Place, role, form and significance of peer review in National Research Evaluation Systems

By Michael Ochsner

National Research Evaluation

During the last decades, due to a shift to new public management policies and increasing pressure on efficiency and accountability, most universities have implemented comprehensive research evaluation procedures (Geuna & Martin, 2003; Whitley & Gläser, 2007). Also, on the national level, the importance of competitive project funding has been increasing since the 1970 and is still rising in most countries (Lepori et al., 2007; Lepori et al., 2018). While there is some discussion whether there is convergence or persisting diversity regarding research evaluation procedures in different countries (Lepori et al., 2007), it is known that research evaluation differs between countries. Several typologies have been proposed to describe or systematise research evaluation procedures across countries (Coryn et al., 2007; Geuna & Martin, 2001, 2003; Hicks, 2012; Jonkers & Zacharewicz, 2016; von Tunzelmann & Mbula, 2003; Whitley, 2007). Yet, the typologies have some drawbacks: they only focus on some aspects, like financing or performance-based funding, include only a few countries for which data is available or exclude the SSH. ENRESSH therefore set out to investigate how research is evaluated in Europe with a special focus on SSH and from the perspective of the researchers. The four-year mixed-methods project consists of a multistage procedure to investigate the evaluation of SSH research in the participating countries (for a description of the procedure, see Galleron et al., 2017). The first phase consisted of a two-round Delphi survey among experts in research evaluation. Its aim was to gain an overview of commonalities and differences in research evaluation across countries as well as to get a common understanding of terms and definitions when studying evaluation procedures. Its product was a typology of national research evaluation systems. The second phase adopted a qualitative approach. Drawing on the results of the first phase, different types of evaluation procedures and a common grid of features of such evaluation procedures were identified and country rapporteurs filed a report on the national evaluation system in their country, i.e. the country-specific combination of the different types of evaluation procedures. The second phase is still ongoing. The results from the work accomplished so far show that there is no such thing as "national research evaluation": There is neither one single or one dominant research evaluation procedure in place per country nor a coherent set of combined procedures but rather a complex combination of many evaluation procedures with different aims, objects, scope and governing bodies. Thus, each country has its own complex *national research evaluation system.* Evaluation procedures not only differ widely across countries, it is even not always clear to distinguish different procedures because outcomes of one procedure can be used for another, leading to difficulties comparing evaluation practices across countries. Furthermore, even experts disagree about how research is evaluated in their countries (see Galleron et al., 2017), for example because formal definitions of a procedure might differ from actual implementation. Sometimes research evaluation systems evolved over time adding and changing procedures without relating them to each other, sometimes different evaluation procedures are combined by

design so that one procedure remediates negative steering effects of another procedure (Ochsner et al., 2018).
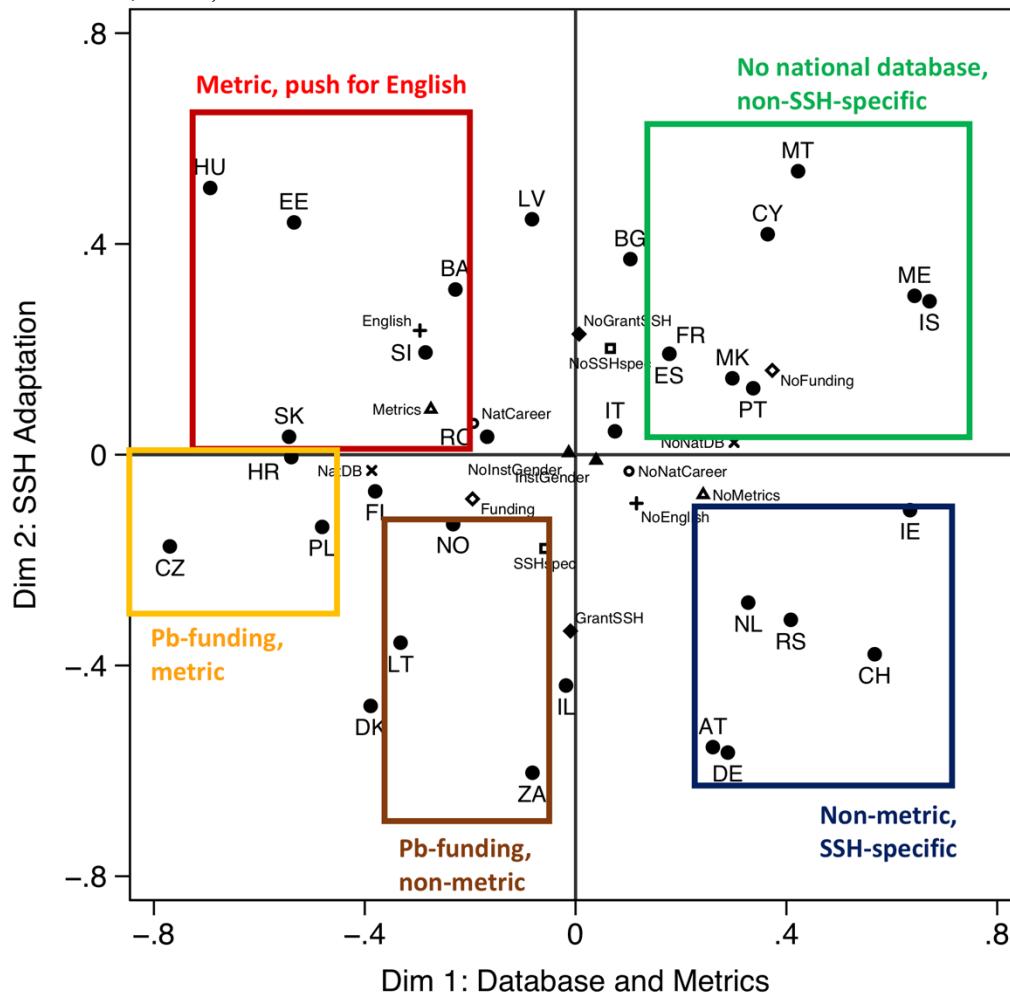

Figure 1. Map of National Research Evaluation Systems from Multiple Correspondence Analysis

Notes. Full circles represent countries, all other symbols represent dummy variables of characteristics of research evaluation systems. English/NoEnglish: system incentivises (or not) English language publications; (No)Funding: evaluation results affect funding; (No)GrantSSH: SSH-specific grant programmes; (No)InstGender: evaluation procedures reflect gender issues; (No)Metrics: main method of evaluation are metrics; (No)NatCareer: national career promotion procedure; (No)NatDB: national publication database; (No)SSH-spec: SSH-specific institutional evaluation procedures.

Source: Ochsner et al. (2018)

The ENRESSH project identified five "ideal types"[7] of research evaluation systems (see Figure 1): "no national database, non-SSH" representing countries without a national publication database, having mainly non-metric evaluation procedures in place and do not have

---

[7] The term "ideal type" is used in the Weberian sense (Weber 1904/1949): ideal types serve to systematize differences in evaluation procedures. They are not real but abstract representations of the phenomena to describe. Real evaluation systems can share characteristics of multiple ideal types.

European Network for Research Evaluation in the Social Sciences and Humanities. COST action 15137. www.enressh.eu

56

SSH-specific adaptations; "non-metric, SSH-specific" characterised by not having a publication database, not basing evaluation on metrics, not incentivising publications in English, and having dedicated funding programs for SSH disciplines; "performance-based funding, non-metric" consisting of a performance-based funding model (PRFS) that allows for SSH-specific adaptations and is based on metrics derived from a national publication database where the funding link is either established through informed peer review or the metric performance-based funding model is combined with an evaluation procedure based on peer review to counter-balance the metric nature of the PRFS; "performance-based funding, metric" representing PRFS based on a national database and a metric evaluation that allows for SSH adaptations, not incentivising publications in English; and "metric, push for English" characterised by metric evaluations based on a national publication database linked to funding and not allowing SSH adaptations while incentivising publications in English. Note that countries within the same type do not completely correspond with the ideal type and can be dissimilar to each other on some dimensions. Figure 1 shows the map of national research evaluation systems based on a multiple correspondence analysis using eight variables describing research evaluation procedures (for details of the methodology, see Ochsner et al., 2018). Countries thereby cluster regionally, suggesting that historical and political structures play a role. Furthermore, it is also remarkable that research-intensive countries rely less on metric models but rather adhere to adaptive designs while other countries try to increase their position in the international ranking game by adopting metric models that favour publications in English (Ochsner et al., 2018). Similar results emerged from another ENRESSH project that investigated the role of books in evaluation procedures (Giménez-Toledo et al., 2019).

Peer review in national evaluation systems

Peer review is an important method in the evaluation of research. In the above-mentioned survey on research evaluation in the different ENRESSH countries, experts from most countries report that aspects of peer review are in place in the evaluation system. Some countries like Switzerland, Germany, Austria, the Netherlands, Serbia and Ireland base their evaluation on peer review. However, also countries known for their metric performance-based funding scheme like Norway might have another evaluation procedure in place that relies on peer review. At the same time, even metric-based systems can include an aspect of peer review: taking up the example of Norway again, the metric performance-based funding model includes a scheme of publication classification that consists of two levels. The decision which publication channel is considered as the most prestigious level (the so-called level 2) is taken by experts in the field. This vital component in the scheme thus represents a form of peer review. Similarly, journal and publisher lists can be developed based on judgements of the scholarly community (Giménez-Toledo, 2016; Giménez-Toledo et al., 2019). While in these cases the works are not directly judged by experts, the experts nevertheless judge the publication channel.

Almost all countries have a competitive project funding scheme implemented (for the increasing share of government funds distributed through competitive project funding, see Lepori et al., 2007; Lepori et al., 2018). The submitted research projects are in all instances evaluated by experts.

Thus, peer review takes an important part in many evaluation procedures of national research evaluation systems. The role, form, and significance of peer review in the evaluation procedure, however, can differ strongly, even within the same country across different evaluation procedures. Thereby, the role, form and significance of peer review are mostly independent from the type of evaluation procedure that can differ regarding purpose (formative versus summative), perspective (ex-ante versus ex-post) and level (publication, project, scholar, research unit, institution, discipline in a country).

The following *roles* of peer review in evaluation procedures can be distinguished: Peer review as the primary method of evaluation; peer review grading as part of a set of indicators; and peer review for assigning levels of publication channels or number of points for publication types.

Peer review can take the following *forms* in evaluation procedures: peer review in panels; independent reviews informing a committee; informed peer review in panels (reviewers can make use of metric information); experts judging metrics such as CVs, objectives or citation data; expert ratings through surveys (e.g. for journal lists); or discussions between experts and the evaluated. Evaluation procedures can also use a combination of different forms of peer review.

The *significance* of the peer review element in an evaluation procedure ranges almost gradually from deciding how many points are assigned to one single output in an evaluation to deciding over a career or project. Thereby peer review can be used to discuss weaknesses and strengths for the positioning and further development of a candidate or a department in formative evaluation procedures just as well as it can be used to distribute funding or taking decisions over a career by attributing or refusing promotion.

This short overview shows that peer review is a complex phenomenon that differs between contexts. It takes an important function whatever role, form and significance it has as even metric procedures base their insights on peer review: it is peer review that decides whether an output is published in a specific journal, peers decide which channels are more prestigious ones or research is directly evaluated by some form of peer review. While peer review is often criticised, it is impossible to evaluate research without any peer judgement – and it is difficult to imagine a useful outcome of an evaluation without any peer influence. Thus, it is important to understand that also metric procedures are only superficially objective and themselves dependent on peer review (see also Donovan, 2007). Instead of investing in metric evaluation to replace peer review or to avoid its negative aspects, efforts should focus on a better understanding of how peer review works and how to combine peer review and metrics instead of playing them off against each other. More research is needed on the roles, forms and significance of peer review in different evaluation procedures within and across countries and on how to improve peer review regarding the issues that have been identified, such as subjectivism, potential biases, mainstreaming and penalising interdisciplinary research. Metric procedures can take a role in improving the peer review process but not replace it (see also Donovan, 2007; Reale et al., 2018).

References

Coryn, C. L. S., Hattie, J. A., Scriven, M., & Hartmann, D. J. (2007). Models and mechanisms for evaluating government-funded research: An international

European Network for Research Evaluation in the Social Sciences and Humanities. COST action 15137. www.enressh.eu

58

comparison. *American Journal of Evaluation, 28*(4), 437–457. https://doi.org/10.1177/1098214007308290

Donovan, C. (2007). The qualitative future of research evaluation. *Science and Public Policy, 34*(8), 585–597. https://doi.org/10.3152/030234207X256538

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, 44*, 35–41. https://doi.org/10.22163/fteval.2017.274

Geuna, A., & Martin, B. R. (2001). University research evaluation and funding: An international comparison. *SPRU Electronic Working Paper Series* (Vol. 71). SPRU. http://www.sussex.ac.uk/Units/spru/publications/imprint/sewps/sewp71/sewp71.pdf

Geuna, A., & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva, 41*(4), 277–304. https://doi.org/10.1023/B:MINE.0000005155.70870.bd

Giménez-Toledo, E. (2016). Assessment of journal & book publishers in the humanities and social sciences in Spain. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research assessment in the humanities. Towards criteria and procedures* (pp. 91–102). Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_8

Giménez-Toledo, E., Mañana-Rodríguez, J., Engels, T. C. E., Guns, R., Kulczycki, E., Ochsner, M., Sivertsen, G., & Zuccala, A. A. (2019). Taking scholarly books into account, part II: A comparison of 19 European countries in evaluation and funding. *Scientometrics, 118*(1), 233–251. https://doi.org/10.1007/s11192-018-2956-7

Hicks, D. (2012). Performance-based university research funding systems, *Research Policy 41*(2), 251–261. https://doi.org/10.1016/j.respol.2011.09.007

Jonkers, K., & Zacharewicz, T. (2016). *Research Performance Based Funding Systems: a Comparative Assessment* (No. EUR 27837 EN). European Commission.

Lepori, B., Reale, E., & Spinello, A. O. (2018). Conceptualizing and measuring performance orientation of research funding systems. *Research Evaluation, 27*(3), 171–183. http://doi.org/10.1093/reseval/rvy007

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, 34*(6), 372–388. https://doi.org/10.3152/030234207X234578

Ochsner, M., Kulczycki, E., & Gedutis, A. (2018). The Diversity of European Research Evaluation Systems. *Proceedings of the 23rd International Conference on Science and Technology Indicators* (pp. 1234–1241). Leiden: Leiden University. https://openaccess.leidenuniv.nl/handle/1887/65217

Reale, E., Avramov, D., Canhial, K., Donovan, C., Flecha, R., Holm, P., Larkin, C., Lepori, B., Mosoni-Fried, J., Oliver, E., Primeri, E., Puigvert, L., Scharnhorst, A., Schubert,

A., Soler, M., Soòs, S., Sordé, T., Travis, C., & Van Horik, R. (2018). A review of literature on evaluating the scientific, social and political impact of social sciences and humanities research. *Research Evaluation*, *27*(4), 298–308. https://doi.org/10.1093/reseval/rvx025

von Tunzelmann, N., & Mbula, E. K. (2003). *Changes in research assessment practices in other countries since 1999: Final report.* http://www.ra-review.ac.uk/reports/Prac/ChangingPractices.pdf

Weber, M. (1904/1949). 'Objectivity' in social science and social policy. In E. A. Shils & H. A. Finch (Eds. and trans.), *The methodology of the social sciences* (pp. 49–112). Free Press.

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 Netherlands. https://doi.org/10.1007/978-1-4020-6746-4

Whitley, R., & Gläser, J. (Eds.). (2007). *The changing governance of the sciences. The advent of research evaluation systems.* Springer Netherlands. https://doi.org/10.1007/978-1-4020-6746-4

European Network for Research Evaluation in the Social Sciences and Humanities. COST action 15137. www.enressh.eu

60

ENRESSH Work Group 1

Overview of Peer Review Practices

in the SSH

ENRESSH Report

Michael Ochsner, ETH Zurich and FORS, Lausanne, Switzerland
Nina Kancewicz-Hoffman, Institute of Literary Research, Polish Academy of Sciences, Poland
Marek Hołowiecki, Adam Mickiewicz University Poznań, Poland
Jon Holm, National Research Council, Norway

With chapters by Mirjam Aeschbach, Gemma Derrick, Tim C. E. Engels, Elea Giménez-Toledo, Raf Guns, Jon Holm, Marek Hołowiecki, Sven E. Hug, Nina Kancewicz-Hoffman, Karolina Lendák-Kabók, Lai Ma, Jorge Mañana-Rodríguez, Michael Ochsner, Ginevra Peruginelli, Janne Pölönen, Elías Sanz-Casado, Tony Ross-Hellauer, Jadranka Stojanovski, Marc Vanholsbeeck

Chapters reviewed by Elea Giménez-Toledo, Aldis Gedutis, Jon Holm, Karolina Lendák-Kabók, Michael Ochsner, Ginevra Peruginelli, Janne Pölönen, Linda Sīle, Yulia Stukalina, Mimi Urbanc

WG to which the report/deliverable is related: WG 1

Grant period to which the report/deliverable is related: Grant Period 4

Version 2.1, April 2020

Executive summary

Peer review is an important method of research evaluation, and it seems that the only adequate way to evaluate SSH research involves some form of peer review. Even if bibliometrics and other quantitative ways of evaluation may provide information on some aspects of SSH research like productivity and publication strategies of research units, metrics-based indicators should be used with caution in SSH due to low coverage of SSH fields in the standard publication databases and a mismatch between dimensions of quality as defined by peers and standard bibliometric indicators. Still, peer review faces many issues and challenges. This report identifies the challenges particularly relevant for the SSH, such as different and thus often conflicting research paradigms or epistemological styles of reviewers and applicants or authors; difficulty in many SSH disciplines to define and evaluate research methodology compared to STEM disciplines; the lack of the idea of linear progress and a much longer time span necessary to evaluate academic impact of publications; the diversity of publication outputs and specific importance of books or monographs; the importance of local languages; challenges related to recent developments in research and its evaluation related to growing interdisciplinarity and the Open Science agenda. To this, the general challenges of peer review are added, such as the risk of gender bias, conservative bias, workload for all parties involved.

The report concludes that peer review fulfils different functions and that peer review practices not only need to acknowledge different disciplinary particularities but also their evaluative context. Rather than playing metrics and peer review off against each other, the focus should be on their optimal use and combination within different evaluation situations. This is especially important when it concerns the SSH because the disciplines falling under this umbrella term share the concurrency of different paradigms and a context-dependent, sometimes interpretative mode of knowledge generation and the use of a wide range of dissemination channels. This leads to a particular challenge regarding the burden of reviewers because SSH disciplines often act in a local context in national languages and include small disciplinary communities.

The SSH disciplines should develop their own ways to adequately evaluate their research, and peer review takes an important part in that. The past has shown that automatically copying evaluation procedures from STEM disciplines did not always work out well. However, the SSH community is well resourced to analyse and remediate the current tensions in research policies between funders' expectations of societal impact and the value of academic autonomy, between the ambition of mainstreaming of SSH research and the care for specific SSH methods and practices, and not least the threatened legitimacy of science in the post-factual society. The task of the SSH community should not only be to defend the integrity of scholarly disciplines, but to contribute to the development of new practices of research assessments that may build bridges between different communities of researchers and between the world of research and society at large.

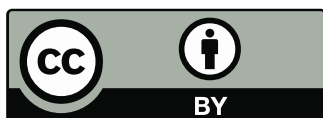European Network for Research Evaluation in the Social Sciences and Humanities. COST action 15137. www.enressh.eu

2

Acknowledgements

Recommended citation of the whole report:

Ochsner, M., Kancewicz-Hoffman, N., Hołowiecki, M., & Holm, J. (Eds.). (2020). *Overview of peer review practices in the SSH. ENRESSH Report.* European Network of Research Evaluation in the Social Sciences and Humanities. https://dx.doi.org/10.6084/m9.figshare.12032589

Recommended citation of a chapter in the report:

Derrick, G., & Ross-Hellauer, T. (2020). Peer Review in SSH: In Need of Development? In M. Ochsner, N. Kancewicz-Hoffman, M. Hołowiecki, J. Holm (Eds.), *Overview of peer review practices in the SSH. ENRESSH Report* (pp. 10–14). European Network of Research Evaluation in the Social Sciences and Humanities. https://dx.doi.org/10.6084/m9.figshare.12032589

European Network for Research Evaluation in the Social Sciences and Humanities. COST action 15137. www.enressh.eu

3

Table of Content

European Network for Research Evaluation in the Social Sciences and Humanities. COST action 15137. www.enressh.eu

4

European Network for Research Evaluation in the Social Sciences and Humanities. COST action 15137. www.enressh.eu

5