to introduce fairness and protect excellence. Systematically and in recurring consultation with the universities, these methods have been further developed to serve what is now named the Research Excellence Framework (REF). The front page of the REF in 2021 showed that the original allocation method has become the main purpose: The REF is "the system for assessing the quality of research in UK higher education institutions". With evaluation as the main purpose, peer review may be considered preferable over metrics, as in the recommendation for the REF stated by *The Metric Tide* report (Wilsdon et al., 2015):

> Metrics should support, not supplant, expert judgement. Peer review is not perfect, but it is the least worst form of academic governance we have, and should remain the primary basis for assessing research papers, proposals and individuals, and for national assessment exercises like the REF.

This statement is relevant for the design of a research assessment system but not necessarily for the design of an institutional funding system. As we shall see, the intentions behind a research assessment system may be compromised if strongly connected to allocation of funding and reputation. However, the other two main types of PBFS also have problems when applied with strong effects on funding and reputation—see Sect. 11.4.

11.2 Historical Background and Motivation

PBFS emerged with the era of New Public Management (NPM) and are motivated by some of the core ideas: less central steering combined with required accountability, performance and outcome measurement, and increased efficiency through competition. If we add the aim of increasing research quality, the core ideas of NPM are often reflected in official motivations for PBFS. As an example, the official motivation for PBFS at the European level was to establish "more effective national research systems – including increased competition within national borders and sustained or greater investment in research", as expressed by the European Commission (2012) in a communication with guidelines for *A Reinforced European Research Area Partnership for Excellence and Growth*.

In many countries, PBFS evidently serve more purposes than the core NPM idea of efficiency through competition. The systems create statistical overview and insight into the research activities, they aim to make the funding allocation criteria fairer and more transparent, and they may reinforce the willingness of governments to sustain or increase funding of the higher education sector. The ages before NPM and PBFS were different. Higher education was for the few and research was performed by the elite with close relations to government. Lobbyism, not equity, was the funding mechanism. NPM now allows for connecting funding to explicit ambitions and targets for large public investments in research. Against this background, it is understandable that several countries have seen their parliaments unanimously agree on continuing their PBFS despite critical voices from academia. PBFS are by many policymakers and administrators perceived as the preferable way of funding and steering an otherwise autonomous sector of society.

If PBFS are to remain in place, a way forward could be to replace the NPM perspective with Public Value Management (Stoker, 2006) as the governance model. A typical NPM aim such as "more effective national research systems" in the above cited example could then be translated to for example "more support for the characteristic and unique forms of societal value creation at universities". The aim of "increased competition within national borders" could be replaced by "increased collaboration among universities and with society in creating these values". This change would reflect new notions of research quality as discussed in other chapters in this book.

11.3 The Effects of PBFS Are Not Easy to Determine

For indicator-based PBFS and the use of indicators in performance agreements, the paradox of performance-based funding is often referred to as "Goodhart's law": When a measure becomes a target, it ceases to be a good measure (Goodhart, 1975). Although this sentence seems reasonable as an expression of what happens when indicators are introduced in the context of a PBFS, Goodhart's law has not been

confirmed in studies of effects of PBFS. In an early overview, Butler (2010) noted that the literature on possible effects "is full of words like 'likely', 'potential', and 'possible' without much evidence". The most complete and recent comparative overview of PBFS in Europe (Zacharewicz et al., 2019) finds that there is an "absence of an assessment of the impact of the different types of performance-based funding systems". In a systematic review, Thomas et al. (2020) observe that one-fourth of the literature on PBFS comprises studies about their possible effects. Taken together, they are inconclusive.

The reason for this continued absence of conclusive evidence of the effects is probably the complexity of the task (Auranen and Nieminen, 2010; Hicks, 2012; Hammarfelt and de Rijcke, 2015; de Rijcke et al., 2016; Gläser and Laudel, 2016; Aagaard and Schneider, 2017). As indicated above, PBFS differ among and constantly change (Sivertsen, 2023), and it is difficult to isolate the incentives and value systems from other influences. Observable changes in performance may be due to other funding and evaluation arrangements such as competitive external funding.

11.4 Designing Better Systems

Although the general effects of PBFS are difficult to determine, as explained above, some ideas for how they can be designed to better serve Public Value Management (see Sect. 11.2) can be drawn from experiences with the designs, implementations, developments, and discussions in different countries (Sivertsen, 2023). The legitimacy and functioning of a system seem to depend on these three factors: (1) the type and design of the system; (2) the strength of its influence on funding and reputation; and (3) the involvement of the funded organisations in collaboration about the design, implementation, management, and evaluation of the system.

The type and design of the PBFS are often discussed as the most important factors, e.g., in the debate in the United Kingdom about the evaluation-based versus the indicator-based alternatives (Taylor, 2011; Geuna and Piolatto, 2016; Wilsdon et al., 2015; Sivertsen, 2017; Harzing, 2018). However, the experience is that all three major types of PBFS have pros and cons. The preference among them depends on whether the PBFS is expected to serve other purposes in addition to funding allocation. Other factors than the type and design seem to be more important for the legitimacy and functioning of a PBFS.

The second factor, the strength of the influence on funding and reputation, seems very important. The validity of the methods for evaluation may be reduced because evaluees and evaluators fear negative outcomes. Qualitative information about strengths and weaknesses can become vague, generally positive and less reliable. Documentation and statistics may become selective, and indicators can turn into displaced goals. The paradox is that, to reach the aims of a PBFS more effectively, their effects on funding and reputation should be minimised.

The examples of Australia and of Uppsala University in Sweden may enlighten the problem. The Australian periodical research evaluation exercise, Excellence in

Research Australia (ERA), ran as a PBFS in two rounds until 2012, after which it was disconnected from university funding in the third round in 2018. The rating of universities now only influenced their reputation. The average rating continued to increase from round to round (Australian Research Council, 2021). The example shows that evaluators care about possible negative consequences of the exercise and that the consequences for recognition can be just as important as funding.

Uppsala is interesting in the same contexts although it is not about PBFS. The university ran self-initiated evaluations with external peer review in 2007 and 2011, both with effects on internal resource allocation. The average rating of departments increased between the two rounds, but they were not of much use as advice for further development. Then, in 2017, Uppsala University (2017) changed the game completely and introduced formative organisational evaluation without resource allocation. The outcome was now a much larger number of specific and consequential recommendations for organisational progress than the earlier evaluations had provided.

The third factor that may influence the legitimacy and functioning of a PBFS is the involvement of the funded organisations in collaboration about the design, implementation, management, and evaluation of the system. This factor is easy to neglect since funding is the sole responsibility of governments on behalf of society. Why should the funded organisations be involved at all in this political task? Experience shows that it is favourable to collaborate with them because the outcome of a PBFS also depends on the degree of alignment between the external expectations and financial incentives that are built into the PBFS at the system level and the internal values, aims, and incentive structures that the funded organisations wish to sustain to provide public value.

Policy Implications

- Many countries have implemented performance-based research funding systems (PBFS) as a way of funding and steering an otherwise autonomous sector of society. These systems create statistical overview and insight into research activities, they have the potential to make the funding allocation criteria fairer and more transparent as an alternative to lobbyism, and they may reinforce the willingness of governments to sustain or increase funding of the higher education sector. However, they are also often met with resistance from the academic community and may have unintended constitutive effects.
- The type and design of a PBFS is important, but there is no ideal solution that solves the problems. All three major types have pros and cons: evaluation-based, indicator-based, and funding contingent on performance agreements. How they are designed is the more important question, and any preference of type depends on whether the PBFS is expected to serve other purposes in addition to funding allocation.

- These systems may become compromises between aims and methods, which is the paradox of performance-based funding. The validity of the methods, and thereby the usefulness of PBFS, may be reduced when performance assessment, measurement, or agreement is connected to funding. Minimising their effects on funding and reputation may support the achievement of their aims, particularly if formative learning is one of them.
- Governance is important. The usual focus in New Public Management on comparability and competition to achieve efficiency might be counterproductive. The funded organisations should be involved in collaboration about the design, implementation, management, and evaluation of the system. External expectations and financial incentives need to be aligned with characteristic forms of public value creation at universities.

References for Further Reading

Aagaard, K., and Schneider, J.W. (2017). Some considerations about causes and effects in studies of performance-based research funding systems. Journal of Informetrics, 11(3), 923–926.

Auranen, O., and Nieminen, M. (2010). University research funding and publication performance—An international comparison. Research Policy, 39(6), 822–834.

Australian Research Council (2021). ERA EI Review Final Report 2020–2021. Canberra: Commonwealth of Australia.

Butler, L. (2010). Impacts of performance-based research funding systems: A review of the concerns and the evidence. In: Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings. OECD Publishing.

de Rijcke, S., Wouters, P.F., Rushforth, A.D., Franssen, T.P., and Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—A literature review. Research Evaluation, 25(2), 161–169.

European Commission (2012). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A Reinforced European Research Area Partnership for Excellence and Growth.

Geuna A., and Piolatto M. (2016). Research assessment in the UK and Italy: Costly and difficult, but probably worth it (at least for a while). Research Policy, 45(1), 260–271.

Gläser, J., and Laudel, G. (2016). Governing science: How science policy shapes research content. European Journal of Sociology, 57(1), 117–168.

Goodhart, C. (1975). Problems of monetary management: The U.K. experience. Papers in Monetary Economics, 1. Sydney: Reserve Bank of Australia.

Hammarfelt, B., and de Rijcke, S. (2015). Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. Research Evaluation, 24(1), 63–77.

Harzing, A.W. (2018). Running the REF on a rainy Sunday afternoon: Can we exchange peer review for metrics? STI 2018 Conference Proceedings, pp. 339–345.

Hicks, D. (2012). Performance-based university research funding systems. Research Policy, 41(2), 251–261.

Martin, B., and Whitley, R. (2010). The UK Research Assessment Exercise: A Case of Regulatory Capture. In: Whitley, R., Gläser, J., and Engwall, L. (Eds.) Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation, Oxford Academic, pp. 51-80.

Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. Palgrave Communications, 3, 17078.
Sivertsen, G. (2023). Performance-based research funding and its impacts on research organizations. In: Lepori B., Jongbloed B., and Hicks D. (Eds). Handbook of Public Research Funding, Edward Elgar Publishing, pp. 90-107.
Stoker, G. (2006). Public value management: A new narrative for networked governance? The American Review of Public Administration, 36, 41–57.
Taylor, J. (2011). The assessment of research quality in UK universities: Peer review or metrics? British Journal of Management, 22(2), 202–217.
Thomas, D.A., Nedeva, M., Tirado, M.M., and Jacob, M. (2020). Changing research on research valuation: A critical literature review to revisit the agenda. Research Evaluation, 29(3), 275–288.
Uppsala University (2017). Quality and renewal 2017. Research environment evaluation at Uppsala University. https://uu.diva-portal.org/smash/get/diva2:1153914/FULLTEXT01.pdf.
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, B. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. 10.13140/RG.2.1.4929.1363.
Zacharewicz, T., Lepori, B., Reale, E., and Jonkers, K. (2019). Performance-based research funding in EU Member States—A comparative assessment. Science and Public Policy, 46(1), 2019, 105–115.

Chapter 12
The Limits of Universities' Strategic Capacity for Steering Research

Siri B. Borlaug and Jens Jungblut

Abstract Today's universities are under pressure from multiple directions. The massive growth in students and staff has led to, among others, a high competition between individual academics and universities for external funding. Studies highlight that the quest for external funding has led to a narrowing down of research topics in terms of breadth and scope, and some call for universities to more actively use their strategic capacity and leadership to create environments that also provide scholars with incentives for scholarly renewal, innovation, and research beyond what is in high demand from external funders. However, the definition of strategic capacity, challenges related to it, and how universities and their leadership may use it are not always clearly defined. This chapter addresses these issues.

12.1 A Changing Governance Scene

After decades of developing steering and accountability relationships between universities and government, we now observe a switch in the rhetoric of governance of universities regarding research. There is, on the one hand, a growing recognition that universities have limited internal steering capacity towards their subunits and academics, as research priorities have been outsourced to external funders. On the other hand, there is a movement towards reforming research assessment where qualitative assessments and evaluations get increasingly more room (Sivertsen and Rushforth 2024). These movements also signal that university leadership could and should get a more prominent and important role in steering research activities.

S. B. Borlaug (✉)
Oslo Business School, Oslo Metropolitan University, Oslo, Norway

Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway
e-mail: siri.borlaug@oslomet.no

J. Jungblut
Department of Political Science at University of Oslo, Oslo, Norway
e-mail: jungblut@stv.uio.no

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_12

However, *studies of the research system and the organisational characteristics of universities raise questions regarding the universities' capabilities to actively employ strategic approaches in terms of research priorities*. In the following, we will explain why the strategic capacity of universities and their leadership to steer the work in their institutions is more limited regarding research work, and what this means for designing successful governance arrangements between the state and the higher education sector.

12.2 What Is Strategic Capacity?

We may talk about strategic capacity on different levels. Firstly, university governance takes place at the intersection of higher education policies, public sector regulations, and academic norms or traditions (Musselin 2021). In this, *strategic capacity of universities is linked to having a certain degree of autonomy* regarding both procedural and substantive matters (Berdahl 1990). Today, the level of autonomy of universities in Europe varies between countries as legal frameworks determine the extent to which universities can make their own decisions and pursue their own strategic initiatives (Pruvot et al. 2023). At the same time, several studies highlight that there is a difference between formal autonomy and the lived or "real" autonomy and that most of the time *the lived autonomy is more limited than the formal one* (Christensen 2011; Maassen et al. 2017). The reason for this difference is manifold, but a part of it is driven by the fact that universities are granted organisational autonomy, while the state employs accountability measures such as competitions for performance-based funding to steer the universities indirectly (Degn and Sørensen 2015).

Second, *for universities, the term strategic capacity describes how an institution lines up its internal subunits like departments, faculties, centres*, etc.*, to achieve common goals* (Thoenig and Paradeise 2016). This presupposes a tighter coupling within the organisation—more "complete" organisations – which indeed is a challenging task given that different departments and subunits can have diverging missions and interests. While universities' strategies set directions and priorities for the organisation, the effectiveness of these strategies depends on internal relations and interactions between units and levels, and how performance related to the strategies is assessed and valued. Some even argue that strengthened hierarchical governance is driving increased organisational specialisation and professionalisation and that this results in fragmentation within universities with stronger faculties, which in turn makes it even harder to implement coherent strategic action (Maassen and Stensaker 2019).

12.3 Steering Loosely Coupled Organisations

One of the main challenges for steering universities is that they are prime examples of *loosely coupled organisations, in which subunits as well as formal structures and activities often are independent of one another, making rational and coherent action less likely* (Elken and Vukasovic 2019; Orton and Weick 1990). While this does not mean that subunits of an organisation are non-responsive to external signals, it underscores that universities are difficult to steer in a rational, top-down manner (Bleiklie et al. 2015). Due to these characteristics, steering in universities, by necessity, must combine top-down control with other steering approaches.

In balancing these aspects, universities must consider, on the one hand, standardisation, professionalisation, and specialisation of organisational management and administrative functions and, on the other hand, safeguard organisational flexibility, adaptability, and integrative capacity needed to enhance productivity and effectiveness of teaching and research (Maassen et al. 2017; Leisyte et al. 2009; Maassen and Stensaker 2019).

This trade-off plays out differently in teaching and research. In most universities, education is much more administratively regulated with the aim to ensure a reliable and comparable provision. In addition, national governments use quality assurance systems to influence procedures within universities to assess and control the provision of education. *This gives university leadership more power to steer their subunits and control the framework in which academics perform their teaching—a connection that is much looser in research* which depends more on the creativity of individual researchers and where disciplinary differences create fundamentally different ways in which research is conducted.

There is a growing recognition in the literature that universities have limited steering capacity both towards their subunits and academics but also with regard to their responses to external demands (e.g. Maassen and Stensaker 2019; Maassen et al. 2017). While rhetorically many university governance reforms have strengthened the role of the leadership, historically grown democratic elements in university governance remain important leading to specific local mixtures between the two approaches (de Boer and Maassen 2020). This mixture varies between and sometimes even within countries. One example here is hired versus elected university leadership.

12.4 External Limitations to Strategic Steering Capacity

Another reason for universities' limited steering capacity is related to consequences of *national, performance-based evaluation systems.* Notably, such systems have had positive effects. They have among others contributed to information about research activities which have been important in terms of transparency and for steering, and they have incentivised especially very low performing units to increase

their output (Aagaard et al. 2015). However, studies show that they also limit the universities' room to manoeuvre. Some highlight that universities' actions are guided by the desire to fulfil indicators and targets of evaluation systems (Thomas et al. 2020; Musselin 2021). Given that many countries have reduced the block grant funding of universities and increased competitive funding, this further strengthens the impact of performance-based systems (Hicks 2012).

Together this has had several, perhaps unintended, effects on different levels. For universities, this has implied an increased competition over resources and reputation and some universities have even employed strategies of gaming indicators to boost their financial returns (e.g. in the United Kingdom). This problem also holds for the level of individual academics as researchers are also not passive recipients of indicators but rather actively engaged with them potentially leading to goal displacement (de Rijcke et al. 2016). This might entail pursuing projects that are deemed relevant by the funders instead of following one's own scientific curiosity.

These strategies are an example of the de-coupling between indicator-based steering and resulting quality of research. For example, investigating the Netherlands, Sweden, and the United Kingdom, Teelken (2015) shows how performance management altered individual behaviour of academics. She finds for all three countries an increase in formalisation and more focus on scientific publication output, but also a less explicit relationship with the actual quality and content of research.

Externally funded research grants are also used by universities as signals of both prestige and research quality, and they are therefore important for the external perception and branding of universities. For some universities, this is additionally enhanced by a focus on performance in international rankings, while others—especially smaller and more teaching-oriented institutions—are instead under pressure to fulfil local or regional needs. These diversified missions are not always properly reflected in national funding competitions or indicators used to distribute funding. Thus, the need to respond to such indicators can further decrease the room to manoeuvre for universities and their leadership.

The increased focus on external competitive research funding further implies that the number of submitted proposals to funding agencies has increased, and for some attractive grants the success rate is now below 10% (Langfeldt 2024). Studies also show that it is *often the same researchers that are granted external funding from different funding sources, leading to a concentration in selected researchers or research groups and their respective topics* (Aagaard et al. 2020). As external funding usually also comes with additional academic positions, one consequence is that *the choice of what to research and whom to recruit is often outsourced from university leadership or even the leadership of subunits to external decision-makers and peer reviewers of grant proposals* (Whitley et al. 2018). This may lead to that national and local needs may not get sufficient attention unless it is specifically prioritised by funding agencies.

Another challenging trend is that *external grants often drive a scaling up in personnel through hiring temporary researchers*. These researchers are recruited based on criteria of the research project and may not match the need of the department or the unit, in terms of competence required for teaching, which in turn makes it harder

for these researchers to find permanent positions once projects end (Borlaug et al. 2019). External funding thus contributes to move the authority to shape the research agenda and hiring policy from leaders of departments or faculties to principal investigators of external grants (Edler et al. 2014; Kondakci and Van den Broeck 2009).

Given the effects that external funding has on university leadership's room to manoeuvre, *several scholars underline universities' strategic capacity and wiggle room as an area that needs more attention* (Musselin 2021; Mignot-Gérard et al. 2023; Franssen et al. 2023). This has emerged in particular in light of *challenges to uphold research areas that are not fulfilling standardised criteria of quality or excellence* such as being successful in acquiring external funding and publishing in highly ranked international scientific journals. Even without ticking these boxes, academic work in these contexts can still be relevant and necessary for society. A more diverse set of evaluation criteria would help to protect these environments, for example by using more peer review-based evaluations and including societal interaction assessments.

12.5 Different Conditions for Strategic Capacity

Not all universities can employ strategies with the same effectiveness. Studies show that *differences in organisational resources and reputation influence the universities' strategic capacities* (Thoenig and Paradeise 2016). Well-off institutions have the capacity to support internal strategic initiatives, while those who struggle more financially mainly follow external priorities set by funding agencies or government ministries. Universities with high strategic capacity are often highly commercial universities like Stanford or Oxford which also have income from, e.g., donations or endowments. This puts them in another situation than for instance purely public universities which in many cases are more dependent on developments in their respective national higher education system (Whitley 2008). The resource situation also varies between universities within one country. Large (and old) universities have often more strategic capacity compared to small and young universities, as the latter often have less resources and are more dependent upon national or even local / regional developments.

The *strategic capacity also varies within a university*. Pfeffer and Salancik (1974) studying budget negotiations between departments show that in decision-making both objective, bureaucratic criteria, like the number of students, and political criteria, like the internal power of the department, matter for budget allocation. Thus, there is often an interconnection as *wealthy departments with more external funding have a better reputation and therefore also more power internally to influence strategies.* Additionally, their relative strong performance may act as a buffer against external pressures from higher-up in the organisational hierarchy (Mignot-Gérard et al. 2023) making it harder to enact strategies against their will.

12.6 What Means Do Universities and Their Subunits Have for Enacting Strategies?

Despite these challenges and imbalances, universities, subunits, and their leadership are not powerless. They can use different tools to support the implementation of (research) strategies even within their limitations.

Evaluations are one such tool, as they can be used to provide arguments and external validation for change processes. External and internal evaluations of research, study programmes, or units may aid leaders and middle managers in setting priorities by providing legitimacy for their interventions, disrupting existing equilibria, or introducing new ideas or ways of seeing things. However, these processes also must strike a balance between disruption and appropriateness as newly introduced ideas and suggestions have to be perceived as legitimate and suitable to a given environment to be fully embraced by it (March and Olsen 2011).

Lately, there have been national and international initiatives to *move from summative and narrow evaluations reflecting past performance of units to more formative evaluations*, which to a larger extent include a broader set of goals and take a more forward-oriented look at units. For example, several research institutions have signed the CoARRA initiative (see Chap. 7)—the Agreement of Reforming Research Assessment, in which they commit to ensure that their research assessments will recognise and reward the plurality of contributions researchers make, respect epistemic differences, and reward open science, research integrity, and societal relevance.

There are also several examples of different forms of formative evaluations. In Norway, development agreements between the ministry and the universities have come to play a central role in the governance of universities. Many universities apply the goals in the agreements to legitimise internal priorities and some even apply them in their own internal steering towards faculties, departments, or centres (Elken and Borlaug 2023).

Another example is the so-called strategy evaluation protocol—SEP in the Netherlands. This assessment is explicitly formative and aimed at learning. The evaluation is done by peers and conducted at the level of organisational subunits (e.g. departments or faculties), rather than the university as a whole (Franssen et al. 2023). An important feature of the SEP is that there is no funding attached to the evaluation. As such it is up to the university and the units to decide how they use the information gained, which gives them flexibility to link it to their respective strategies.

The emphasis on learning and development seems to be a fruitful way forward for ensuring quality and variety. On the subunit level, peer-reviewed evaluations may provide a leeway for middle managers to engage in so-called sense-giving and sensemaking processes. Here *evaluations and similar exercises may be a legitimising device for setting research priorities.* Thus, they can be used to strengthen the implementation of strategies even in the absence of hierarchical steering (Degn 2018; Franssen et al. 2023).

Given the ongoing interest in and discussions about universities' internal authority and strategic capacity, these insights are important. They show that, *while being a powerful steering tool, funding is not the only and maybe sometimes even the wrong means to implement strategies and achieve change in research and research practices.* Moreover, the appropriateness of strategic interventions and steering is important to ensure not only support from academics but ultimately also the effectiveness. This in turn highlights that a too strong focus on hierarchical steering will encounter problems in cases where interventions are not perceived as appropriate. Moreover, especially research is an activity that is hard to steer in a hierarchical manner as it depends on inspiration, innovation, and knowledge of those actively working with it.

Having said this, funding remains a key tool for steering especially in situations where governance is more indirect and where setting incentives is the most common practice. This holds true both for the relation between governments and universities and for the relation between universities and their subunits.

To increase the strategic capacity of universities and delimit the many challenges following external funding, such as the rise of temporary positions, we observe a tendency in some countries to argue for a switch in funding streams, which is increasing block grant funding while reducing external competitive funding. While such a switch can empower universities it may also create challenges. For instance, external grants have several important functions in academia: They in part support (if that is their aim) inter- and transdisciplinary research (Lyall et al. 2013), they facilitate national and international research collaboration, and they can help to concentrate resources necessary for larger investigations. While the latter, as we have noted above, can also be problematic, it is still an important function especially in those disciplines that depend on large teams or expensive infrastructures.

While strategic capacity is often discussed on the level of university leadership, it is important to also direct attention to the middle level including faculties, departments, or centres as their lived autonomy and steering capacity vary considerably given that universities are loosely coupled organisations. This creates a diverse ecology of subunits all with their own interests and priorities, which in turn makes it harder for university leadership to implement detailed and uniform strategies. A one-size-fits-all approach will not be successful when trying to develop and implement strategies for a diverse set of subunits. A successful implementation of any strategy depends on the cooperation of the affected subunits and academics, their perception of and involvement in the strategy, and its process of creating matter.

Policy Implications

- Universities are loosely coupled organisations, and the strategic capacity of universities and their leadership with regard to research is in general more limited than in education.
- Given the greater dependence on subunits and individual academics as well as the diversity among them, the leadership have to find ways to create acceptance for strategies and support for their implementation within the organisation.

- Using funding as incentive is one way of doing that. However, the ability to use this strategically can be limited by the incentive system that the government uses towards the universities. Not all universities will have sufficient resources to use internal incentive structures.
- Formative evaluations are an important tool that can help support the implementation of strategies as peer review is a well-established procedure in academia and the results of evaluations can be used to create justification and support for the implementation of strategies.

References for Further Reading

Aagaard, K., Bloch, C. & Schneider, J.W. (2015) Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator, *Research Evaluation,* 24(1).

Aagaard, K., Kladakis, A., & Nielsen, M. W. (2020). Concentration or dispersal of research funding? Quantitative Science Studies, 1(1), 117-149.

Berdahl, R. (1990). Academic freedom, autonomy and accountability in British universities. Studies in Higher Education, 15(2), 169-180. https://doi.org/10.1080/03075079012331377491

Bleiklie, I., Enders, J., & Lepori, B. (2015). Organizations as Penetrated Hierarchies: Environmental Pressures and Control in Professional Organizations. Organization Studies, 36(7), 873-896.

de Boer, H., & Maassen, P. (2020). University governance and leadership in Continental Northwestern Europe. Studies in Higher Education, 45(10), 2045-2053. https://doi.org/10.1080/03075079.2020.1823640

Borlaug, S. B., Langfeldt, L., Gulbrandsen, M., & Ramberg, I. (2019). Impacts of the SFF scheme on the Norwegian research system: Sub-report I to the SFF evaluation panel.

Degn, L., & Sørensen, M. P. (2015). From collegial governance to conduct of conduct: Danish universities set free in the service of the state. Higher Education, 69, 931-946.

Degn, L. (2018). Academic sensemaking and behavioural responses–exploring how academics perceive and respond to identity threats in times of turmoil. Studies in Higher Education, 43(2), 305-321.

Edler, J., Frischer, D., Glanz, M., & Stampfer, M. (2014). Funding individuals—changing Organisations: The impact of the ERC on universities Organizational Transformation and Scientific Change: The Impact of Institutional References FrRestructuring on Universities and Intellectual Innovation (pp. 77-109): Emerald

Elken, M., & Vukasovic, M. (2019). The Looseness of Loose Coupling: The Use and Misuse of "Loose Coupling" in Higher Education Research. Theory and Method in Higher Education Research, 5, 53-71. https://doi.org/10.1108/S2056-375220190000005005

Elken, M., & Borlaug, S. B. (2023). Implementation of ambiguous governance instruments in higher education. Higher Education, 1-16.

Franssen, T., Borlaug, S. B., & Hylmö, A. (2023). Steering the direction of research through organizational identity formation. Minerva, 61(4), 495-519.

Hicks, D. (2012). Performance-based university research funding systems. Research policy, 41(2), 251-261.

Kondakci, Y., & Van den Broeck, H. (2009). Institutional imperatives versus emergent dynamics: a case study on continuous change in higher education. Higher Education, 58(4), 439-464.

Langfeldt, L, (2024) The many dilemmas of grant peer review. In G. Sivertsen and L. Langfeldt (eds). *Challenges in Research Policy*. Springer. [Chapter 5]

Leisyte, L., Enders, J., & de Boer, H. (2009). The balance between teaching and research in Dutch and English universities in the context of university governance reforms. Higher Education, 58(5), 619-635.

Lyall, C., Bruce, A., Marsden, W., & Meagher, L. (2013). The role of funding agencies in creating interdisciplinary knowledge. Science and Public Policy, 40(1), 62-71.

March, J. G., & Olsen, J. P. (2011). The Logic of Appropriateness. In R. Goodin (Ed.), The Oxford Handbook of Political Science (pp. 478–497). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199604456.013.0024

Mignot-Gérard, S., Sponem, S., Chatelain-Ponroy, S., & Musselin, C. (2023). Kaleidoscopic collegiality and the use of performance research metrics. The case of French universities. Higher Education, 85(4), 887-918.

Musselin, C. (2021). University governance in meso and macro perspectives. Annual review of sociology, 47, 305-325.

Maassen, P., Gornitzka, Å., & Fumasoli, T. (2017). University reform and institutional autonomy: A framework for analysing the living autonomy. Higher Education Quarterly, 71(3), 239-250. https://doi.org/10.1111/hequ.12129

Maassen, P., & Stensaker, B. (2019). From organised anarchy to de-coupled bureaucracy: The transformation of university organisation. Higher Education Quarterly, 73, 456–468. https://doi.org/10.1111/hequ.12229

Orton, J. D., & Weick, K. E. (1990). Loosely Coupled Systems: A Reconceptualization. Academy of Management Review, 15(2), 203-223.

Pfeffer, J., & Salancik, G. R. (1974). Organizational Decision Making as a Political Process: The Case of a University Budget. Administrative Science Quarterly, 19(2), 135-151. https://doi.org/10.2307/2393885

Pruvot, E. B., Estermann, T., & Popkhadze, N. (2023). University Autonomy in Europe IV. The Scorecard 2023. Retrieved from: https://eua.eu/downloads/publications/eua%20autonomy%20scorecard.pdf

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

Sivertsen, G. & Rushforth, A. (2024) The ongoing reform of research assessment. In G. Sivertsen and L. Langfeldt (eds). Challenges in Research Policy. Springer. [Chapter 7]

Teelken, C. (2015). Hybridity, coping mechanisms, and academic performance management: Comparing three countries. Public Administration, 93(2), 307-323

Thoenig, J. C., & Paradeise, C. (2016). Strategic capacity and organisational capabilities: A challenge for universities. Minerva, 54, 293-324.

Thomas, D. A., Nedeva, M., Tirado, M. M., & Jacob, M. (2020). Changing research on research evaluation: A critical literature review to revisit the agenda. Research Evaluation, 29(3), 275-288.

Whitley, R. (2008). Construction universities as strategic actors: Limitations and variations (No. 557). Manchester Business School Working Paper.

Whitley, R., Gläser, J., & Laudel, G. (2018). The impact of changing funding and authority relationships on scientific innovations. Minerva, 56, 109-134.

Chapter 13
The Consequences of Paying to Publish

Thed van Leeuwen and André Brasil

Abstract Open access publishing has been the most prolific aspect of the transition towards open science. In this transition, increasingly national governments, national and international funding agencies, and institutional leadership have initiated policies to promote and stimulate the development to open access as the norm in scholarly publishing. However, this has not always led to the best outcomes.

13.1 Introduction

In the last part of the previous century, the world witnessed the rise of a movement aiming to organise access towards scientific knowledge in a more egalitarian manner. The initial aims of the open science movement, as it was coined, revolved around ensuring fair access to scientific, or scholarly, literature (Fecher and Friesike 2013). Exorbitant prices are often practised for this type of literature, making access impossible for many scholars in low- and middle-income countries.

Originating primarily from university librarians, this movement gained substantial momentum through subsequent engagement from research funders and science policymakers. This progression led to the design of national open access mandates, institutional open access policies, and specific requirements within research grants. This chapter investigates the effectiveness of some of these policies, with a particular focus on paradoxical outcomes unforeseen during their development and implementation.

T. van Leeuwen (✉) · A. Brasil
Centre for Science and Technology Studies (CWTS) at Leiden University, Leiden, Netherlands
e-mail: leeuwen@cwts.leidenuniv.nl; a.l.brasil@cwts.leidenuniv.nl

G. Sivertsen, L. Langfeldt (eds.), *Challenges in Research Policy*, SpringerBriefs in Political Science, https://doi.org/10.1007/978-3-031-69580-3_13

13.2 Policies

As countries started issuing national policies and mandates on open science, different approaches were taken to open access for scholarly publishing. Some policies induced publishing in journals that do not charge for access to the papers they publish, in what is referred to as Gold Open Access. Such openness is often made possible through the payment of often expensive Article Processing Charges (APCs) by authors. That was the road of choice for countries such as the United Kingdom and the Netherlands. Other countries, like Denmark, chose a different path, encouraging researchers to deposit a version of their manuscripts in a freely accessible online repository, such as an institutional repository or a subject-specific archive. National perspectives often resulted in likewise policies on the institutional level, although these often also covered other aspects of open science, such as open data and open source code.

In 2018, a consortium of international funders, known as cOAlition S, launched a plan outlining how to publish in open access (OA) format (cOAlition S 2018). The so-called Plan S prioritised the Gold Open Access model over other types of OA, particularly over what is known as Hybrid Open Access. While in the Gold model all the journal's content is freely accessible to anyone interested without any costs or restrictions, Hybrid OA journals continue to charge for subscriptions, and they make only a share of the articles they publish open, through APC payments. The Hybrid model, designed to be a temporary resource to aid publishers in the transition to the Gold preferred standard, is still widespread 5 years after the launch of Plan S, and publishers continue to profit from both the subscription income and APC payments for many papers submitted to their Hybrid journals.

While Plan S has proposed to use a cap, a maximum cost acceptable for one single open access publication, it is still up to the publisher, often owners of highly prolific internationally oriented journals, to determine the value of APCs charged for either Gold or Hybrid OA publications. *Thus, the scholarly publishing system is moving from a pay-to-read to a very expensive pay-to-publish model, sometimes both.* As a result, inclusion in scientific publishing remains limited to those who can afford the persisting high costs.

13.3 Studies

A recent study (Brasil et al. 2022) on open science policies and the phenomenon of APCs shows that the national open science policy of the Netherlands clashes with the actual development initiated by the Dutch universities. In 2013/2014, the national government issued an open access policy prioritising Gold as the default open access format. However, universities have a high degree of autonomy in the country, being well organised in contexts such as university library settings, evaluation cycles, and overall governance of the institutions. Based on such autonomy,

Dutch universities, together with the royal library in the so-called UKB, negotiated with the publishing industry specific deals that allowed the country's scholars to publish in journals within the subscriptions with these publishers. These deals are known as transformative agreements, and their development in the country started in 2016/2017, favouring publication in Hybrid Open Access format. The two contradictory initiatives have influenced publishing practices of scholars from various scholarly domains, and they have also sparked discussions about the benefits of the Hybrid and the Gold models.

One of the common elements between the two OA formats is that articles published are open, freely accessible without any paywalls. However, publishing is not free, and often authors pay article processing charges (APCs) to publish. These APCs were the object of a second study (Brasil and van Leeuwen 2022), which investigates Gold OA publishing at a country level from a global comparative perspective. The study shows that, while the original motivation for the open science movement was better access for the scholars of low- and middle-income countries to reading scholarly literature from the Northern hemisphere, the current development of Gold Open Access publishing is driving into the direct opposite direction.

Scholarly publishing in Gold Open Access journals via the payments of APCs is becoming more and more expensive. A direct translation of numbers of publications times APC rates shows that publishing becomes more costly, but when normalisation for national welfare situation is conducted, by applying the OECD-based PPP index (purchasing power parity), the situation even worsens (Brasil and van Leeuwen 2022). In short, while the Gold OA contributed to making more scholarly literature available for scholars from the low- and middle-income countries, the high APCs are creating significant challenges for them to publish their work.

For instance, Figs. 13.1 and 13.2 contrast the spending of countries on Gold OA publishing with the impact measured through the mean normalised citation score (mncs). The difference between the visualisations is that the first figure shows spending based on a nominal conversion of the average amount of money spent on APCs to US Dollars. Figure 13.2 improves that perspective by normalising spending according to the PPP index. The period covered is 2015–2018; PPP rates are from 2022 OECD data.

Highlighting extremes, Brazil produces roughly some 25,000 APC-based papers, of which the average APC amounts up to around $1000, which was 5x minimum monthly wage in Brazil, while the Netherlands produces some 12,500 APC-based papers, of which the average APC rate is around $2500, which was 1.6 times minimum wage in the Netherlands!

After the correction for welfare level, by introducing PPP-corrected spending on APCs, the average APC rates for Brazil amount to US$ 2139, while PPP-corrected APC rate for the Netherlands amounts to US$ 2818. By applying purchasing power parity (PPP), we observe that for Brazil, the average APC rate has increased to 10x minimum monthly wage, while for the Netherlands this increased to 1.8x minimum monthly wage. This leads to the conclusion: for Brazilian academics it is much more difficult to live up to international academic-economic standards, and to allow for publishing in journals with an international standing, often published in journals

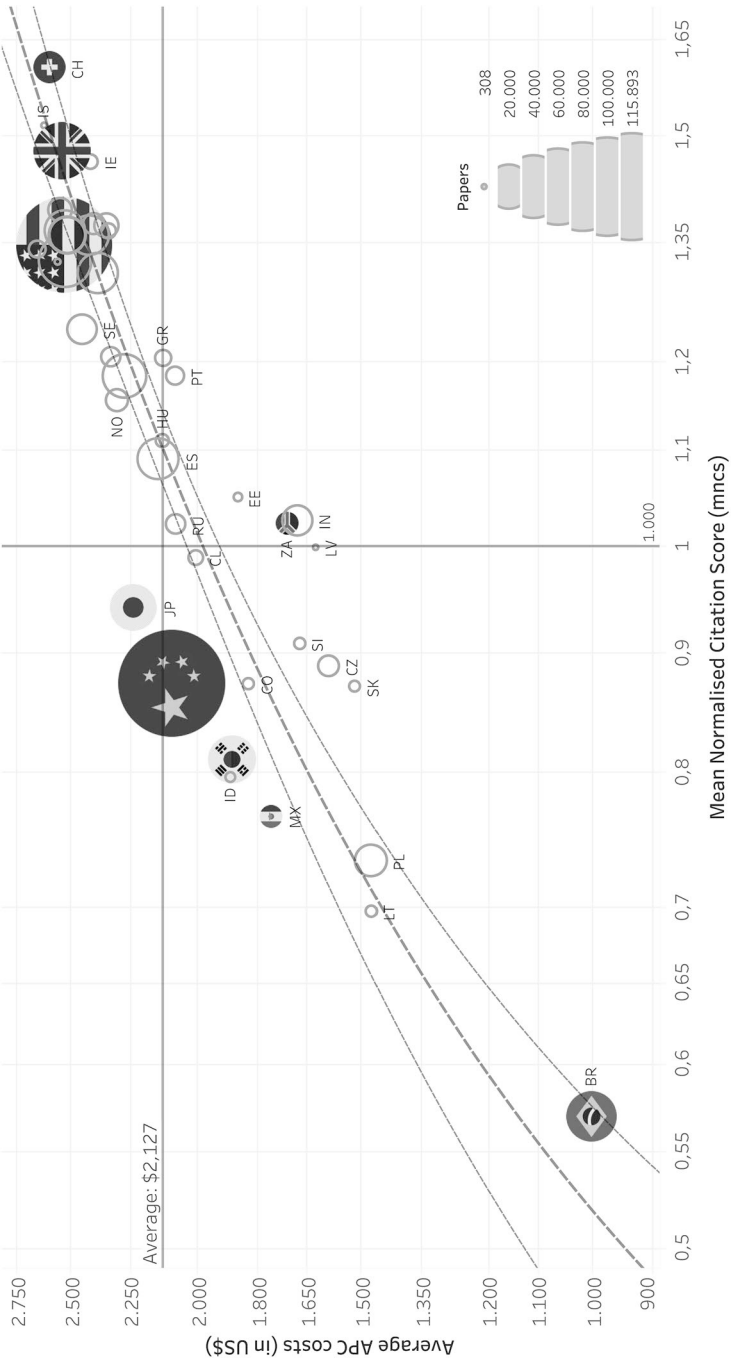


Fig. 13.1 Global spending on APCs per country, 2015–2018

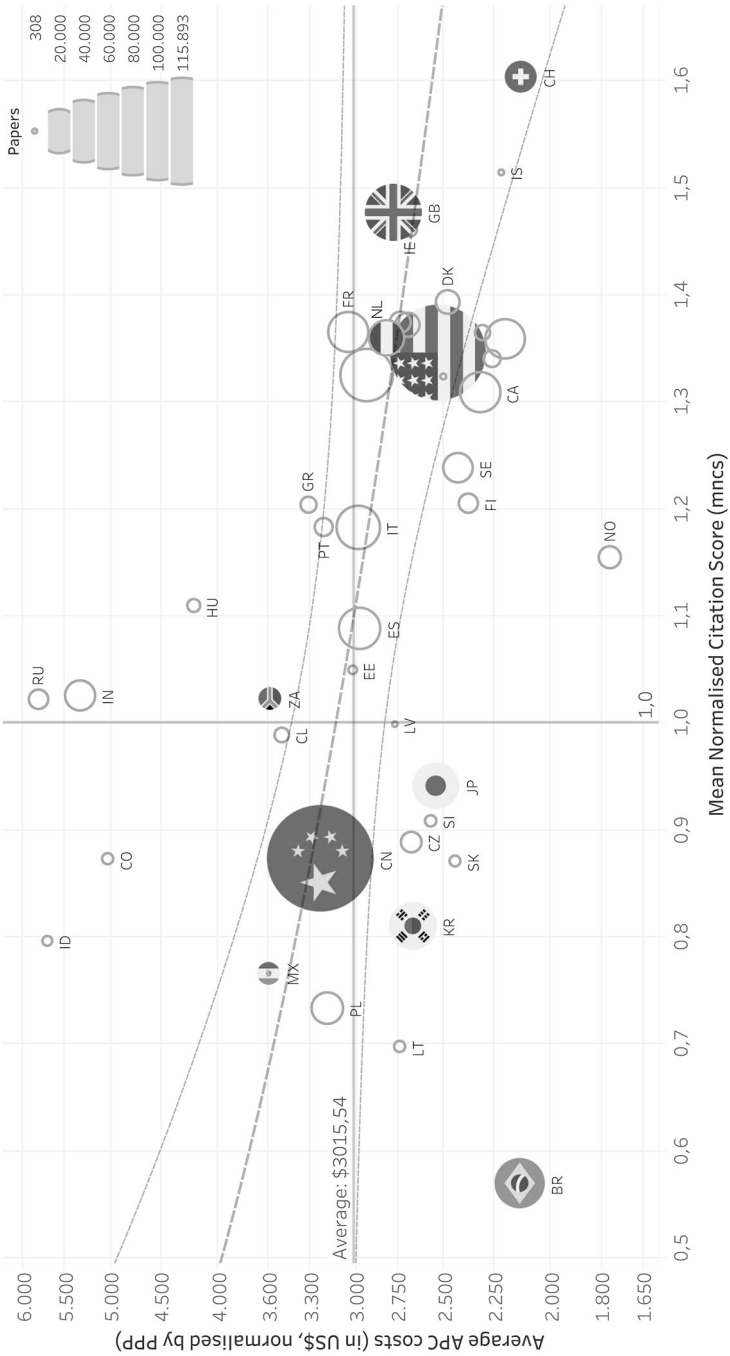


Fig. 13.2 Global spending on APCs per country, 2015–2018, corrected by PPP

processed for internationally oriented bibliographic platforms such as Web of Science, Scopus, or Dimensions, since APCs are simply too high to afford.

Our analysis of the trend lines in Figs. 13.1 and 13.2 has led us to some significant findings. Figure 13.1 shows that higher nominal spending on APCs is associated with a greater average impact of publications. In contrast, when we adjust the expenditures according to each country's PPP, Fig. 13.2 demonstrates an inversely proportional relationship, indicating that the citation impact increases as the normalised cost, which we can see as the actual cost of the investment, decreases. *These findings underscore the importance of considering the real cost of APCs in relation to their impact on publications.*

Returning focus to the Dutch situation, the development of Hybrid OA published material has been made possible due to the increasing number of agreements between the scholarly system, on the one hand, and the publishing industry, on the other hand. The bulk subscription contracts have been the stepping stone to transformative agreements, leading to the *Read & Publish* deals that the universities now have with publishers. These deals opened the way to publish in journals that are still within the subscription situation, but with the possibility to allow open access exceptions, so a hybrid form of open access publishing. However, a national mandate states that Gold OA is the route for the Netherlands to reach full OA from 2024 onwards, and the conflict becomes clear in the following figures.

In Fig. 13.3, we clearly see how the development of Gold OA publishing has been continuous from 2011 onwards. With the launch of the national mandate in 2014 and the climate already prepared for that in the years before, Hybrid OA publishing initially decreased, and only started to increase after the agreements with the publishers started to take from 2015/2026 onwards, with Hybrid OA becoming more important for Dutch academics, and overtaking Gold OA publishing.

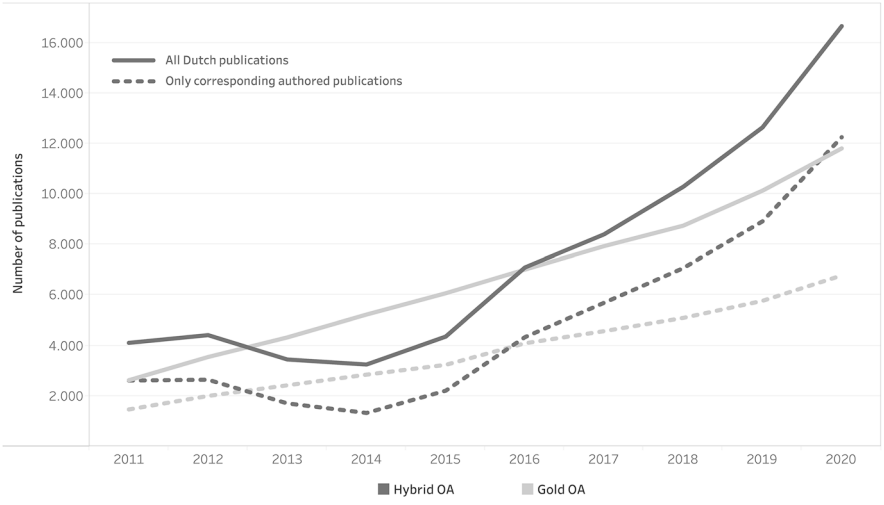


Fig. 13.3 Dutch output in Gold and Hybrid OA, 2011–2020, all publications

To analyse the attractiveness or strategic behaviour in scholarly publishing, the analysis also focused on the publication output in which Dutch authors were corresponding authors. In Fig. 13.3 this is indicated with the dotted lines, showing a similar trend as for all Dutch output. Hence, we cannot conclude that due to the Read & Publish deals, scholarly cooperation has led to an increase of papers coming from the Netherlands, with Dutch corresponding authors. However, we see that the gap between Gold and Hybrid OA is wider, indicating that Hybrid OA publishing has become more popular compared to Gold OA publishing, and on top of that, the Dutch Hybrid OA corresponding authored output surpasses in numbers the numbers of Gold OA published output in 2020.

13.4 Conclusions

What we witnessed in the two studies conducted is a complex power structure, with a variety of actors, both supranationally and nationally, conflicting interests within the national context, funding agencies, and a variety of motivations (academic, commercial, individual). A clear issue in the debate around Plan S was the relationship of the consequences of Plan S with the existing reward and recognition systems, as well as career perspectives of early career researchers in an international context. This aspect also returns in the study on the Dutch system, whereby academic freedom to choose the journals that suit you best is conflicting with prescribed ways of publishing in both the national mandate and Plan S. Finally, the study on global Gold Open Access publishing reflects a development towards a more unequal access to scholarly publishing, along the lines of available financial resources (Sawahel 2022; Zhang et al. 2022). All these outcomes clearly show unexpected consequences of the policies undertaken (Stone 2012). We do not have the same data as we have for the Dutch situation available for other countries, but assume that wherever these Transformative Agreements have been introduced, similar issues have popped up, as often no actions have been taken to also implement accompanying policies regarding recognition and reward policies, career policies, etc.

The open science movement started as an initiative to create more read access to the international serial literature, which was obstructed by sky high subscription rates. We now witness that due to the increasing price rates of open access publishing, in particular regarding the mandated Gold OA format, access to publishing in the international serial literature by the research communities from low- and middle-income countries has decreased. In contrast, read access is no longer a problem, since an increasing number of publications become automatically available (through the open access development itself). An important question in the background of this analysis is to what extent the transformative agreements, as these deals were named, did fail in the end, as no transformation is taking place at all.

In the meantime, cOAlitionS accepted Hybrid OA publishing as a transition model towards the intended Gold OA. However, the recent launch of their new

programme, "Towards Responsible Publishing", indicates a Plan S 2.0 that moves away from traditional publishers to further promote pre-print publishing as the preferred form of scholarly communication. And this happens against a background in which yet another model of open access publishing is becoming increasingly popular, namely the Diamond model. Diamond Open Access is the form of open access publishing in which the direct costs of publishing are not taken care of by the publishing author(s), but by a consortium supplying the money for scholarly open access publishing. A prominent example of this new way of organising open access publishing is the Open Library of Humanities (OLH 2024). This development should be much better prepared and supported from policymaking on different levels. Embedding publishing in Diamond OA journals would incentivise publishing there, while now the majority of academics are hesitant about both publishing in the journals and doing editorial and peer review work for such journals.

Currently developments are taking place globally around the open science agenda, internationally, nationally, and institutionally. International developments initiated by the EU, UNESCO, and funding actors like Science Europe and cOAlition S have a strong effect on national and institutional actions. And apart from the unexpected and unintended consequences described in this policy brief, the upswing of predatory publishing and ghost conferencing are in itself examples of unintended and unexpected negative consequences of the open science development (ALLEA 2023).

So, summarising, the push for Gold OA publishing has created a lot of buzz, as well as created a situation in which new players appeared, old players have reinforced their market positions, and sketchy journals and publishers have popped in the scenery, all of this being unforeseen effects of a myriad of science and funding policies implemented at supranational, national, and institutional levels.

Policy Implications

- Science policy should be more aware of the potential risk of unintended consequences occurring after policies have been initiated and implemented.
- Gold OA publishing has proved to not be the solution in the transition towards open science and open access publishing, as the publishing industry has increased its grip on scholarly publishing.
- The fact that openness is closely connected to other aspects of research management is reflected in the choice by Dutch academics to publish in hybrid OA journals rather than in Gold OA journals, as that is where the highly prestigious journals are located.

References for Further Reading

ALLEA (2023). ALLEA statement on curbing predatory practices in open access scholarly publishing. https://doi.org/10.26356/PREDATORYPUBLISHING

Brasil, A., Chen, C. M.-L., & van Leeuwen, T. (2022, September 7). Analysing Open Access publishing beyond Plan S: Motivations for publishing in Hybrid and Bronze OA formats. 26th International Conference on Science, Technology and Innovation Indicators (STI 2022), Granada, Spain. https://doi.org/10.5281/zenodo.6966750

Brasil, A., & van Leeuwen, T. (2022). The unseen costs of article processing charges: The different realities of Brazil and the Netherlands. 26th International Conference on Science, Technology and Innovation Indicators (STI 2022), Granada, Spain. https://doi.org/10.5281/zenodo.6966707

cOAlition S (2018). Plan S. https://www.coalition-s.org/

Fecher, B. and Friesike, S., Open Science: One Term, Five Schools of Thought (May 30, 2013). https://doi.org/10.2139/ssrn.2272036

OLH (2024). The Open Library of Humanities (openlibhums.org).

Sawahel, W. (2022). Is the emerging open-access model another closed system? University World News, Africa Edition, 16 June 2022.

Stone, D. (2012). The Policy Paradox. The art of political decision making. WW Norton, New York.

Zhang, L., Wei, Y., Huang, Y., and Sivertsen G. (2022). Should open access lead to closed research? The trends towards paying to perform research, Scientometrics, 127, 7653–7679.