2.3 National Research Evaluation Systems

Michael Ochsner, Emanuel Kulczycki, Aldis Gedutis, and Ginevra Peruginelli

Abstract: In current societies, research takes an important role as a driver of economic and social development. Therefore, research becomes more strategically relevant than ever. Consequently, research evaluation procedures have been implemented to monitor and steer research. This chapter offers an overview of national research assessment practices and reveals that they differ across countries and even research institutes as evaluations evolved over time. It comes to the conclusion that evaluation designers and research policymakers should establish an explicit link between policy goals and a specific research evaluation procedure taking the national evaluation system into account.

Keywords: research evaluation, evaluation systems, country comparison, peer review, social sciences and humanities, knowledge society, new public management, performance-based research funding.

Introduction

Recent decades have been marked by the transformation to a knowledge society. Knowledge generation has been seen as the major driver of economic development as well as an important means for reaching social goals. In such a knowledge society, universities and research institutions began to play an important role (see, e. g., Välimaa and Hoffman, 2008), which led to political demands for their accountability. Politicians, taxpayers, research agencies and managers became more interested in how the money provided to universities and research institutions is spent because the benefit of science was no longer just seen as the provision of highly qualified workers (e. g., Hoenack, 1993) but also as the provision of actual services and economic outputs (Gibbons et al., 1994; Etzkowitz and Leydesdorff, 1998). The development towards the knowledge society was linked with a shift in how public institutions were managed: rather than assuring high professional standards of public service procedures, public institutions now had to provide a service to the customer. Thus, the way in which such institutions were controlled changed considerably. Instead of procedures, outcomes were evaluated (Child, 2005): was there a "return on investment"? This came with institutionalised distrust (Deem, Hillyard and Reem

Michael Ochsner, D-GESS, ETH Zurich, Zurich (Switzerland) and FORS, Lausanne (Switzerland), ochsner@gess.ethz.ch
Emanuel Kulczycki, Adam Mickiewicz University Poznań, Poznań (Poland), emek@amu.edu.pl
Aldis Gedutis, Klaipėda University, Klaipėda (Lithuania), aldis.gedutis@ku.lt
Ginevra Peruginelli, Istituto di Informatica Giuridica e Sistemi Giudiziari – IGSG-CNR, Florence (Italy), peruginelli@igsg.cnr.it

https://doi.org/10.1515/9783110646610-011

2007): public servants had to be controlled regarding their efficiency and efficacity. This development also reached the universities (Deem et al., 2007; Hamann, 2016; Readings, 1996; Rolfe, 2013) and changed the way accountability was achieved here. In the past, quality assurance was guided by the principle of scientific freedom (peer review and rigid appointment procedures), but this no longer sufficed for being accountable to the public in the context of increased public and private spending on higher education and the rising importance of research for society and the economy. New Public Management thus asks for systematic evaluation procedures for publicly funded research on the institutional or national level (see, e. g., Geuna and Martin, 2003). At the same time, the share of competitively distributed funding for research increased considerably (Lepori, Reale and Spinello, 2018; Lepori et al., 2007). Thus, we argue in this chapter that research today is evaluated on different levels and in different time frames by various actors, leading to complex systems of evaluation procedures.

Given the logic of New Public Management, evaluation procedures were implemented in a top-down manner by governments and university administrations with the goal of the measurement of the direct achievements of research. Peer review was often criticised for being subjective (see, also critically, Daniel, Mittag and Bornmann, 2007; Peters and Ceci, 1982) and a more objective approach needed to be applied. Therefore, most of those procedures were – and still are – based on bibliometric and scientometric methods that pretend to facilitate a comparison of performances across departments, fields or countries.

These indicators, however, reflect only the research practices in a few disciplines of the natural and technical sciences and do not work well for many disciplines, being especially inadequate for the social sciences, humanities and arts (van Leeuwen, 2013; Nederhof, 2006). Moreover, the use of the indicators has been shown to have negative steering effects on researchers (see, e. g., de Rijcke et al., 2016). Therefore, not all countries implemented a bibliometric evaluation procedure. A well-known example is the UK where repeatedly a discussion was held on whether peer review could be replaced by indicators (see, e. g., Wilsdon et al., 2015), but the cornerstone of the evaluation stayed officially peer review.

Typologies of Evaluation Procedures

Several scholars set out to classify different types of research evaluation procedures or to present overviews of how research is evaluated across countries (Coryn et al., 2007; Galleron et al., 2017; Geuna, Hidayat and Martin, 1999; Geuna and Martin, 2001; 2003; Hicks, 2010; 2012; Jonkers and Zacharewicz, 2016; Lepori et al., 2007; Lepori, Reale and Spinello, 2018; Ochsner, Kulczycki and Gedutis, 2018; von Tunzelmann and Kraemer Mbula, 2003).

Three differentiations of evaluation procedures have been established (see also Whitley, 2007): the stage of evaluation (ex-ante vs. ex-post evaluation); link to fund-

ing (summative vs. formative evaluation); and method of evaluation (metric vs. peer review and different levels of evaluation). The first differentiates between procedures that evaluate research at the proposal stage, i. e. before the research has been carried out (ex-ante evaluation, e. g. for project funding) and procedures that evaluate research already conducted (ex-post evaluation, e. g. institutional evaluation). The second differentiates between procedures that allocate funding (summative) and procedures that aim at improving processes without any consequences regarding funding (formative). The third differentiates across different methods of evaluation, the most prominent being metric, indicator-based evaluation and peer review-based evaluation. Classifications differ as to which of the three differentiations they take into account. Most consider only one or two of those aspects.

The first group of classifications takes funding allocation to institutions into account and differentiates according to different methods of evaluation (Coryn et al., 2007; Geuna, Hidayat and Martin, 1999; Geuna and Martin, 2001; 2003; Hicks, 2012; von Tunzelmann and Kraemer, Mbula 2003). The studies of this first group of classification suggest that performance-based research funding systems (PRFSs) are implemented to enhance research excellence. Yet, both Hicks (2012) and Geuna and Martin (2003) raise the question of whether PRFSs are helping to achieve this goal as, for example, they are not encouraging interactions with industry (Hicks, 2012, p. 259) and they are costly and come with diminishing returns after the initial increase at the time of implementation (Geuna and Martin, 2003, p. 303).

The second group of classifications focuses on project funding, i. e. ex-ante evaluation (Lepori et al., 2007; Lepori et al., 2018; Zacharewicz et al., 2018). They show that competitively funded research projects gain rapidly in importance. They find that, first, there is considerable diversity across countries regarding funding instruments, agencies and beneficiaries. Second, there are many commonalities. In all countries, project funding is the second main channel of public funding of research and the share of competitive funding is growing. Furthermore, there is a common shift towards funding instruments oriented towards specific topics.

The typologies presented so far concern only a few specific evaluation procedures in the respective countries and do not take fully into account the three aspects of evaluation procedures identified above. If the goal of research evaluation is to influence research practice (if it wasn't, why would you evaluate in the first place; see, e. g., Hicks, 2012; Jonkers and Zacharewicz, 2016; Zacharewicz et al., 2018), it is not efficient to look at isolated research evaluation procedures or funding schemes. Researchers are influenced by many evaluation procedures and if policy aims at influencing research practice, it needs to take this diversity into account. Therefore, we focus in the next section especially on the third group of typologies, that focus on research evaluation systems (Galleron et al., 2017; van Gestel and Lienhard, 2019; Giménez-Toledo et al., 2019; Ochsner et al., 2018).

National Research Evaluation Systems

The combination of evaluation procedures in a country is complex. So complex that even experts can disagree about how research is evaluated in their countries (Galleron et al., 2017; Ochsner et al., 2018). Formal definitions of evaluation procedures can differ from actual practice, evaluation procedures evolve over time, and, most importantly, different evaluation procedures can be combined to balance out potential negative steering effects of the procedures (Ochsner et al., 2018). The studies showed that there is no dominant evaluation procedure in a country nor a coherent set of procedures. Rather, each country has a national evaluation system in place, i.e. a complex combination of different evaluation procedures with different aims, objects, scope and governing bodies (see also a similar conclusion regarding evaluation in law studies in van Gestel and Lienhard, 2019).

Ochsner et al. (2018) focus on eight characteristics regarding three types of evaluation procedures (institutional evaluation, project funding and national career promotion) and identify five ideal types. They are not real but rather abstract representations of evaluation systems. Actual national evaluation systems are rather combinations of the five ideal types. The first type is named "no national database, non SSH-specific" (not having a national publication database, using mainly non-metric evaluation procedures and not allowing for SSH adaptations). The second ideal type is named "non-metric, SSH-specific" (not having a publication database, not relying on metrics for their evaluations, not incentivising publications in English and having dedicated funding programs for SSH disciplines). The third ideal type is called "performance-based funding, non-metric" (having a PRFS in place that allows for SSH-specific adaptations and is based on metrics derived from a national publication database; funding link being either established through informed peer review or through a combination of a metric PRFS with an evaluation based on peer review to counterbalance the metric nature of the PRFS). The fourth type is named "performance-based funding, metric" (PRFSs being based on a national database and a metric evaluation that allows for SSH adaptations and not incentivising publications in English). Finally, the fifth type is named "metric, push for English" (metric evaluation based on a national publication database linked to funding and not allowing for SSH adaptations and incentivising publications in English). Table 1 shows how countries can be attributed to the five ideal types. It is remarkable that countries cluster regionally, which suggests that historical and political structure play a role in how research is evaluated. Similar results are reported regarding the role of books in evaluation procedures (Giménez-Toledo et al., 2019).

Table 1: Five ideal types of national evaluation systems and classification of countries.

Ideal Type	Countries closest to the ideal type	Countries difficult to classify, closest type chosen
No national database, non SSH-specific	Cyprus, France, Iceland, Macedonia, Malta, Montenegro, Portugal, Spain	Bulgaria, Italy
Non-metric, SSH-specific	Austria, Germany, Ireland, the Netherlands, Serbia, Switzerland	
Performance-based funding, non-metric	Lithuania, Norway, South Africa	Denmark, Israel
Performance-based funding, metric	Czech Republic, Croatia, Poland	Finland
Metric, push for English	Bosnia-Herzegovina, Estonia, Hungary, Slovakia, Slovenia, Romania	Latvia

Note: variables used for the classification (yes/no): institutional evaluation results affect funding; main method of institutional evaluation are metrics; system incentivises English language publications; institutional evaluation procedures reflect gender issues; existence of national publication database; SSH-specific institutional evaluation procedures; SSH-specific project funding programmes; existence of national career promotion procedure.

These results are based on how evaluation systems are perceived by experts. In the follow-up study, the experts from the countries collected national regulations on evaluation procedures to further systematise actual evaluation policies. The discussions showed that the three procedures used so far do not suffice to adequately describe the evaluation systems. The analysis of regulations identified seven different types of evaluation procedures: accreditation, formative national evaluation, performance-based national evaluation, excellence initiatives, national career promotion, government project funding and evaluation of academies of sciences or research institutes. Each country has its own mix of two to six evaluation procedures on the national level (see Ochsner, 2020), showing that national research evaluation procedures are complex and diverse and different types of evaluation procedures serve different goals.

Conclusion

Research is a complex endeavour and, therefore, research evaluation practices are diverse. There is no such thing as "the evaluation procedure" in a country but each country has a distinct set of evaluation procedures making up a national evaluation system. The situation of research is different in each country and so are research policies. Evaluation procedures need to reflect the needs of the research landscape in the country, its research policy and the academic structure in the country.

Consequently, it makes sense that each country has its own set of evaluation procedures in place rather than to try to standardise evaluation procedures across countries. Nevertheless, countries striving to achieve similar goals with their evaluation procedures can learn from each other's experiences.

Still, in practice, an explicit link of policy goals with a specific set of evaluation procedures in a country is missing. Many evaluation procedures seem to have evolved more or less arbitrarily, which can be seen by the geographical clustering among types of evaluation systems. We therefore encourage evaluation designers and research policymakers to establish an explicit link between policy goals and a specific research evaluation procedure, taking the national evaluation system into account.

Acknowledgements

This chapter is based upon work from COST Action CA 15137 'European Network for Research Evaluation in the Social Sciences and the Humanities (ENRESSH)', supported by COST (European Cooperation in Science and Technology).

References

Child, J 2005, *Organisation: contemporary principles and practices*, Blackwell, Malden, MA and Oxford.

Coryn, CLS, Hattie, JA, Scriven, M & Hartmann, DJ 2007, 'Models and Mechanisms for Evaluating Government-Funded Research: An International Comparison', *American Journal of Evaluation,* vol. 28, no. 4, pp. 437–457.

Daniel, HD, Mittag, S & Bornmann, L 2007, 'The potential and problems of peer evaluation in higher education and research', in A Cavalli (ed.), *Quality assessment for higher education in Europe*, pp. 71–82, Portland Press, London.

de Rijcke, S, Wouters, PF, Rushforth, AD, Franssen, TP & Hammarfelt, B 2016, 'Evaluation practices and effects of indicator use – a literature review', *Research Evaluation*, vol. 25, no. 2, pp. 161–69.

Deem, R, Hillyard, S & Reed, M 2007, *Knowledge, higher education, and the new managerialism. The changing management of UK universities*, Oxford University Press, Oxford.

Etzkowitz, H & Leydesdorff, L 1998, 'The endless transition: a "Triple Helix" of university-industry-government relations', *Minerva*, vol. 36, no. 3, pp. 203–08.

Galleron, I, Ochsner, M, Spaapen, J & Williams, G 2017, 'Valorizing SSH research: Towards a new approach to evaluate SSH research' value for society', *fteval Journal for Research and Technology Policy Evaluation*, vol. 44, pp. 35–41.

Geuna, A, Hidayat, D & Martin, BR 1999, *Resource Allocation and Research Performance: The Assessment of Research*, Report prepared for the Higher Education Funding Council of England, SPRU, Brighton.

Geuna, A & Martin, BR 2001, *University Research Evaluation and Funding: An International Comparison*, SPRU Electronic Working Paper Series (vol. 71), SPRU, Brighton, http://www.sussex.ac.uk/Units/spru/publications/imprint/sewps/sewp71/sewp71.pdf (July 15, 2020).

Geuna, A & Martin, BR 2003, 'University Research Evaluation and Funding: An International Comparison', *Minerva*, vol. 41, no. 4, pp. 277–304.

Gibbons, M, Lomoges, C, Nowotny, H, Schwartzmann, S, Scott, P & Trowm, M 1994, *The new production of knowledge. The dynamics of science and research in contemporary societies*, Sage.

Giménez-Toledo, E, Mañana-Rodríguez, J, Engels, TCE, Guns, R, Kulczycki, E, Ochsner, M, Pölönen, J, Sivertsen, G & Zuccala, AA 2019, 'Taking scholarly books into account, part II: a comparison of 19 European countries in evaluation and funding', *Scientometrics*, vol. 118, no. 1, pp. 233–251.

Hamann, J 2016, 'The visible hand of research performance assessment', *Higher Education*, vol. 72, no. 6, pp. 1–19.

Hicks, D 2012, 'Performance-based university research funding systems', *Research Policy*, vol. 41, no. 2, pp. 251–61.

Hicks, D 2010, 'Overview of models of performance-based research funding systems', in OECD (ed.), *Performance-based Funding for Public Research in Tertiary Education Institutions*, OECD Publishing, Paris, pp. 23–52.

Hoenack, SA 1993, 'Higher Education and Economic Growth', in WE Becker & DR Lewis (eds.), *Higher Education and Economic Growth*, pp. 21–50, Springer, Dordrecht.

Jonkers, K & Zacharewicz, T 2016, *Research Performance Based Funding Systems: a Comparative Assessment*, European Commission, Brussels.

Lepori, B, Reale, E & Spinello, AO 2018, 'Conceptualizing and measuring performance orientation of research funding systems', *Research Evaluation*, vol. 27, no. 3, pp. 171–83.

Lepori, B, van den Besselaar, P, Dinges, M, Poti, B, Reale, E, Slipersæter, S, Thèves, J & van der Meulen, B 2007, 'Comparing the evolution of national research policies: What patterns of change?', *Science and Public Policy*, vol. 34, no. 6, pp. 372–88.

Nederhof, AJ 2006, 'Bibliometric monitoring of research performance in the social sciences and the humanities: A review', *Scientometrics*, vol. 66, no. 1, pp. 81–100.

Ochsner, M 2020, *WG1. Aligning research evaluation with policy goals: Risks and opportunities. Presentation of the final results at the ENRESSH final meeting*, ENRESSH, Paris, https://enressh.eu/wp-content/uploads/2018/04/Stakeholders_PolicyGoals_v04.pdf (July 15, 2020).

Ochsner, M, Kulczycki, E & Gedutis, A 2018, 'The Diversity of European Research Evaluation Systems', in Proceedings of the 23rd International Conference on Science and Technology Indicators, CWTS, Leiden, pp. 1234–41.

Peters, DP & Ceci, SJ 1982, 'Peer-review practices of psychological journals: The fate of published articles, submitted again', *The Behavioural and Brain Sciences*, vol. 5, no. 2, pp. 187–95.

Readings, B 1996, *The university in ruins*, Harvard University Press, Cambridge, MA.

Rolfe, G 2013, *The university in dissent. Scholarship in the corporate university*, Routledge, Abingdon.

van Gestel, R & Lienhard, A, eds. 2019, *Evaluating academic legal research in Europe. The advantage of lagging behind*, Edward Elgar, Cheltenham.

van Leeuwen, TN 2013, 'Bibliometric research evaluations, Web of Science and the Social Sciences and Humanities: a problematic relationship?', *Bibliometrie – Praxis und Forschung*, vol. 2, no. 8, pp. 1–18.

von Tunzelmann, N & Kraemer Mbula, E 2003, *Changes in research assessment practices in other countries since 1999: final report*, http://www.ra-review.ac.uk/reports/Prac/Changing Practices.pdf (July 15, 2020).

Välimaa, J & Hoffman, D 2008, 'Knowledge society discourse and higher education', *Higher Education*, vol. 56, no. 3, pp. 265–285.

Whitley, R 2007, 'Changing governance of the public sciences', in R Whitley & J Gläser (eds.), *The changing governance of the sciences. The advent of research evaluation systems*, pp. 3–27, Springer, Dordrecht.

Wilsdon, J, Allen, L, Belfiore, E, Campbel, P, Curry, S, Hill, S, Jones, R, Kain, R, Kerridge, S, Thelwall, M, Tinkler, J, Viney, I, Wouters, P, Hill, J & Johnson, B 2015, *The metric tide. Report of the independent review of the role of metrics in research assessment and management*, HEFCE, London.

Zacharewicz, T, Lepori, B, Reale, E & Jonkers, K 2018, 'Performance-based research funding in EU Member States – a comparative assessment', *Science and Public Policy*, vol. 46, no. 1, pp. 105–115.