

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 Sile,
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.

Keywords

Peer Review, Evaluation, Criteria, Societal Impact, Books, Funding, Open Access

Acknowledgements

This report is a result of work from COST Action CA 15137 “European Network for Research Evaluation in the SSH (ENRESSH)”, supported by COST (European Cooperation in Science and Technology).

The editors would like to thank all participants from Work Group 1 “Conceptual Frameworks of Research Evaluation in the SSH” who participated in one of the sessions regarding this report (alphabetical order):

Georgios Afxentiou, Judit Bar-Ilan, Ondřej Daniel, Katya De Giovanni, Ioana Galeron, Aldis Gedutis, Haris Gekić, Elea Giménez-Toledo, Agnė Girkontaitė, Raf Guns, Jon Holm, Marek Hołowiecki, Todor Hristov Dechev, Dragan Ivanović, Nina Kancewicz-Hoffman, Andreja Isenič Starčič, Arnis Kokorevics, Litaugas Kraniuskas, Emanuel Kulczycki, Karolina Lendák-Kabók, Lai Ma, Jorge Mañana-Rodríguez, Ketrina Mijo, Michael Ochsner, Elena Papanastasiou, Sanja Peković, Ginevra Peruginelli, Janne Pölönen, Hulda Proppé, Tony Ross-Hellauer, Elías Sanz-Casado, Ana Ramos, Linda Sīle, Karel Šima, Dagmar Simon, Jolanta Šinkūnienė, Gunnar Sivertsen, Jack Spaapen, Jadranka Stojanovski, Mimi Urbanc, Marc Vanholsbeeck, Maja Vehovec, Albena Vutsova, Alesia Zuccala



This work is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

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>

Table of Content

Executive summary	2
Acknowledgements	3
PART I: General Framework: state of the art of peer review in the SSH in Europe	6
Introduction: Aim and scope of the report	6
By Nina Kancewicz-Hoffman, Michael Ochsner, Marek Hołowiecki & Jon Holm	
Peer Review in SSH: In Need of Development?.....	10
By Gemma Derrick & Tony Ross-Hellauer	
PART II: Issues and Discussions Specific to and Most Relevant for SSH Peer Review .	15
Evaluation Criteria and Methodology	15
By Michael Ochsner	
Definitions of societal impact and its evaluation in context	23
By Gemma Derrick	
Does excellence have to be in English? Language diversity and internationalisation in SSH research evaluation.....	32
By Nina Kancewicz-Hoffman & Janne Pölönen	
PART III: Guidelines, Procedures and Formal Criteria Versus their Practical Application	42
Review of guidelines and recommendations for evaluation	42
By Nina Kancewicz-Hoffman	
Ambiguity in identification of scholarly peer-reviewed publications.....	50
By Janne Pölönen, Tim C. E. Engels & Raf Guns	
Place, role, form and significance of peer review in National Research Evaluation Systems.....	55
By Michael Ochsner	
Practices of peer review in the SSH I: A systematic review of peer review criteria.....	61
By Sven E. Hug, Marek Hołowiecki, Lai Ma, Mirjam Aeschbach & Michael Ochsner	
Practices of peer review in the SSH II: Peer review and other manuscript selection processes for books in the SSH	67
By Elea Giménez-Toledo & Jorge Mañana-Rodríguez	
Practices of peer review in the SSH III: peer review in the legal domain. Three parallel case studies in Italy, Spain and Croatia.....	71
By Ginevra Peruginelli, Elías Sanz-Casado & Jadranka Stojanovski	

PART IV: Current Challenges for Peer Reviewing: Towards More Open and Gender-Sensitive Peer Reviewing Practices in the SSH	78
A gender and geopolitical perspective on peer review	78
By Karolina Lendák-Kabók & Michael Ochsner	
Peer review in the context of the new modes of knowledge production, dissemination and evaluation	87
By Marc Vanholsbeeck	
Peer review in the context of Open Science: towards more open peer reviewing practices?	94
By Marc Vanholsbeeck	
PART V: Conclusion	99
By Michael Ochsner, Nina Kancewicz-Hoffman, Jon Holm & Marek Hołowiecki	
Bibliography	101
ENRESSH publications.....	101
Full bibliography	103

PART I: General Framework: state of the art of peer review in the SSH in Europe

Introduction: Aim and scope of the report

By Nina Kancewicz-Hoffman, Michael Ochsner, Marek Hołowiecki & Jon Holm

The evaluation of research is a delicate issue in the social sciences and humanities. There is a lot of opposition from SSH scholars against evaluation procedures. This has several reasons, for the humanities they were subsumed to four main points (Hug et al., 2014): a) the methods originate from the natural and life sciences, b) lacking consensus on quality criteria, c) fear of negative steering effects, and d) reservations against (simple) quantification. While these points are evident in the case of the use of bibliometric approaches, they are also relevant when it comes to peer review.

Regarding the first point, peer review has been developed in the context of journal articles. Both, social sciences and humanities have a much wider range of publication outputs, such as books, reports, feuilletons and exhibitions, many of which do not only address a scientific public. Secondly, the SSH are not characterised by a strong internal organisation protecting and promoting what is considered a mainstream approach (van den Akker, 2016) nor by the idea of linear progress (Lack, 2008). Rather, SSH disciplines are diverse with regard to theoretical and empirical approaches. There is often no consensus for quality criteria even within one discipline because different paradigms compete with and enrich each other (see, e.g., van den Akker, 2016; Mallard et al., 2009). Thirdly, negative steering effects, such as conservatism, strengthening of old-boys' networks, discrimination of marginal topics or approaches become an issue in diverse research environments. Finally, often peers need to rate objects of evaluation and thus create measurements that are not necessarily valid but have an impact beyond the evaluation exercise. These are just some of the issues regarding peer review in the social sciences and humanities that need to be addressed in a context when peer review regains more importance in research evaluation.

In general, peer review appears to be enjoying a come-back as the preferred means of assessment in different evaluation situations after years of fascination with bibliometrics and quantitative methods. Three main drivers can be identified behind this return: Bibliometrics and other quantitative methods have proven inadequate for the evaluation of SSH research (Hicks, 2004; Nederhof, 2006; Ochsner et al., 2012) and showed adverse steering effects, also for the STEM fields, which shifted the balance to the critical voices against (simplistic) bibliometric applications across all fields (see, e.g., Burrows, 2012; Hammarfeldt, 2017; Lawrence, 2002; MacRoberts & MacRoberts, 2017; Molinié & Bodenhausen, 2010). The DORA declaration is a further result of this critical examination of bibliometric evaluation methods. Second, there is a tendency on the part of policy makers and funders to apply the same evaluation procedures, methods and sometimes even criteria for all research domains, STEM and SSH. This tendency is partly a response to expectations on the part of the STEM community and linked to the rise of the importance of interdisciplinary projects (see, e.g., König, 2016). Third, the open science movement includes new ways of publishing, reviewing and new forms of metrics. These developments concern STEM and SSH disciplines alike and the STEM are already defining the discourse. It is therefore vital for the SSH

community to document, understand and monitor the practice of peer review and its relation to evaluation and research practice to make sure that these new developments can be used for the benefit of SSH disciplines.

The aim of this interim report is to describe conceptual and practical issues of peer review in SSH in Europe. The discussion of these issues is rooted in a subgroup of Work Group 1 within the COST-Action CA-15137 ENRESSH that has investigated peer review processes and practices.

The scope: Peer Review is an intrinsic part of evaluation process. It is difficult to analyse without considering it in the context of evaluation in a wider sense. In the report, we focus on those aspects of evaluation which are related to peer review.

The report is addressed to policy makers, reviewers but also to the research community at large – in the SSH and beyond – as it enables all stakeholders to identify issues requiring attention and further inquiry.

The report is structured as follows: Part I starts with setting a general framework for the report. Derrick and Ross-Hellauer remind us that the past has shown that automatically copying evaluation procedures from STEM disciplines did not always work out well and that SSH needs to develop its own theoretical approaches and practical applications. Therefore, they warn in their chapter against a colonisation of SSH by STEM values and notions of quality and pledge for a conscious re-appropriation of evaluation procedures adequate for the SSH.

The following set of chapters addresses issues specifically relevant for SSH (Part II). Ochsner focuses on approaches for recognising scientific quality in SSH via peer review and proposes specific recommendations: Reviewers should rate the objects of evaluation across a broad range of criteria rather than giving a holistic judgement. Also, the criteria should clearly differentiate between criteria for scientific quality and criteria concerning policy goals or relation to or impact on society. Derrick focuses on the evaluation of societal impact and related biases within the evaluation process that can act against the promotion of SSH research and calls for more research on how models and definitions of impact are constructed in practice. Kancewicz-Hoffman and Pölönen indicate that acknowledging multilingualism in research communication and collaboration is a condition for adequate peer review and stress a need for more research on practices of peer review in multilingual contexts.

Part III focuses on current practices of peer review. In her chapter, Kancewicz-Hoffman reviews a selection of recent documents setting out practices of peer review and identifies main challenges for SSH concluding that more systematic observation of peer review practices in SSH is necessary. The following chapters address some of those challenges. Pölönen, Engels and Guns show in their study that there is ambiguity in what is seen as a “peer-reviewed” publication: This ambiguity not only concerns self-reported lists of publications but also PRFS that are based on formal criteria employed to publication lists taken from information systems. In a short overview based on a large survey of National Evaluation Systems in Europe, Ochsner indicates the centrality of peer review in all instances of evaluation. He shows that the role, form, and significance of peer review, however, can differ strongly across different evaluation procedures and between countries, in most cases independently from the type of evaluation procedure. Hug and al. report work in progress on a systematic review of studies on peer review criteria on manuscript and grant

applications. Their preliminary findings indicate a need for more research on peer review criteria in general and in the natural sciences and humanities in particular. They also stress the importance of a comparative perspective on the evaluation cultures of different research disciplines. Giménez-Toledo focuses on peer review of scholarly books and its differences with peer review in journals; she concludes that although peer review is one among a number of the selection procedures applied by book publishers, it does ensure scientific quality. Peruginelli, Sanz-Casado and Stojanovski present three national surveys of peer review procedures in law journals pointing out to similarities in approach to peer review in the three countries. They note a common tendency for a more rigorous and transparent peer review process and note an opportunity for open, “social” peer review created by technological innovations.

Part IV addresses recent challenges to peer review related to geopolitical and gender perspectives, interdisciplinarity and the open science turn. Lendák-Kabók and Ochsner report on gender and geopolitical perspectives on peer review and investigate how male and female Early Career Investigators from Eastern and Western Europe perceive and react to peer review. Vanholsbeeck discusses challenges of SSH integration to European impact-driven interdisciplinary research referring to the concept of exoterisation of research. Reporting on a study on senior SSH scholars’ attitudes towards research evaluation, he shows that although, in their opinion, peer review keeps its importance as an evaluation tool, they are at the same time aware of complexities of peer review practices in the context of current policies related to interdisciplinarity, internationalisation and societal impact. He recommends educating researchers in new skills needed in the conduct of more open, gender sensitive and interdisciplinary peer reviewing practices. Finally, Vanholsbeeck discusses the current status of Open Peer Review.

The report ends with conclusions drawn from the chapters and the combined bibliography in Part V.

References

- Burrows, R. (2012). Living with the H-Index? Metric assemblages in the contemporary academy. *The Sociological Review*, 60(2), 355–372. <https://doi.org/10.1111/j.1467-954X.2012.02077.x>
- Hammarfelt, B. (2017). Four claims on research assessment and metric use in the humanities. *Bulletin of the Association for Information Science and Technology*, 43(5), 33–38. <https://doi.org/10.1002/bul2.2017.1720430508>
- Hicks, D. (2004). The four literatures of social science. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 473–496). Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-2755-9_22
- Hug, S. E., Ochsner, M., & Daniel, H.-D. (2014). A framework to explore and develop criteria for assessing research quality in the humanities. *International Journal for Education Law and Policy*, 10(1), 64.
- König, T. (2016). Peer review in the social sciences and humanities at the European level: The experiences of the European Research Council. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research assessment in the humanities. Towards criteria and*

- procedures* (pp. 151–163). Springer International Publishing.
https://doi.org/10.1007/978-3-319-29016-4_12
- Lack, E. (2008). Einleitung – Das Zauberwort “Standards”. In E. Lack & C. Markschie (Eds.), *What the hell is quality? Qualitätsstandards in den Geisteswissenschaften* (pp. 9–34). Campus.
- Lawrence, P. A. (2002). Rank injustice. *Nature*, 415(6874), 835–836.
<https://doi.org/10.1038/415835a>
- MacRoberts, M. H., & MacRoberts, B. R. (2018). The mismeasure of science: Citation analysis. *Journal of the Association for Information Science and Technology*, 69(3), 474–482. <https://doi.org/10.1002/asi.23970>
- Mallard, G., Lamont, M., & Guetzkow, J. (2009). Fairness as appropriateness: Negotiating epistemological differences in peer review. *Science, Technology, & Human Values*, 34(5), 573–606. <https://doi.org/10.1177/0162243908329381>
- Molinié, A., & Bodenhausen, G. (2010). Bibliometrics as weapons of mass citation / La bibliométrie comme arme de citation massive. *CHIMIA International Journal for Chemistry*, 64(1), 78–89. <https://doi.org/10.2533/chimia.2010.78>
- Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, 66(1), 81–100.
<https://doi.org/10.1007/s11192-006-0007-2>
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2012). Indicators for research quality in the humanities: Opportunities and limitations. *Bibliometrie - Praxis und Forschung*, 1, 4.
<https://doi.org/10.5283/bpf.157>
- van den Akker, W. (2016). Yes we should; Research assessment in the humanities. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research assessment in the humanities. Towards criteria and procedures* (pp. 23–29). Springer International Publishing.
https://doi.org/10.1007/978-3-319-29016-4_3

Peer Review in SSH: In Need of Development?

By Gemma Derrick & Tony Ross-Hellauer

Peer review, whether as a political tool or one to facilitate academic self-governance, is a powerful driver of knowledge production. As its primary role of ensuring the validity and quality of research, it has been used in a variety of settings including: pre-publication evaluation of scientific manuscripts; decision making of grant applications; the assessment of research departments (such as used in national audit exercises); reviews of research disciplines by funding councils; and as a method of international benchmarking. In all these situations, the operationalisation of peer review is different with group-peer review situations that rely on the benefits of open deliberation by a range of research and non-academic experts (Derrick, 2018); and scientific manuscript being a blinded process negotiated by one single actor, the editor (Derrick & Ross-Hellauer, forthcoming). Considering these differences, here we concentrate on peer review for scholarly manuscripts only and how its conceptualisation and operationalisation relates to SSH.

Peer review of scholarly manuscripts is the formal quality assurance mechanism whereby works are made subject to the scrutiny of others, whose feedback and judgements are then used to improve them and make final decisions regarding selection for publication. We can discern five distinct functions that peer review is used to perform: (1) give feedback/guidance for improvement, (2) judge soundness/robustness, (3) judge novelty, quality and/or potential impact, (4) judge suitedness for venue, (5) make recommendations regarding publication. These processes need not be entwined, however. For example, some journals have adopted models where reviewers are asked to focus only on technical soundness rather than perceived importance (e.g., PLOS ONE and PeerJ). At other venues, post-publication review is applied so that recommendations regarding publication no longer apply (e.g., F1000Research). Peer review is usually single or double-blind. In SSH, in contrast to STEM, double-blind review is most common for journals, although single-blind is more usual for books and monographs (British Academy, 2007, p. 10). Various models of open peer review have been proposed and applied in various contexts (Ross-Hellauer, 2017). A great variety of possible models and factors are possible, but usually under this name is understood review where reviewer identities are revealed to authors (open identities) and/or reviews are published alongside publications (open reports). An alternative understanding, which seems more prevalent in SSH (Ross-Hellauer, 2017) is of “open participation” where reviews are “crowdsourced” via the Web. These models of innovation seem to be more prevalent at STEM venues than in SSH. The traditional system of review has been subject to various criticisms for concerns of bias, unnecessary delay and unreliability. Its general sustainability is also questioned – the challenges of ever-increasing levels of knowledge production, quantified just in sheer number of publications, combined with the fact that peer review is a non-rewarded activity for most researchers, means that finding and engaging reviewers is a major challenge for journal editors (Lotriet, 2012, p. 27).

That innovation and research in peer review has traditionally been driven by STEM concerns, and viewed through a STEM lens, is problematic because a review of the literature reveals that publication and peer review practices differ in SSH in important ways. Before discussing these differences, we should be clear that SSH is not homogenous – there are important epistemological and methodological differences amongst subjects. Some

disciplines, for example, are more science-like (e.g., economics, psychology) in using hypothesis-driven methods and somewhat positivistic epistemologies – these disciplines can tend to exhibit more STEM-like publication patterns (article publications in English-language international journals). However, in other SSH subjects, especially in the humanities, much more value is placed on books as the primary mode of research communications. Where articles are preferred, SSH subjects make more use of local journals and regional languages (Kulczycki et al., 2018). SSH researchers also tend to produce fewer articles, and SSH articles tend to have fewer co-authors (Ossenblok et al., 2012; Ossenblok, et al., 2014). In contrast to STEM where high submission rates often mean high rates of desk-rejection, in SSH, relatively fewer articles are refused prior to peer review (British Academy, 2007, p. 10). Times taken to review (Huisman & Smits, 2017, pp. 641–642) and from submission to publication (Björk & Solomon, 2013, p. 914) are both substantially longer than in most STEM fields however. Considering that manuscripts are estimated to be submitted to between three to six journals prior to appearing in their venue of publication (Azar, 2004), it is worth considering the extent to which extended rejection, re-submission and eventual publication cycles in SSH delay progress. This should of course be viewed in terms of the pace of conversation in many SSH fields. In contrast to fast-moving empirical sciences, SSH often have slower rates of citation after publication, and the true impact of works can be much longer in revealing itself (British Academy, 2007, p. 7).

As said, monographs play a much greater role in many SSH subjects (Giménez-Toledo et al., 2019). Peer review of books should not be assumed to mirror that of journal publications, although unfortunately there is a dearth of material regarding this (although this field seems to be emerging, see, e.g., Kulczycki, et al., 2018). We can say that peer review for books is also less formal in terms of processes and criteria, with variation according to whether peer review is applied (if it is applied) to the book proposal, individual chapters or full manuscript; whether review is by independent peers, editorial boards, or acquisition editors, whether commercial as well as academic criteria are taken into account, and who is responsible for organising review (publisher, series or book editor, academic board) (Derricourt, 2012; Verleysen & Engels, 2013).

Investigations of attitudes towards peer review amongst researchers generally find that peer review is highly-valued in general, but not without reservations (Mulligan, Hall, & Raphael, 2013; Ross-Hellauer et al., 2017; Rowley & Sbaffi, 2018; Ware, 2008). Rowley and Sbaffi's (2018) study was unique in closely examining the role of discipline and found that SSH scholars were less likely than their STEM counterparts to believe peer review can judge novelty or importance, detect plagiarism or fraud, detect factual inaccuracies, determining an article's fit to the journal. They were also more likely to agree about the existence of bias towards authors based on gender, world-region and level of seniority.

Such differences may reflect differences in the purposes of peer review across disciplines. It can be argued that there are important epistemological differences in styles of knowledge generation that mean that the judgements involved in SSH are often of a different kind in comparison to that in STEM. Human behaviour and human experience have levels of complexity beyond that of a drug trial or a physics experiment, where variables can be strictly controlled and success or otherwise declared. Even in more empirically oriented fields like economics, sociology or psychology, factors like small sample sizes (of biased composition) or the confounding factors of field work can mean that the degrees of interpretative

freedom are higher in SSH than in STEM. This means that strictly positivist criteria for success, including criteria like strict methodological reporting or reproducibility, risk privileging only that which is most STEM-like about SSH, and devaluing other areas. In terms of peer review, such quasi-positivism could endanger sympathetic formulative assistance. This will be especially true in a context of over-reliance on interactional expertise, where a “peer’s” expertise might still involve different preconceptions about what knowledge is. Mallard, Lamont and Guetzkow (2009) interviewed 81 panellists serving on five multidisciplinary SSH fellowship competitions and found four distinct epistemological styles guiding decision-making and that conflicts arose when one style was applied in evaluating a proposal which adopted another. Such fundamental epistemological suppositions will impact how results are valued, how meaning is interpreted, and hence reviewer evaluations. They concluded that reviewers should use “cognitive contextualization,” adopting “epistemological styles most appropriate to the field or discipline of the proposal under review.” Where this is not possible, we should not see the value of a reviewer’s work as nil, but we might nonetheless require a declaration of epistemological Conflict of Interest, as is suggested by Shimp (2004).

In simple terms, peer review of SSH material is often not engaged in careful checking of the correctness of procedure and theory in an experimental setting, as can be the case in STEM. This is not to say that interpretation and judgements are not also intimate parts of even the hardest of sciences, but that the interpretive flexibility of those determinations can be said to lie within a smaller range. Human behaviour and experiences, not to mention the myriad ways in which they can be manifested, are levels of complexity higher. As Derricourt says, “in softer social sciences and a wider range of the humanities, the questions on a submitted paper might be whether it fits into the conventions of the discipline, whether the reviewer agrees or disagrees with the approach and argument, and how important or interesting the reviewer finds it. These are more editorial questions than questions of authentication.” (Derricourt, 2012, p. 145)

These elements are all examined in depth in a forthcoming publication by the current authors, which stems from an ENRESSH Short Term Scientific Mission project “Peer review in SSH: in need of development”, which explores the suitability of current peer review, and demands on peer review, for SSH disciplines (Derrick & Ross-Hellauer, forthcoming). It conceptualises peer review as an act of boundary-work found necessary to demarcate scientific knowledge, which required the formalisation of a reviewer function, acting as an expert. However, it also conceptualises the STEM-SSH divide, not as a categorical distinction, but as a fluid spectrum that runs parallel to a scale of the object of study expressed by Flyvbjerg’s (2001) distinction between objects to humanistic foci of study. This spectrum allows for the existence of peripheral overlaps between fields of study and therefore a mechanism by which regulatory advice through peer reviewers required as part of the peer review system acts instil forms of expertise that are not central to the missions and values of SSH research. In fact, we argue that this sharing different forms of expertise within this spectrum, due to the growth of knowledge production (Yan, 2016), interdisciplinary research (van Noorden, 2015) and the unsustainable nature of peer review (Ross-Hellauer et al., 2017), constitutes a form of gradual colonisation of SSH by STEM values and notions of quality. Left unchecked, the black box nature of peer review catalyses the colonisation of SSH, resulting in a systematic devaluation that forces SSH researchers to

submit to, and adopt rather than consciously and openly assess notions of excellence offered by reviewers as part of the scholarly peer review process. In this way, the existing peer review system is feeding a *Teufelskreis/vicious cycle* that alters how SSH can self-govern and regulate notions of quality and value independent of STEM via the peer review process.

References

- Azar, O. H. (2004). Rejections and the importance of first response times. *International Journal of Social Economics*, 31(3), 259–274. <https://doi.org/10.1108/03068290410518247>
- Björk, B.-C., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. *Journal of Informetrics*, 7(4), 914–923. <https://doi.org/10.1016/j.joi.2013.09.001>
- British Academy. (2007). *Peer Review: The challenges for the humanities and social sciences*. <https://www.thebritishacademy.ac.uk/publications/peer-review-challenges-humanities-and-social-sciences>
- Derrick, G. (2018). Impact from the evaluators' eye. In G. Derrick, *The evaluators' eye* (pp. 1–19). Springer International Publishing. https://doi.org/10.1007/978-3-319-63627-6_1
- Derrick, G. & Ross-Helauer, T. (forthcoming). *Decolonising the social sciences and humanities through peer review*.
- Derricourt, R. (2012). Peer review: Fetishes, fallacies, and perceptions. *Journal of Scholarly Publishing*, 43(2), 137–147. <https://doi.org/10.1353/scp.2012.0000>
- Flyvbjerg, B. (2001). *Making social science matter: Why social inquiry fails and how it can succeed again*. Cambridge Univ. Press.
- 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>
- Huisman, J., & Smits, J. (2017). Duration and quality of the peer review process: The author's perspective. *Scientometrics*, 113(1), 633–650. <https://doi.org/10.1007/s11192-017-2310-5>
- Kulczycki, E., Engels, T. C. E., Pölönen, J., Bruun, K., Dušková, M., Guns, R., Nowotniak, R., Petr, M., Sivertsen, G., Istenič Starčič, A., & Zuccala, A. (2018). Publication patterns in the social sciences and humanities: Evidence from eight European countries. *Scientometrics*, 116(1), 463–486. <https://doi.org/10.1007/s11192-018-2711-0>
- Lotriet, C. J. (2012). Reviewing the review process: Identifying sources of delay. *Australasian Medical Journal*, 5(1), 26–29. <https://doi.org/10.4066/AMJ.2012.1165>

- Mallard, G., Lamont, M., & Guetzkow, J. (2009). Fairness as appropriateness: Negotiating epistemological differences in peer review. *Science, Technology, & Human Values*, 34(5), 573–606. <https://doi.org/10.1177/0162243908329381>
- Mulligan, A., Hall, L., & Raphael, E. (2013). Peer review in a changing world: An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology*, 64(1), 132–161. <https://doi.org/10.1002/asi.22798>
- Ossenblok, T. L. B., Engels, T. C. E., & Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science: A comparison of publication patterns and incentive structures in Flanders and Norway (2005-9). *Research Evaluation*, 21(4), 280–290. <https://doi.org/10.1093/reseval/rvs019>
- Ossenblok, Truyken L.B., Verleysen, F. T., & Engels, T. C. E. (2014). Coauthorship of journal articles and book chapters in the social sciences and humanities (2000-2010): Coauthorship of journal articles and book chapters in the social sciences and humanities (2000-2010). *Journal of the Association for Information Science and Technology*, 65(5), 882–897. <https://doi.org/10.1002/asi.23015>
- Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research*, 6, 588. <https://doi.org/10.12688/f1000research.11369.2>
- Ross-Hellauer, T., Deppe, A., & Schmidt, B. (2017). Survey on open peer review: Attitudes and experience amongst editors, authors and reviewers. *PLOS ONE*, 12(12), e0189311. <https://doi.org/10.1371/journal.pone.0189311>
- Rowley, J., & Sbaiffi, L. (2018). Academics' attitudes towards peer review in scholarly journals and the effect of role and discipline. *Journal of Information Science*, 44(5), 644–657. <https://doi.org/10.1177/0165551517740821>
- Shimp, C. (2004). Scientific peer review: A case study from local and global analyses. *Journal of the Experimental Analysis of Behavior*, 82(1), 103–116. <https://doi.org/10.1901/jeab.2004.82-103>
- van Noorden, R. (2015). Interdisciplinary research by the numbers. *Nature*, 525(7569), 306–307. <https://doi.org/10.1038/525306a>
- Verleysen, F. T., & Engels, T. C. E. (2013). A label for peer-reviewed books. *Journal of the American Society for Information Science and Technology*, 64(2), 428–430. <https://doi.org/10.1002/asi.22836>
- Ware, M. (2008). Peer review in scholarly journals: Perspective of the scholarly community – Results from an international study. *Information Services & Use*, 28(2), 109–112. <https://doi.org/10.3233/ISU-2008-0568>
- Yan, E. (2016). Disciplinary knowledge production and diffusion in science. *Journal of the Association for Information Science and Technology*, 67(9), 2223–2245. <https://doi.org/10.1002/asi.23541>

PART II: Issues and Discussions Specific to and Most Relevant for SSH Peer Review

Evaluation Criteria and Methodology

By Michael Ochsner

Introduction

Peer review is the most important means for the assessment of academic research in the SSH and serves to decide which works, proposals or careers are funded or whether research or a career is evaluated as successful or excellent. Yet, while there are many studies on the potential biases and subjectivity of peer review (see, e.g., Bornmann et al., 2008; Bornmann et al., 2011; Lamont, 2009; Langfeldt, 2006), less is known about what quality of research means and how peers (can) identify it. Knowledge on these issues will help to improve peer review practices. While there are many aspects relevant for a successful peer review process, such as peer selection, technical support of peers, organisation of peer review (blind, double-blind, panel decision etc.), this chapter will focus on the aspect of how quality can best be recognised in peer review. It is structured as follows: the first part gives an overview of concepts of research quality in the SSH, as this is the concept peer review is supposed to “measure”; the second part focuses on assessment and issues related to peer review as an instrument for the evaluation of research; the last part will draw conclusions and gives recommendations for the peer review process regarding evaluation criteria and methods.

What is Quality?

Lamont’s book “How professors think” (Lamont, 2009) describes how experts take the role of gatekeepers – or even of “guardians of science”, as Daniel (1993) put it. A citation of an interviewee of Lamont summarises how peer reviewers most often judge a work or career: “There are different [...] kinds of excellence [but I am] pretty confident that I’d know it when I see it” (Lamont, 2009, 159). Even though scholars judge the quality of their colleagues’ and students’ work on a daily basis, the knowledge about what quality is remains tacit. Just as car drivers cannot explain how they got out of a complicated situation they managed routinely, the experts cannot tell exactly how they judge a research or a career (for tacit knowledge, see Polanyi, 1967). Yet, unlike the car, in a situation of deciding upon careers or scarce funding, it is important to have a clear rationale for a fair and just evaluation. However, while there is an abundance of literature on concepts of quality in higher education or of research quality in research funding procedures (for an overview, see for example Langfeldt & Scordato, 2016), there are only a few studies that investigate systematically what characterizes “good” research from the point of view of those who can judge best what “good” research is: the scholars themselves. This is especially true for the SSH. An overview on projects on SSH scholars’ notions of research quality in European countries conducted by Work Group 1 of the ENRESSH COST-Action showed that there is a growing interest in understanding what research quality is, but there are only a few studies that investigate in a bottom-up manner how researchers understand and identify the quality of SSH research (see Ochsner, Hug & Galleron, 2017). Yet, when academic peer review is making judgements about the quality of research, there should be an understanding of what

quality means, going beyond generic terms like *plausibility*, *scientific value* and *originality* (Polanyi, 1967) or the – almost congruent – criteria from the UK’s RAE 2008, *rigour*, *significance* and *originality* that proved to be very imprecise in practice (Johnston, 2008). Research on scholars’ notions of quality shows that research quality is a complex, multi-dimensional construct (see, e.g., Bazeley, 2010; Hemlin, 1993; Hug et al., 2013; Ochsner & Dokmanović; Ochsner et al., 2013). Reducing research quality to some aspects might lead to adverse effects, such as goal displacement (see, e.g., Butler, 2007; De Rijke et al., 2016; Sousa & Brennan, 2014) or task reduction (de Rijke et al., 2016). Such behavioural changes have mainly been investigated regarding the use of indicators. But also peer review is based on criteria even though they remain often tacit or are reduced to policy goals and scholars try to anticipate how they will be evaluated adapting their behaviour accordingly. Furthermore, even if formal indicators are used to standardise peer judgement or render it more objective, they can be used in selective ways by peers in evaluations to support the decision made without them (see, e.g., Gozlan, 2016).

Some often-used criteria in evaluations are not criteria for scientific quality but rather for a way of doing research as research on scholars’ notions of quality shows: Interdisciplinarity, collaborative research, internationality and societal impact are not seen as indicative of scientific quality but rather of a modern way of doing research. It can be of high as well as of low quality (Ochsner et al., 2013). Hence, it is important to distinguish two types of evaluation criteria: criteria applied by research funders and research policy makers linked to strategic decisions on the one hand and criteria reflecting scientific quality on the other. Yet, both types of criteria are often used as quality criteria, also in peer review processes (see, e.g., Wissenschaftsrat, 2004; Krull & Tepperwien, 2016). Such a confounding of different types of evaluation criteria based on different quality conceptions between evaluators and scholars might lead to communication issues and to opposition against evaluation procedures as well as to bias in peer review, i.e. if peers should evaluate interdisciplinary research but in their review they judge the research according to their disciplinary standards (see Langfeldt, 2006). Rather, the different types of criteria should be evaluated separately so that the different nature of the criteria becomes transparent.

Societal impact is a special case as it is not directly related to research quality but to another aim of research, i.e. to lead to effects outside of academia, and thus should be evaluated separately (see, e.g., Koninklijke Nederlandse Akademie van Wetenschappen [KNAW], 2011; VSNU, NWO & KNAW, 2016). It has its own quality criteria and maybe even the experts or peers might be different, for example proponents of civil society, politics or the economy. Yet, there is not much knowledge on how peers evaluate societal impact, or even how they define it. Rather, scholars seem to prefer evaluating research quality over societal impact (Albert et al., 2012) and when they need to evaluate societal impact, they are even less confident about what it means (Derrick & Samuel, 2017).

Assessment of Research by Peer Review

If research quality is a complex, multidimensional construct, how can it be assessed in practice? In the last years, several initiatives investigated how the quality of SSH research – or sometimes research performance in the SSH – could be assessed adequately (see, e.g., Andersen, 2013; Giménez-Toledo et al., 2013; Gogolin & Stumm, 2014; Ochsner et al., 2016, 2017; for an overview Ochsner, Galleron & Ionescu, 2017). The suggested

assessment procedures usually involve SSH scholars in the process either as experts in peer review systems or as experts in defining output lists (e.g. publication lists or list of publishers). This involvement of peers in the process is important because only about 50% of the relevant quality criteria for humanities research can be adequately measured with indicators (Ochsner et al., 2012) and likely this is similar for the social sciences. If peers are involved, however, it is important that the process assures a fair assessment, i.e. an equal treatment using the same rules for all units to be assessed. Peer review as a method for research evaluation has been criticised and there are several studies on different biases of peer review, such as low interrater reliability, mediocre predictive validity, factors other than scientific quality like gender or institution of the applicant explaining outcome etc. (see, e.g., Bornmann & Daniel, 2008a; Bornmann et al., 2008, 2010; Mutz et al., 2015; Tamblyn et al., 2018). However, there are several methodological issues involved with these criticisms of the peer review process (see also Langfeldt et al., 2015): All of them compare outcomes without having a clear concept of what the outcome is. First, it is not clear whether a high interrater reliability is desirable as it might be a result of unfortunate choice of experts who follow the same paradigm and might not accept research drawing from another paradigm. Without investigating how the peers arrived at the different ratings, the result is not indicative of the quality of the peer review process. Second, high predictive validity, usually measured as difference in citation rates from articles published compared to rejected but published elsewhere or comparing citation impact from scholars having received a certain funding compared to those having not received the funding, might just be an effect of being published in the specific journal or having received the prestigious funding. The outcome indicator citations is linked to and dependent on many things (see Bornmann & Daniel, 2008b) and is in most cases not a valid measure for a functioning peer review process, certainly not in the SSH. Third, a bias that has been identified in a peer review process might not be the problem of the peer review but of conditions external to the peer review: e.g. researchers at prestigious institutions might have more time to write a proposal, women might be less self-confident and submit understated proposals; because men work more often at prestigious institutions, both conditions would lead to a bias towards men even though the peer reviewers would not favour men as such (see, e.g., Ceci & Williams, 2011; Enserink, 2015).

More important for the quality of the peer review process is the *intrarater* reliability: the likelihood that the same reviewer assigns the same score to the same application at different points in time, similar to a test-retest reliability, e.g. a reviewer's rating is independent of the order the reviewer reads different proposals (see Ochsner, Hug & Daniel, 2017). Thorngate et al. (2009) conclude their comprehensive research on decision making by stating that merit should be judged separately along specified criteria in order to achieve consistent results. So-called "holistic" judgements (i.e., "I know it when I see it") apply different weighting functions to different applicants, which opens the door widely for double standards and biases (Thorngate et al., 2009, p. 26). Furthermore, focusing on a broad range of criteria helps to avoid preferring aspects with similar gradings and neglecting aspects that follow a different pattern; people tend to look for "consistency", but an assessment based on redundant information is always inferior to an assessment based on more information, as Tversky and Kahnemann (1974) explain in a more general context: "Highly consistent patterns are most often observed when the input variables are highly redundant or

correlated. Hence, people tend to have great confidence in predictions based on redundant input variables. However, an elementary result in the statistics of correlation asserts that, given input variables of stated validity, a prediction based on several such inputs can achieve higher accuracy when they are independent of each other than when they are redundant or correlated. Thus, redundancy among inputs decreases accuracy even as it increases confidence [...] (Tversky & Kahnemann, 1974, p. 1126). For example, if a reviewer finds that an article is well-written and presents interesting findings but has some doubts about the methodological rigour, it is likely that the two consistent evaluations of “well-written” and “interesting findings” will dominate his or her final holistic judgement while the “methodological issues” will go under the radar. The more informative combination would be the “well-written” but “methodologically not rigorous”, simply because the category of “well-written” might evoke the category “interesting results”, simply because the writing style makes the results look interesting. If each criterion is judged separately, “methodological issues” are less likely go under the radar as the criteria catalogue will draw attention to it. Additionally, explicit criteria serve transparency: which criteria were used for the judgement and how were they weighted? Feedback on criteria might then help unsuccessful authors or applicants improve their next submissions. All these points are important for the judgment of merit to be fair and consistent (Thorngate et al., 2009), i.e. that all applications, manuscripts or other objects of evaluation are assessed according to the same standards. This will prevent different forms of bias, such as conservative, gender or institution bias. By providing a framework for consistent judgement of research quality across peers, it also helps preventing the more technical issues such as low interrater reliability or low predictive validity. A rating across multiple criteria helps disentangling differences between weightings of different criteria from different judgements on a single criterion. It will also show that low predictive validity (measured by citations) might not be the best quality measure for a peer review procedure by showing why an output or project was selected, while citations are linked to many other things than quality, e.g. a mainstream topic or size of institution.

Conclusions

In the SSH, peer review is the most common and most important way to assess outputs, careers, projects or institutions. While there are many aspects relevant for a successful peer review process, this chapter focused on the aspect of how scientific quality can best be recognised in peer review, one of the most important aims of academic peer review. Despite its general acceptance as an assessment method for many evaluation situations, peer review faces some criticisms, such as low interrater reliability, mediocre predictive validity and different kind of biases (conservative bias, gender bias, institutional bias). This chapter argues that these biases are linked to the fact that there is no clear methodology that links the concept “scientific quality” with the procedure. While research shows that scientific quality is a complex and multidimensional construct, this complexity is rarely taken into account in review practices. At the same time, research on decision making shows that merit should be judged separately along specified criteria in order to achieve consistent and fair results. Holistic approaches to evaluation open the doors widely for different biases. Therefore, peer review processes should include a catalogue of explicit criteria that guide the judgements of the peers. Each criterion should be rated separately. This has also the

advantage that indicators can be assigned to specific criteria for which they can provide additional information to peers (informed peer review), which can increase the acceptability for the use of indicators among scholars and might reduce subjectivity (see Ochsner et al., 2014). Also, criteria for scientific quality and criteria for policy goals, such as interdisciplinarity or societal impact, should be evaluated separately.

References

- Albert, M., Laberge, S., & McGuire, W. (2012). Criteria for assessing quality in academic research: The views of biomedical scientists, clinical scientists and social scientists. *Higher Education*, 64(5), 661–676. <https://doi.org/10.1007/s10734-012-9519-2>
- Bazeley, P. (2010). Conceptualising research performance. *Studies in Higher Education*, 35(8), 889–903. <https://doi.org/10.1080/03075070903348404>
- Bornmann, L., & Daniel, H.-D. (2008a). The Effectiveness of the Peer Review Process: Inter-Referee Agreement and Predictive Validity of Manuscript Refereeing at *Angewandte Chemie*. *Angewandte Chemie International Edition*, 47(38), 7173–7178. <https://doi.org/10.1002/anie.200800513>
- Bornmann, L., & Daniel, H.-D. (2008b). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80. <https://doi.org/10.1108/00220410810844150>
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2008). How to detect indications of potential sources of bias in peer review: A generalized latent variable modeling approach exemplified by a gender study. *Journal of Informetrics*, 2(4), 280–287. <https://doi.org/10.1016/j.joi.2008.09.003>
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). A Reliability-Generalization Study of Journal Peer Reviews: A Multilevel Meta-Analysis of Inter-Rater Reliability and Its Determinants. *PLoS One*, 5(12), e14331. <https://doi.org/10.1371/journal.pone.0014331>
- Bornmann, L., Mutz, R., Marx, W., Schier, H., & Daniel, H.-D. (2011). A multilevel modelling approach to investigating the predictive validity of editorial decisions: do the editors of a high profile journal select manuscripts that are highly cited after publication? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(4), 857–879. <https://doi.org/10.1111/j.1467-985X.2011.00689.x>
- Butler, L. (2007). Assessing University Research: A Plea for a Balanced Approach. *Science and Public Policy*, 34(8), 565–574. <https://doi.org/10.3152/030234207X254404>
- Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, 108(8), 3157–3162. <https://doi.org/10.1073/pnas.1014871108>
- Daniel, H.-D. (1993). *Guardians of science: Fairness and reliability of peer review*. VCH Verlagsgesellschaft. <https://doi.org/10.1002/3527602208>

- 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>
- Derrick, G., & Samuel, G. (2017). The future of societal impact assessment using peer review: pre-evaluation training, consensus building and inter-reviewer reliability. *Palgrave Communications*, 3, palcomms201740. <https://doi.org/10.1057/pal-comms.2017.40>
- Enserink, M. (2015). Dutch sexism study comes under fire. *Science*, 364(6446). <https://doi.org/10.1126/science.aad4618>
- Giménez-Toledo, E., Tejada-Artigas, C., & Mañana-Rodríguez, J. (2013). Evaluation of scientific books' publishers in social sciences and humanities: Results of a survey. *Research Evaluation*, 22(1), 64–77. <https://doi.org/10.1093/reseval/rvs036>
- Gozlan, C. (2016). Les sciences humaines et sociales face aux standards d'évaluation de la qualité académique. *Sociologie*, 7(3), 261–21. <https://doi.org/10.3917/socio.073.0261>
- Hemlin, S. (1993). Scientific quality in the eyes of the scientist. A questionnaire study. *Scientometrics*, 27(1), 3–18. <https://doi.org/10.1007/BF02017752>
- 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>
- Johnston, R. (2008). On structuring subjective judgements: Originality, significance and rigour in RAE2008. *Higher Education Quarterly*, 62(1-2), 120–147. <https://doi.org/10.1111/j.1468-2273.2008.00378.x>
- Koninklijke Nederlandse Akademie van Wetenschappen. (2011). *Quality Indicators for Research in the Humanities*. KNAW, Royal Netherlands Academy of Arts and Sciences. https://www.knaw.nl/en/news/publications/quality-indicators-for-research-in-the-humanities/@@download/pdf_file/20111024.pdf
- Krull, W., & Tepperwien, A. (2016). The four “I”s: Quality indicators for the humanities. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research Assessment in the Humanities. Towards criteria and procedures* (pp. 165–179). Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_13
- 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., Bloch, C. W., & Sivertsen, G. (2015). Options and limitations in measuring the impact of research grants – evidence from Denmark and Norway. *Research Evaluation*, 24(3), 256–270. <https://doi.org/10.1093/reseval/rvv012>

- Langfeldt, L., & Scordato, L. (2016). *Efficiency and flexibility in research funding. A comparative study of funding instruments and review criteria*. NIFU Report No. 2016:9. Oslo: NIFU.
- Mutz, R. D., Bornmann, L., & Daniel, H.-D. (2015). Testing for the fairness and predictive validity of research funding decisions: A multilevel multiple imputation for missing data approach using ex-ante and ex-post peer evaluation data from the Austrian Science Fund. *Journal of the Association for Information Science and Technology*, 66(11), 2321–2339. <https://doi.org/10.1002/asi.23315>
- Ochsner, M., & Dokmanović, M. (2017). Quality criteria and research obstacles in the SSH in Macedonia. In *2nd international conference on research evaluation in the social sciences and humanities, RESSH 2017* (pp. 69–71). https://www.uantwerpen.be/images/uantwerpen/container41447/files/A5-BOOK_RESSH2017_170717-interactive.pdf
- Ochsner, M., Galleron, I., & Ionescu, A. (2017). *List of projects on SSH scholars' notions of research quality in participating countries*. ENRESSH Report Version 1.4. https://enressh.eu/wp-content/uploads/2017/09/Report_Quality_Projects.pdf
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2012). Indicators for research quality in the humanities: Opportunities and limitations. *Bibliometrie - Praxis und Forschung*, 1, 4. <https://doi.org/10.5283/bpf.157>
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2013). Four types of research in the humanities: Setting the stage for research quality criteria in the humanities. *Research Evaluation*, 22(2), 79–92. <https://doi.org/10.1093/reseval/rvs039>
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2014). Setting the stage for the assessment of research quality in the humanities. Consolidating the results of four empirical studies. *Zeitschrift für Erziehungswissenschaft*, 17(6), 111–132. <https://doi.org/10.1007/s11618-014-0576-4>
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (Eds.). (2016). *Research assessment in the humanities. Towards criteria and procedures*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-29016-4>
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2017). Assessment Criteria for Early Career Researcher's Proposals in the Humanities. In *Proceedings of the 16th International Conference on Scientometrics and Informetrics ISSI*, 16.-20. October at Wuhan University (pp. 105–111). Wuhan University.
- Ochsner, M., Hug, S. E., & Galleron, I. (2017). The future of research assessment in the humanities: bottom-up assessment procedures. *Palgrave Communications*, 3, 17020. <https://doi.org/10.1057/palcomms.2017.20>
- Polanyi, M. (1967). *The tacit dimension*. Routledge & Kegan Paul.
- Sousa, S. B., & Brennan, J. L. (2014). The UK Research Excellence Framework and the transformation of research production. In C. Musselin and P. Teixeira (Eds.),

Reforming higher education (pp. 65–80). Springer. https://doi.org/10.1007/978-94-007-7028-7_4

Tamblyn, R., Girard, N., Qian, C. J., & Hanley, J. (2018). Assessment of potential bias in research grant peer review in Canada. *Canadian Medical Association Journal*, 190(16), E489–E499. <https://doi.org/10.1503/cmaj.170901>

Thorngate, W., Dawes, R. M., & Foddy, M. (2009). *Judging merit*. Psychology Press.

Tversky, A., & Kahnemann, D. (1974). Judgement under uncertainty: heuristics and biases. *Science*, 185, 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>

VSNU, NWO, KNAW. (2016). *Standard Evaluation Protocol 2015–2021. Protocol for Research Assessments in the Netherlands*. VSNU, NWO and KNAW. <https://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/SEP2015-2021.pdf>

Wissenschaftsrat. (2004). *Recommendations for rankings in the system of higher education and research*. Wissenschaftsrat. https://www.wissenschaftsrat.de/download/archiv/6285-04-engl.pdf?_blob=publicationFile&v=2

Definitions of societal impact and its evaluation in context

By Gemma Derrick

Introduction

Without a doubt, one of the major sources of contention within the modern research reward system, or the valuation of research outputs has been the concept of the societal impact of research. It is a term that launched a thousand articles and research books, with renewed enthusiasm and insights in how research can return on its societal contract intertwined with rhetoric around the increased accountability of science and transparency of public research investment.

Loosely defined around the impact that research has beyond the academic system, the concept of societal¹ impact has goaded research valuation processes as well as evaluation criteria globally. This has mostly been because of renewed interest in public accountability and framing research investment as a means to a societal end such as economic growth, job creation and addressing societal challenges. In addition, as this chapter will argue, that it is also because of the academic community's increased understanding, and begrudged acceptance of societal impact as a component of research excellence that has allowed its implementation in formal research evaluations on all levels globally. Many researchers viewing both societal and academic impact² as two sides of the same coin (D'Este et al., 2018) which is in direct contrast to concerns that the creation, and therefore the evaluation, of excellent societal impact is independent to academic excellence. This article embraces a broad conceptualization of societal impact that manifests itself differently within national research policies that have moved to formalize its assessment.

Hand in hand with a wider acceptance comes a new generation of early career researchers, more amenable and socialized towards the importance of research impact beyond academia in considerations of excellence, and the development of more sophisticated tools with which to measure outcomes, but also monitor impact development. The result has been a more global understanding of research excellence driven by more agnostic motivations of societal relevance and the necessity to learn about the concept of societal impact, and mechanisms of how to operationalize it on an individual level, and evaluate it on an organization and national levels.

Nonetheless, despite this enthusiasm for the greater recognition of societal impact within the academic reward system, for the sake of its assessment, its definition will differ depending on the context of its assessment. Recent interest in mapping and characterizing impacts over fields, countries and public/private partnerships contexts (Benneworth & Jongbloed, 2010; Bornmann, 2012, 2013) has seen an evolving conceptualization of

¹ In this context the term 'societal' is used to refer to impact beyond academia and includes economic, cultural, social and notions of public engagement.

² Academic impact refers, although not exclusively, to traditional outputs of academic activity such as books, articles and PhD creation.

societal impact. One that extends from linear perceptions of bench to bedside³ to a multi-level and multi-actor perception of the use and evolution of research into usable, extra-academic evidence (Bayley & Phipps, 2019; Phipps et al., 2016). Derrick (2018) refers to impact as an “*ambiguous concept*” where its current absence in the process of formal and explicit academic socialization on the individual level results in questions around the process of assessment through peer review since peer review is governed by academic expertise, rather than more generalized social expertise (Derrick, 2018). Likewise, Nutley et al. (2007) concluded that “*research is a somewhat elusive concept, difficult to operationalize, political in essence and hard to assess in a robust and widely accepted manner*” (Nutley et al., 2007, p. 295). Its definition, therefore, is forever in flux and sensitive to: changes in notions of public value; the academic community’s growing appreciation and acceptance of societal impact as a criterion; as well as the experience gained in operationalization as an assessment criterion in practice. However, despite this growing interest, developments in definitions and formalized inclusions in many research assessment procedures, the applicability of societal impact for the social science and humanities (SSH) is still a matter of debate (Wroblewska, 2019).

This chapter will explore conceptualizations of societal impact that are relevant to different modes and foci of evaluation. It will specifically distinguish between ex-post versus ex-ante modes of evaluation of societal impact. It will include a discussion of the foci of many definitions of societal impact used across Europe on national, organizational and national levels. As such, it highlights some of the main challenges facing societal impact assessment, and the current tools utilized for its assessment with a particular focus on peer review.

Ex-post societal impact

This section will explore the existence and utilization of models and definitions of societal impact that are based towards an ex-post understanding of research impact. Ex-post definitions of societal impact are those that have already occurred, or ‘after the fact’. This conceptualization opens the assessment to the possibility of primarily quantitative indicators used to indicate that societal impact has occurred, but does not go so far as to assign a value to the impact achieved.

The main example of the utilization of ex-post impact assessment is the UK’s 2014 (and subsequent 2021) Research Excellence Framework (REF2014/2021). The REF in 2014 represented the world’s first formal, ex-post assessment of how research had had an impact beyond academia linked to the allocation of research funding. The use of this criterion has positioned societal impact as a serious notion of academic excellence (i.e. on par with the importance of traditional notions of academic excellence and esteem) on the individual level, as one that is of strategic importance on the organizational level. As a framework, the UK’s REF2014/2021 represents perhaps the world’s most developed agenda for evaluating the wider benefits of research and its perceived political success has influenced many other countries in how they define, and approach its assessment. The definition used to drive the submissions was very broad and developed, through a series of pilots, to be as

³ A term used to describe the process by which the results of research done in the laboratory are directly used to develop new ways to treat patients. <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/bench-to-bedside>

all-encompassing as possible. The REF2014 definition “impact is defined as an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia” has since evolved to a REF2021 definition that uses the experience of the REF2014 to build upon and expand definitions of impact to include notions of ‘public engagement’ as well as ‘impact on teaching beyond your institution’. The second addition to the definition is particularly pertinent for the SSH fields of Education, and Educational Research. Despite the changes in the definition, the assessment process is still dictated by the sub-criteria ‘significance’ and ‘reach’⁴, and the assessment is still performed on a 5-star scale (from Unclassified to 4-stars).

While evolution of the REF2014 impact definition to a broader, more inclusive understanding of societal impact are welcome, the emphasis within this definition on a ‘change’ and ‘effect’ restrict the types of evidence that can be used in narratives to demonstrate research value beyond academia. As such, this emphasis risks de-valuing the contribution of SSH research to notions of public value, as the assessment of societal impact becomes more goal-orientated, and therefore amenable to more quantitative indicators of value. In fact, from an evaluation of the REF2014 case studies, research has shown that there was an advantage associated with those case studies that promoted a single, long term, quantifiable outcome (Watermeyer & Hedgecoe, 2016). This was found to be the case in Business and Management (Hughes et al., 2019), political science (Dunlop, 2018) and STEM-featured units of assessment such as medical and biomedical science (Greenhalgh & Fahy, 2015) and physics (Watermeyer, 2016). Commonly, these outcomes as “effects” are associated with societal impacts that are more downstream in nature and are represented within an ex-post evaluation through quantitative indicators.

Despite this conflict, the adoption of an ex-post societal impact criterion in the UK has influences many other countries as well. In Norway, for example, the definition of societal impact has been heavily influenced by the UK definition provided above. However, as Wroblewska (2019) highlights, this exercise produced case studies with quite dissimilar characteristics and this was partly explained by the lack of a link between evaluation results and the distribution of funding; as well as a mismatch between the purposes (summative vs formative), processes (goal-orientated vs experimental) and culture (performative vs reflexive) existing within Norwegian universities.

Ex-ante societal impact

Unlike ex-post impact, ex-ante societal impact is understood as an estimate of the likely future impact of research. It primarily involves the employment of a level of foresight and involves a prospective analysis of what the impact of the research under question *might* be in the future. In assessment practice, this may result in proxy indicators including: track record; reach of collaborative networks; as well as nuanced indicators such as trust in the applicant to fulfil the promised work that are influenced by the health of existing personal connections (“*I know this person, and can trust them to do the work*”), used as substitutes for a societal impact evaluation rather than driving an independent assessment of excellence.

⁴ ‘Significance’ refers to the intensity of the influence or effect of the impact (Derrick, 2018)

In this section, we rather explore those definitions of societal impact than imitate an ex-ante conceptualisation and, using the information presented above as a contrast, explore the definitions currently in use in different national contexts. It outlines the current use of ex-ante definitions and the success and limitations of approaches and tools currently used on individual, research project and national levels.

Definitions and mechanics (Derrick, 2018) of ex-ante societal impact are used in various forms by many government-based funding agencies. In the UK, the all Research Councils use a common definition regardless of disciplinary differences. This definition, “*the demonstrable contribution that excellent research makes to society and the economy*”, creates an explicit link between research excellence and the contribution that it may make to society and the economy. Likewise, in relation to the EU research council definition “the economic and/or societal impact expected from the project, including the identification of customer and societal benefits; definition of the process to be followed leading to concrete application; initial steps of analysis of the advantages of the project’s outcomes over existing products, policies, or processes; and, where applicable, brief explanation of the activities to be undertaken in terms of clarification of IPR position and strategy, testing in real world contexts, plans for contacts with commercial and/or societal partners”.

There is an emphasis on the non-academic impacts, but the inclusion criteria have been broadened to use more explicit examples; “the economic and/or societal impact expected from the project, including the identification of customer and societal benefits”. It also uses concrete examples of expectations about how these societal impacts are to be embedded into the research project at the proposal stage; “...clarification of IPR position and strategy, testing in real world contexts, plans for contacts with commercial and/or societal partners”. This level of detail is in stark contrast to the abstraction evident in the NSF (National Science Foundation) definition of the ‘Broader Impacts’ criteria (“the potential to benefit society and contribute to the achievement of specific, desired societal outcomes”). As with the definition provided by the UK’s Research Councils, allows for proposals to specify and imagine the societal impacts that may emerge from their research, and elect the pathways to manage (and monitor) the generation of these impacts. Finally, the Research Council of Norway (Forskningradet) use a criterion in evaluations towards the potential societal impact of the research. In this evaluation, evaluators are urged to assess the extent to which the planned outputs of the proposed project address important present and/or future scientific challenges and, further, the extent to which the planned outputs will address UN Sustainable development goals. In addition, evaluators they must consider the extent that the potential societal impacts are clearly formulated and plausible⁵ which demonstrates where the danger of proxies being applied to evaluate such as track record, as well as non-independent tendencies (Derrick, 2018), to assess ex-ante impact claims.

The assessment and selection of the appropriate tool to assess societal impact will depend largely on its conceptualization and definition. In this way, conceptualizations that rely heavily on assumed linear modes of research-into-practice and the instrumental use of research will be more orientated into the use of indicators that assess if, and how a study is used (and cited) in a policy document or guideline.

⁵ https://www.forskningradet.no/en/Research_funding_in_2019/1254037516684

Challenges for assessing Social Sciences and Humanities impact

The trends and pressure to broaden the definition of academic excellence to include notions of societal impact has been received differently within different academic disciplines. Without a doubt, the challenges of identifying and evaluating societal impact in SSH has been felt acutely. Indeed, some authors have argued that the increasing influence of societal impact in demonstrating the value of research investment pose a threat to SSH (Benneworth 2015; Benneworth et al., 2016). This threat contributes to the perceived ‘crisis of humanities’ which relates to a decreased importance and prestige of the humanities, a decline in funding and/or a fall in recruitment numbers to university SSH courses (Benneworth et al., 2016).

As a broad discipline, SSH brings particular challenges in the assessment of societal impact that go beyond the general concerns related to attribution, causality, and time lag. Impact that stems from SSH research is not as amenable to the measures, definition and models of impact more readily adhered to by STEM subjects (Derrick & Ross-Hellauer, forthcoming; Ross-Hellauer & Derrick, 2019). This makes it hard to measure and clearly state the value added by SSH research, especially if it is in competition with indicator-driven outcomes that are, not necessarily easier, but more commonly associated with STEM subjects. A relatively new concern that arose from a systematic assessment of how peer review panels navigated the assessment of societal impact within the REF2014, may resonate with the challenges faced by SSH in demonstrating impact (Derrick, 2018). The notion of ‘centrality’⁶ or else, an assessment of how important a piece of research was in realizing a final, impact goal (Derrick, 2018), was shown to be an important consideration for panels to consider to mitigate the issues associated with attribution and causality. However, in these evaluations which were based on narratives or Impact Case studies (ICS), there is a risk that the nuanced nature of SSH work and its contribution may be overshadowed by other, more grandiose claims of ‘centrality’.

However, contrary to these issues of STEM and SSH in competition for notions of public value, is the understanding of societal impact relativism in the SSH research assessment. In this situation, the public value that SSH researchers as a community assign to their field may be blind, fall beyond, or be biased against, those societal impacts that are more amenable to indicators, or else linear notions of impact creation. This is especially the case in the situation of a peer review tool being used to assess the societal impact of SSH. Here panel members might implicitly punish claims of impact that adhere to an indicator-based demonstration of value, in favour of impact claims that are more in line with their own understandings and resistance to the implementation of societal impact as formalized criteria. This behaviour embeds notions of value within the SSH discipline by using its researchers as participants in the evaluation process, mainly through peer review. This works against the adoption of societal impact into community-held notions of academic excellence by excluding notions of impact that are measurable that are perceived to go against the more nuanced understandings of SSH contribution to public value. What is needed, therefore, is a re-examination of what societal impact means from SSH and various

⁶ ‘Centrality’ has been observed in ex-post assessments of societal impact but has not yet been associated with ex-ante assessments of Impact.

indicators that are, and are not, relevant for their measure alongside heightening the community's appreciation of their own value to the public (de Jong et al., 2016).

The future of societal impact evaluation: peer review?

As the 'gold standard' in research assessment, many of the assessment exercises described above use peer review in evaluations of societal impact. Although more commonly associated with the evaluation of traditional notions of research impact, the authority and political legitimacy of peer review (Derrick & Samuel, 2016; Derrick, 2018) lie in the assumption that the appropriate level of expertise, disciplinary relevance and fairness of judgement will be employed during the assessment process. Since the trust between applicant and evaluation panel is based on these assumptions, it is therefore a more relevant evaluation tool for traditional, socialized notions of academic excellence. For societal impact, as well as for traditional forms of academic excellence, models and definitions of the criteria are constructed in practice (Derrick, 2018). As such, the newness of the societal impact as a criterion and the inexperience of review panels to assess this ambiguous criterion (Derrick, 2018), results in the drawbacks commonly associated with peer review such as bias, becoming more pronounced in the assessment process. There is also a risk of evaluators embedding the practice of treating more socialized assessments of research and academic excellence as proxies for assessments of societal impact. Alongside the newness of the criterion comes the risk that evaluators defer their personal conceptions related to political, economic or social considerations as more reliable yardsticks for evaluation (Derrick & Samuel, 2018). This deference queries the very nature of a 'peer' in peer review where even the level of expertise and experience of evaluators to the assessment of societal impact are diffuse in nature. All countries that have adopted formalized societal impact criteria, both ex-post and ex-ante, have also adopted peer review as the evaluation tool. In addition, these countries have attempted to compensate for the lack of expertise in impact-generation by appointing non-academic stakeholders to the panel, research has shown that these evaluators face considerable difficulty in influencing the committee culture of panels, and therefore the evaluation outcome (Derrick, 2018). As a result, some researchers (Derrick, 2018) have questioned the relevance of peer review for the assessment of ambiguous criteria such as societal impact, especially during a time when it occupies a liminal space within research excellence (Watermeyer & Chubb, 2019). As societal impact continues to become implemented in research policy as a central, formalized criterion in many countries, so too will evaluator behaviours and formalized evaluation mechanics evolve within peer review panels. Interestingly some countries, such as Australia and Finland, have explicitly separated the evaluation practice and the definitions of academic and societal impact. In other words, in these countries separate panels evaluating academic and societal impact. This model of assessment is worthy of further exploration.

Conclusion

Both the evaluation of ex-ante and ex-post societal impact has grown in importance resulting in a number of countries adopting its broad definition into formalized evaluation procedures. However, the conceptualization of a broader, societal impact from research remains in flux, and is sensitive to changes in notions of public value. The choice, therefore, remains between embracing the serendipitous and unpredictable nature of impact as a

concept to be resolved during evaluative practice; or stabilizing the practice by creating mechanics (definitions, tools and indicators) that see societal impact as a pragmatic, largely predictable outcome from research. Both options have consequences for SSH research. In both its ex-ante and ex-post forms, the assessment of societal impact highlights some specific challenges for SSH in demonstrating and realizing its wider public value. These differences in conceptualizations of SSH-value beyond academia can play out in peer review panels, which, as the evaluation tool of choice, can act to embed or, in the very least, pronounce biases within the evaluation process that can act against the promotion of SSH research. As the adoption of societal impact as a criterion becomes more widespread, this challenge is likely to decrease in importance. Nonetheless, more research focusing on how panels construct models and definitions of impact in practice are needed in order to swiftly address any challenges that act against the interests of the SSH community.

References

- Bayley, J. E., & Phipps, D. (2019). Building the concept of research impact literacy. *Evidence & Policy: A Journal of Research, Debate and Practice*, 15(4), 597–606. <https://doi.org/10.1332/174426417X15034894876108>
- Benneworth, P. (2015). Tracing how arts and humanities research translates, circulates and consolidates in society. How have scholars been reacting to diverse impact and public value agendas? *Arts and Humanities in Higher Education*, 14(1), 45–60. <https://doi.org/10.1177/1474022214533888>
- Benneworth, P., Gulbrandsen, M., & Hazelkorn, E. (2016). *The Impact and Future of Arts and Humanities Research*. Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-40899-0>
- Benneworth, P., & Jongbloed, B. W. (2010). Who matters to universities? A stakeholder perspective on humanities, arts and social sciences valorisation. *Higher Education*, 59(5), 567–588. <https://doi.org/10.1007/s10734-009-9265-2>
- Bornmann, L. (2012). Measuring the societal impact of research: Research is less and less assessed on scientific impact alone – we should aim to quantify the increasingly important contributions of science to society. *EMBO Reports*, 13(8), 673–676. <https://doi.org/10.1038/embor.2012.99>
- 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>
- de Jong, S. P. L., Smit, J., & van Drooge, L. (2016). Scientists' response to societal impact policies: A policy paradox. *Science and Public Policy*, 43(1), 102–114. <https://doi.org/10.1093/scipol/scv023>
- Derrick, G. (2018). Impact from the evaluators' eye. In G. Derrick, *The evaluators' eye* (pp. 1–19). Springer International Publishing. https://doi.org/10.1007/978-3-319-63627-6_1

- Derrick, G. E., & Samuel, G. N. (2016). The evaluation scale: exploring decisions about societal impact in peer review panels. *Minerva*, 54(1), 75–97. <https://doi.org/10.1007/s11024-016-9290-0>
- Derrick, G. E., & Samuel, G. S. (2018). Exploring the degree of delegated authority for the peer review of societal impact. *Science and Public Policy*, 45(5), 673–682. <https://doi.org/10.1093/scipol/scx091>
- Derrick, G. & Ross-Helauer, T. (forthcoming). *Decolonising the social sciences and humanities through peer review*.
- D’Este, P., Ramos-Vielba, I., Woolley, R., & Amara, N. (2018). How do researchers generate scientific and societal impacts? Toward an analytical and operational framework. *Science and Public Policy*, 45(6), 752–763. <https://doi.org/10.1093/scipol/scy023>
- Dunlop, C. A. (2018). The political economy of politics and international studies impact: REF2014 case analysis. *British Politics*, 13(3), 270–294. <https://doi.org/10.1057/s41293-018-0084-x>
- Greenhalgh, T., & Fahy, N. (2015). Research impact in the community-based health sciences: An analysis of 162 case studies from the 2014 UK Research Excellence Framework. *BMC Medicine*, 13(1). <https://doi.org/10.1186/s12916-015-0467-4>
- Hughes, T., Webber, D., & O’Regan, N. (2019). Achieving wider impact in business and management: Analysing the case studies from REF 2014. *Studies in Higher Education*, 44(4), 628–642. <https://doi.org/10.1080/03075079.2017.1393059>
- Nutley, S. M., Walter, I., & Davies, H. T. O. (2007). *Using evidence* (1st ed.). Bristol University Press. <https://doi.org/10.2307/j.ctt9qgwt1>
- Phipps, D., Cummins, J., Pepler, D. J., Craig, W., & Cardinal, S. (2016). The Co-produced Pathway to Impact Describes Knowledge Mobilization Processes. *Journal of Community Engagement and Scholarship*, 9(1), Article 5. <https://digitalcommons.northgeorgia.edu/jces/vol9/iss1/5>
- Ross-Hellauer, A., & Derrick, G. (2019). *Decolonising the social sciences and humanities through peer review*. The Third Research Evaluation in the SSH Conference (RESSH 2019), Valencia.
- Watermeyer, R. (2016). Impact in the REF: Issues and obstacles. *Studies in Higher Education*, 41(2), 199–214. <https://doi.org/10.1080/03075079.2014.915303>
- Watermeyer, R., & Chubb, J. (2019). Evaluating ‘impact’ in the UK’s Research Excellence Framework (REF): Liminality, looseness and new modalities of scholarly distinction. *Studies in Higher Education*, 44(9), 1554–1566. <https://doi.org/10.1080/03075079.2018.1455082>
- Watermeyer, R., & Hedgecoe, A. (2016). Selling ‘impact’: Peer reviewer projections of what is needed and what counts in REF impact case studies. A retrospective analysis. *Journal of Education Policy*, 31(5), 651–665. <https://doi.org/10.1080/02680939.2016.1170885>

Wroblewska, M. N. (2019). *Impact evaluation in Norway and in the UK: A comparative study, based on REF 2014 and Humeval 2015-2017*. ENRESSH.
<https://doi.org/10.3990/4.2666-0776.2019.01>

Does excellence have to be in English? Language diversity and internationalisation in SSH research evaluation

By Nina Kancewicz-Hoffman & Janne Pölönen

Introduction

With the increasing globalisation of research and of academia in general, English (or rather Global English as it is often called to distinguish it from languages spoken in English-speaking countries) has become the language in which research is both performed and communicated. This is related to increasing internationalisation of the institutional environment of research creating pressures but also opportunities for transnational collaborations through funding instruments, scientific journals or professional associations (Heilbron et al., 2017). These new research policies incentivising research globalisation have direct impact on research evaluation processes and criteria and on the role of English in research. This is evidenced for example by changes in the evaluation process within EU Framework Programmes which is now performed only in English although this was not the case 20-30 years ago (Sapiro & Seiler-Juilleret, 2016). At the national level, the internationalisation of research and the use of Global English in research communication are favoured as well. Consequently, evaluation processes at the national level have increasingly international character and are conducted, at least partly, in English enabling the participation of international evaluators to ensure quality but also to introduce national research communities to international collaboration as well as competition.

Both the internationalisation process and the dominant use of English language have been introduced firstly in STEM disciplines and are rooted in their research cultures. The SSH community is characterised by the embedment of research in the local context and by linguistic diversity in producing and disseminating knowledge. Due to those characteristics, “Patterns of internationalization in the SSH therefore differ from internationalization in the natural sciences.” (Heilbron et al., 2017; see also Sivertsen, 2016; Kulczycki et al., 2018). Traditionally, SSH researchers publish in more than one language, although not always one of them is English. A study of multilingualism in scholarly publishing in Poland shows that researchers from all fields publish in more than one language but when STEM researchers publish most often in English and in Polish, SSH researchers, and especially humanities researchers, often publish in more than one (or even two) languages but more often, their dominant language is neither English nor Polish (Korytkowski & Kulczycki, 2019). In a study of multilingualism in SSH scholarly communication, Kulczycki et al. (2020) showed that 53% of 25,365 SSH researchers from seven European countries published peer-reviewed journal articles in a three-year period in more than one language, ranging from 38% in Flanders (Belgium) to 69% in Slovenia.

Moreover, different SSH disciplines, depending on a number of factors, adopt standards of internationalisation and use of English in scholarly communication at a different pace (Sapiro & Seiler-Juilleret, 2016). Although generally both social sciences and humanities disciplines are embedded in the local context, some of the humanities disciplines are more strongly dependent on language as a part of research instrumentarium. In those disciplines, the ability to express ideas in writing and the quality of writing directly influence the quality of research outputs and outcomes (Sapiro & Seiler-Juilleret, 2016). Therefore, forcing all

research communication to be only in English works against its quality and can be seen as “impoverishment” of SSH research (Leão et al., 2018). While use of English as publication language differs between SSH fields, it also differs between countries: in the Western European and Nordic countries, typically a larger share of peer-reviewed output is in English than in the Central and Eastern European countries (Kulczycki et al., 2018). Differences between countries are partly due to political and scholarly traditions, size of the market for local language outputs, as well as incentives created by evaluation and funding systems. In parallel to the growing pressure for internationalisation and the use of English, there are also growing concerns regarding the societal impact and the societal role of SSH. For academic impact in the globalised research environment, the use of English is a requirement. However, communicating knowledge in national/vernacular languages may be more powerful in order to reach wider, non-specialist public (Sapiro & Seiler-Juilleret, 2016; Sivertsen, 2018). Sapiro & Seiler-Juilleret note that the dominance of English in academic publishing may lead to adverse effects when in some countries, younger researchers, in order to adapt to quality requirements and evaluation cultures and expectations, have no experience of writing scholarly work in their national language (Sapiro & Seiler-Juilleret, 2016). Sivertsen concludes that as researchers are expected to communicate their knowledge for different purposes and different publics, they should feel free to choose a language of communication which suits best a given communication situation: “The choice of language depends on the international scholarly relevance of the research versus the societal relevance for the culture and society being studied” (Sivertsen, 2016).

In the SSH, language diversity or multilingualism are here to stay, although most probably with a trend for the growing use of English as a part of the internationalisation process. This poses a number of challenges to evaluation processes and procedures. Those challenges are rarely noted and addressed, as generally – as Sivertsen remarks – “language is invisible” in thinking of research policies of which evaluation is a part. Sivertsen recommends working toward “balanced multilingualism” in research and research policy including evaluation but warns that it will be a long process requiring compromises (Sivertsen, 2018). The ‘Helsinki Initiative on Multilingualism in Scholarly Communication’ makes direct recommendations regarding language diversity in evaluation:

“3. Promote language diversity in research assessment, evaluation, and funding systems.

- Make sure that in the process of expert-based evaluation, high quality research is valued regardless of the publishing language or publication channel.
- Make sure that when metrics-based systems are utilized, journal and book publications in all languages are adequately taken into account.”

(Helsinki Initiative, 2019; Kulczycki et al., 2020)

When comparing institutional evaluation exercises in UK and Portugal, Deem stresses the importance of finding a way to take into consideration research outputs in diverse languages and recommends “care needs to be taken to deal with how work in a variety of languages can be fairly assessed” (Deem, 2016). The EC Expert Group on Assessment of University-Based Research as well as the ESF Peer Review Guide recommend considering all languages equally in the evaluation (European Commission 2010; European Science Foundation [ESF], 2011). However, there are no clear recommendations how to deal with this issue in evaluation practice.

This chapter will highlight those instances where the issue of language of research communication is relevant for the evaluation process in general and for peer review in particular. It will also discuss possible solutions and recommendations.

1. Assessing the quality of research

As writing, formulating a text is in SSH a part of the research process. The command of language by a researcher has a direct influence on the quality of research work. The dominance of scholarly communication in English automatically gives advantage to English native speakers and researchers based in English-speaking countries. If they write and publish in English those researchers have a clear advantage over researchers using English as a second language: “This linguistic inequality creates a bias which has much more negative consequences in the SSH than in the natural sciences because of the greater importance given to written expression and interpretative analysis as well as less firmly standardized (and universally canonized) conceptual formulations” (Sapiro & Seiler-Juilleret, 2016). This is documented by the dominance of US- and UK-based researchers in many contexts. Among others the command of English is considered one of the reasons for the successes of UK-based researchers in the EU Framework Programmes (Heilbron et al., 2017). It is however not clear how this disadvantage could be directly offset in the peer review process. There are two ways to remediate it, that is to consider publications in English and in other languages on equal foot and/or to professionally translate scholarly publications and other forms of scholarly communication.

2. Barriers to research dissemination

The first, necessary condition to consider equally both English and non-English publications in the evaluation process, is to provide equally easy and open access to them. The fact that the majority of highly respected journals in almost any discipline of SSH are published in English, also in Europe, shows to what extent English is a dominant language in scholarly publishing. However, this publishing pattern strongly influences communication and collaboration flows (Heilbron et al., 2017) and substantially limits scientific communication and academic impact of research in languages other than English. In this way, multilingualism in research communication, although considered positive in itself, creates barriers for fellow researchers to dissemination of and access to works published in diverse national languages. For peer review, this means that reviewers and panel members may not be aware of new achievements and publications in languages other than English. If experts refer in their evaluation to metric information provided by commercial indexes predominantly used for evaluation purposes like Web of Science by Clarivate Analytics and Elsevier’s Scopus, their assessment may be even more biased due to low representation of non-English journals.

3. Increasing accessibility of research outputs in languages other than English

There are initiatives facilitating access to publications in national/vernacular languages. One of the oldest among them is the European Reference Index for the Humanities (ERIH), functioning now as ERIH PLUS and covering also social sciences. In fact, the main goal of creating ERIH in 2002 was to provide access to quality research in the Humanities in European languages other than English: “The main aim of ERIH has been from its very

beginnings to enhance *global visibility* of high quality research in the humanities published in academic journals in various European languages all over Europe” (ERIH PLUS, 2014). ERIH continues to play this role with almost 8000 journals registered in the database out of which about 4500 are in 30 languages other than English (ERIH PLUS, 2014).

OPERAS – ‘Open Access in the ERA through scholarly communication’ – is a consortium building a European research infrastructure supporting the dissemination of research publications in diverse languages (OPERAS, 2017). For OPERAS, the multilingualism of research communication in SSH is considered an added value and should not create a barrier to internationalisation “[...] the challenges for OPERAS are to support researchers that want to continue publishing in their own language and to develop transnational scientific cooperation at the same time. Thereof, the proposed intervention areas are: translation, multilanguage discovery tool and the endowment of national languages” (Leão et al., 2018). OPERAS works toward creating services to certify, find and activate research in SSH in a number of languages, however making those services fully operational will take time.

4. Metrics informing experts

In many evaluation contexts experts are informed with metrics, including for example number of peer-reviewed publications and citations, to support the assessment of the quantity and quality of research. Most frequently used information sources, often understood to represent the international excellence, are commercial databases Web of Science by Clarivate Analytics and Elsevier’s Scopus. The problem is that these databases, focusing on a selective subset of mainly international peer-reviewed English language journals, provide a very poor representation of scholarly literature in most SSH fields. They almost completely omit books, national journals and non-scholarly publications (Hicks, 2004). As Sivertsen (2016) points out, ‘coverage in a commercial indexing service should not be used as a criterion for research quality or an indicator of internationalization in the SSH’.

Kulczycki et al. (2020) show that, in particular, Web of Science and Scopus provide an impoverished picture of multilingualism. Web of Science and Scopus covered only 25.0% and 30.7%, respectively, of the 164,218 peer-reviewed journal articles produced in 2013–2015 by SSH researchers from seven European countries (Czech Republic, Denmark, Finland, Flanders [Belgium], Norway, Poland and Slovenia). Moreover, Web of Science and Scopus coverage was only 3.4% and 8.0% of the peer-reviewed articles published in local languages of these countries, and 10.6% and 17.4% of journal articles in other languages not including English.

The solution is to supply expert-evaluators with metrics derived from more comprehensive information sources covering publications in different languages and books, such as the institutional Current Research Information System (CRIS) or a national bibliographic database (van Leeuwen et al., 2016; Sile et al., 2018; Giménez-Toledo et al., 2019). More comprehensive lists of journals and book publishers have also been developed both at international (ERIH PLUS, 2014) and national level to identify peer-reviewed publication channels (Sivertsen, 2016, 2018; Giménez-Toledo et al., 2019). In using such lists in evaluation, however, it is important to remember the following recommendation of the DORA declaration (DORA, 2012):

‘Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist’s contributions, or in hiring, promotion, or funding decisions.’

The role of metrics in evaluation should be to support expert assessment based on research contents, not to ‘substitute for informed judgement’ (Hicks et al., 2015).

5. Translation of scholarly work

Another way of facilitating access to multilingual scholarly work is translation. Sapiro and Seiler-Juilleret remark that translation of scholarly work can also contribute to the quality of research by introducing another layer of reflexivity or a different cultural context (Sapiro & Seiler-Juilleret, 2016).

Obviously, translation has been a part of SSH for centuries. However, considering the present fast pace of knowledge production and communication, especially in English, scholarly translation faces new challenges. Research shows that its impact on knowledge dissemination and exchange is not balanced. “There are many more books translated *from* English than *into* English, whereas for all other languages the reverse holds true. About sixty percent of all book translations worldwide are made from English, while book translation into English represents two to three percent of the national book production in both the US and the UK, which is among the lowest translation rates in the world” (Heilbron et al., 2017). As a result, publications from other regions/languages do not circulate enough. There is a number of issues to be resolved related to translation in the peer review process:

- What is the status of a work published in one language and then translated (into English or into another language) – which is the original, what if versions are different (and often when translating for another cultural context a publication has to be adapted); this poses a number of questions to reviewers;
- How will translations between languages other than English be considered in the evaluation – will international access to the work be ensured; how easy/difficult it will be to find reviewers with necessary knowledge of languages to be able to evaluate;
- Translation of scholarly work requires professionalism, however “translations are not recognized in academic curricula, apart in disciplines such as philosophy or ancient languages, and scholars seldom have translation experience” (Sapiro & Seiler-Juilleret, 2016).
- On the other hand, to ensure quality scholarly translation requires the knowledge of the field and should be done by fellow-scholars. It is also time consuming as in the SSH, most often monographs are translated. The question is whether a translation is considered and evaluated as a part of a researcher-translator professional track record.
- In case of evaluation of manuscripts or grant proposals, special funding is necessary to cover costs of translation, not to mention problems with finding a good professional translator.

OPERAS proposes for the purpose of reviewing a solution based on technology – automated translation for review: “Attention should be paid to the development of such translators, possibly as tool for ‘working translations’ which can facilitate (international) peer

reviews of manuscripts that are not written in English. In particular, this could be very helpful in order to have a referee on global content (using a working translation and not a publishing translation). In the case that the content proves to be valuable, this first referee could be combined with a second one directed not only to content but also to the quality of writing, thus keeping the possibility of publishing in the original language and not necessarily in English” (Leão et al., 2018). This solution combines advantages of less time- and effort-consuming automated translation with the expertise of a peer with knowledge of the language. It needs to be tested whether, especially in disciplines where the language and its precision is a key, this procedure ensures fair and equal assessment.

Both solutions – facilitating access to research outputs in diverse languages and supporting translations of scholarly works – are ways to enable better and fairer evaluation by peers. However, both are at an early stage of development, their progress will require substantial funding and time.

6. Internationalisation criteria and the question of language

Research funders, especially in STEM fields, consider international impact of funded research the most important consequence of provided funding and the main criterion of research quality. Internationalisation understood in a wider sense – as transnational collaboration combined with international dissemination and impact of research outputs – is often a priority in all types of evaluation: institutional evaluation, evaluations for promotion or for funding programmes. However, funders are not always clear what they understand under this term and what they want to achieve. To encourage internationalisation efficiently and to ensure fairness of the expert-based evaluation of the international character of research requires a clear definition of the notion of internationalisation and of its evaluation criteria. For example, Cappacioni notes that *the internationalisation criterion has been the most problematic* in ANVUR institutional evaluation in Italy in 2004-2010 (Capaccioni & Spina, 2018).

The definition of internationalisation and criteria for its evaluation are usually closely related to the language of scholarly communication as international impact is understood as publishing in international journals which in turn is equalled with journals published in English. However, considering publications in English as international by default may be misleading as not all journals published in English are truly international, they may be internal to an English-speaking country like US or UK. On the other hand, in some SSH disciplines journals in other languages (e.g. French, German, Italian, Russian) may be the most relevant international communication channels. Sivertsen recommends developing more nuanced criteria of research quality which will do justice to the role, audiences and type of publication (Sivertsen, 2016). Cappacioni proposes to understand internationalisation “as the ability of disciplinary sectors to have a dialogue with the world of international research.” He makes suggestions for developing ways to measure this criterion or more precisely to establish a ranking of international impact (Capaccioni & Spina, 2018).

Unclear or vague evaluation criteria can be wrongly applied by reviewers and could lead to unexpected or adverse effects as research communities usually quickly adapt to evaluation expectations. For example, prioritising international collaborations and English as the language of knowledge communication can lead to devaluation of publications in national

languages and the reluctance or even inability of researchers to write scholarly work in their national language (Sapiro & Seiler-Juilleret, 2016).

7. Selection of Peer Reviewers

Finding appropriate experts is one of the basic challenges of peer review in general. Multilingual evaluation creates an additional level of difficulty in this respect requiring experts who can evaluate in several languages and in respective cultural contexts. It should be assumed that, as a basic requirement, any multilingual evaluation requires sufficient knowledge of English from peers to participate in the exercise as in a multicultural context proceeding will most probably be in English. Also, some part of publications (or grant proposals in case of transnational/multi-country funding competitions) will be in English. When the evaluation covers research in another language it is advisable that *at least* some experts have *at least* passive knowledge of this language. Ideally the command of language should go hand in hand with the familiarity with the cultural context of research under evaluation (European Science Foundation [ESF], 2015; Deem, 2016).

The paradox of this situation is that when identifying and selecting peer reviewers with the expertise in research in another language and culture one has to take into consideration their possible familiarity with the research community of the country and thus possible conflicts of interest. “It was an explicit recommendation of the FCT and ESF to avoid any bias for or against the focus on Portuguese culture and its specificities, relevant especially in humanities and social sciences. This created an additional challenge for Panel 6 (Humanities) to appoint an expert with knowledge of Lusophone studies who at the same time has no recent connections or collaborations in Portugal. This was solved by appointing specialists in comparative literature and translation studies who ensured knowledge and understanding of national languages and cultures and their study, but this added another layer of difficulty (ESF, 2015; see also Deem, 2016).

Peer review in a multilingual context requires more time and resources than an exercise in one language be it English or another one. There are still not enough studies on practices in this area.

Conclusions

After a period of uncontested and increasing dominance of English in research considered a condition *sine qua non* of internationalisation, multilingualism is now gaining recognition. There is a growing number of initiatives which research and promote multilingualism in the academia, especially in SSH so that pioneering ERIH (started in 2008, now ERIH PLUS, 2014) is now followed by projects like INTERCO-SSH (2013), OPERAS (2017) and the Helsinki Initiative (2019). Each of the projects has a different main focus: when ERIH PLUS (European Reference Index for the Humanities and Social Sciences) started as a journal index, INTERCO-SSH (International Collaboration in the SSH) was a research and policy project, OPERAS (Open Access in the ERA through scholarly communication) aims at creating a research infrastructure with a number of functions and the Helsinki Initiative (Helsinki Initiative on Multilingualism in Scholarly Communication) is a campaign aiming at influencing research institutions and policy makers but also individual researchers. Their common goal is to create better and fairer conditions for the use of a variety of languages in research. They also all note that there is a direct relation between language of research

communication and research collaboration as well as peer review and evaluation in general. However, the awareness of this situation among researchers and research policy makers is still not sufficient. As long as the language issue is not well identified – “invisible” as says Sivertsen (2018) –, the consequences of the dominance of English vs multilingualism cannot be properly analysed and addressed. This is probably why, when there is substantial research regarding diverse aspects of internationalisation including research on the growing role of English, research on practices of peer review in multilingual contexts is scarce. To improve the situation of peer review in multilingual contexts, there is more research needed on:

- Practice of peer review in English-dominated and multilingual environments: what elements of the peer review process are relevant and how they influence the evaluation;
- Dissemination of and access to research outputs in a diversity of languages: how diversity influences peer review outcomes and how peer review in multilingual contexts can be improved;
- Research quality in a multilingual, multicultural context: how to evaluate it;
- Evaluation criteria and corresponding indicators in an international context: are they fair to research in languages other than English;
- Evaluation experts: how to select, prepare and support them.

References

- Capaccioni, A., & Spina, G. (2018). Guidelines for peer review. A survey of international practices. In A. Bonaccorsi (Ed.), *The evaluation of research in social sciences and humanities* (pp. 55–69). Springer International Publishing. https://doi.org/10.1007/978-3-319-68554-0_3
- Deem, R. (2016). Recent research evaluations in the UK and Portugal: Methodologies, processes, controversies and consequences. In C. Sarrico, P. Teixeira, A. Magalhães, A. Veiga, M. J. Rosa, & T. Carvalho (Eds.), *Global challenges, national initiatives, and institutional responses* (pp. 159–186). SensePublishers. https://doi.org/10.1007/978-94-6300-675-0_9
- DORA. (2012). San Francisco Declaration on Research Assessment. <https://sfdora.org/read/>
- ERIH PLUS. (2014). *European Reference Index for the Humanities and Social Sciences*. NSD. <https://dbh.nsd.uib.no/publiseringsskanaler/erihplus/about/index>
- European Science Foundation. (2011). *European peer review guide: Integrating policies and practices into coherent procedures*. http://archives.esf.org/fileadmin/Public_documents/Publications/European_Peer_Review_Guide_01.pdf
- European Science Foundation. (2015). *ESF report on the Foundation for Science and Technology (FCT) evaluation of R&D Units 2013–2014*. European Science Foundation. https://www.fct.pt/apoios/unidades/avaliacoes/2013/docs/ESF_FCT_report_Final.pdf

- European Commission. (2010). *Assessing Europe's university-based research-expert group on assessment of university-based research*. Publications Office of the European Union. https://ec.europa.eu/research/science-society/document_library/pdf_06/assessing-europe-university-based-research_en.pdf
- Giménez-Toledo, E., Mañana-Rodríguez, J., Engels, T. C. E., Guns, R., Kulczycki, E., Ochsner, M. Pölönen, J., Sivertsen, G. & Zuccala, 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>
- Heilbron, J., Boncourt, T., Schögler, R., & Sapiro, G. (2017). *European Social Sciences and Humanities (SSH) in a Global Context. Preliminary findings from the INTERCO-SSH Project*. <http://interco-ssh.eu/wp-content/uploads/2017/02/European-Social-Science-in-a-Global-Contextv2.pdf>
- Helsinki Initiative (2019). *Helsinki Initiative on Multilingualism in Scholarly Communication*. Federation of Finnish Learned Societies, Committee for Public Information, Finnish Association for Scholarly Publishing, Universities Norway & European Network for Research Evaluation in the Social Sciences and the Humanities. <https://www.helsinki-initiative.org/> <https://doi.org/10.6084/m9.figshare.7887059>
- Hicks, D. (2004). The four literatures of social science. H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 473–496). Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-2755-9_22
- Hicks, D., Wouters, P. F., Waltman, L., de Rijcke, S., and Rafols, I. (2015) The Leiden Manifesto for research metrics: Use these 10 principles to guide research evaluation. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>
- INTERCO-SSH. (2013). *International Collaboration in the SSH*. <http://interco-ssh.eu/en/>
- Korytkowski, P., & Kulczycki, E. (2019). *Dominant language of researchers across fields*. The Third Research Evaluation in the SSH Conference (RESSH 2019), Valencia. https://ressh2019.webs.upv.es/wp-content/uploads/2019/10/ressh_2019_paper_21.pdf
- Kulczycki, E., Engels, T. C. E., Pölönen, J., Bruun, K., Dušková, M., Guns, R., Nowotniak, R., Petr, M., Sivertsen, G., Istenič Starčič, A., & Zuccala, A. A. (2018). Publication patterns in the social sciences and humanities: Evidence from eight European countries. *Scientometrics*, 116(1), 463–486. <https://doi.org/10.1007/s11192-018-2711-0>
- Kulczycki, E., Guns, R., Pölönen, J., Engels, T. C. E., Rozkosz, E. A., Zuccala, A. A., Bruun, K., Eskola, O., Starčič, A. I., Petr, M., & Sivertsen, G. (2020). Multilingual publishing in the social sciences and humanities: A seven-country European study. *Journal of the Association for Information Science and Technology*, asi.24336. <https://doi.org/10.1002/asi.24336>
- Leão, D., Angelaki, M., Bertino, A., Dumouchel, S., & Vidal, F. (2018). *Operas multilingualism white paper*. <https://doi.org/10.5281/ZENODO.1324026>

- OPERAS. (2017). Open Access in the ERA through scholarly communication. <https://operas.hypotheses.org/>
- Sapiro, G., Seiler-Juilleret, H. (2016). Disseminating the social sciences and humanities. European Policy Brief, September 2016. INTERCO-SSH. https://ec.europa.eu/research/social-sciences/pdf/policy_briefs/interco-ssh_pb_092016.pdf
- Sīle, L., Pölönen, J., Sivertsen, G., Guns, R., Engels, T. C. E., Arefiev, P., Dušková, M., Faurbæk, L., Holl, A., Kulczycki, E., Macan, B., Nelhans, G., Petr, M., Pisk, M., Soós, S., Stojanovski, J., Stone, A., Šušol, J. & Teitelbaum, R. (2018) Comprehensiveness of national bibliographic databases for social sciences and humanities: findings from a European survey. *Research Evaluation*, 27(4), 310–322. <https://doi.org/10.1093/reseval/rvy016>
- Sivertsen, G. (2016). Patterns of internationalization and criteria for research assessment in the social sciences and humanities. *Scientometrics*, 107(2), 357–368. <https://doi.org/10.1007/s11192-016-1845-1>
- Sivertsen, G. (2018). Balanced multilingualism in research. In *L'évaluation de la recherche scientifique : Enjeux, méthodes et instruments, Actes de la colloque International*, Rabat, 6–7 Decembre 2017 (pp. 88–102). <http://colloque.csefrs.ma/wp-content/uploads/2019/01/Actes-du-colloque-2017-6-11-2018.pdf>
- van Leeuwen, T. N., van Wijk, E., & Wouters, P. F. (2016) Bibliometric analysis of output and impact based on CRIS data: A case study on the registered output of a Dutch university. *Scientometrics*, 106(1), 1–16. <https://doi.org/10.1007/s11192-015-1788-y>

PART III: Guidelines, Procedures and Formal Criteria Versus their Practical Application

Review of guidelines and recommendations for evaluation

By Nina Kancewicz-Hoffman

Introduction

In the research community, peer review continues to be a generally accepted evaluation tool. For SSH, it is even the preferred tool as the use of bibliometrics and quantitative indicators is considered less suitable. But peer review is also criticised for several shortcomings, such as subjectivity, risks of several biases and, regarding SSH, for being based on STEM practices, simply because studies of science focused on STEM disciplines and more research about their practices is available (manuscript peer review focuses on journal articles; notion of internationality is based on English language; language and types of documents that can be submitted in grant peer reviews are often limited).

This chapter reviews guidelines, recommendations and reports on peer review with a special focus on SSH. The aim of this chapter is twofold: to identify conceptual issues related to peer review in SSH and to describe their practical implications. The review is based on a selection of documents (see references with an asterisk in the list of references) developed and circulating internationally as, on the one hand, they express more widely shared opinions and positions than guidelines applied only in one country and, on the other hand, their dissemination and impact is transnational. All but two documents cover peer review regardless of research domains, however, the focus of our analysis is on SSH. The documents refer to the application of peer review in diverse evaluation situations – institutions' past performance and future strategies, evaluation of disciplines, individual careers, research outputs and applications for project funding. Some documents focus specifically on peer review, some cover the wider evaluation process of which peer review is only one part. Only documents published in 2010 or later were considered.

The significance of peer review practice in shaping research domains – its influence on individual careers, research communication, research funding and institutional development – calls for constant efforts toward its improvement. The documents under consideration respond to these expectations aiming to improve the process and to assist research policy makers, funders, publishers and reviewers themselves in performing their task. In this chapter we used as a starting point the guidelines and recommendations which are described in more detail by Capaccioni and Spina (2018). We then expand the analysis using a broader range of documents and structuring our analysis according to several issues raised in them: evaluation of research quality; research modalities; fairness, transparency and objectivity; ethics and integrity; professionalisation of peer review; costs and time.

Evaluation of research quality

The main concern for the majority of guidelines and recommendations is the ability of peer review to perform effectively the evaluation of research quality, the identification of research which is excellent and relevant. *The European peer review guide* includes the task of identifying excellence in research among its seven core principles of peer review

(European Science Foundation [ESF], 2011, p. 13). The understanding of quality and its definition as well as criteria which should be used to evaluate it are considered a major issue in the evaluation process and are addressed in several documents (Académie des Sciences, Leopoldina, Royal Society, 2017; Bonaccorsi, 2018; ESF, 2011; European Science Foundation [ESF], 2015; Wissenschaftsrat, 2010). It is obviously not a simple task to formulate a limited number of criteria covering major aspects of scientific work and, moreover, to phrase them in such a way as to receive relevant and comparable assessments. Most often the criteria include relevance (or significance) measured by citations and sometimes also patents or industrial or economic implementation; originality (or innovation) understood as new experimental, factual or methodological findings; (methodological) rigour or clarity understood as correct application of methods; more recently also societal impact interpreted initially as economic, financial gain. In some cases, an additional criterion is internationality understood as publishing in English and/or participating in international projects is also evaluated.

When peer review is applied to evaluation of individuals e.g. for jobs, promotion or as applicants for grants or in institutional evaluation the use of bibliometric and quantitative indicators by peers as supporting information is normally accepted for STEM disciplines although it is often accompanied by warnings against its possible over- or misuse and finding an appropriate balance between qualitative and quantitative criteria is considered a challenge (Bonaccorsi, 2018; ESF, 2015; Institut de France, 2011; Institut de France, 2014; Wissenschaftsrat, 2010).

This approach to peer review poses a problem to the SSH community which is especially cautious in the use of quantitative indicators and has more nuanced understanding of the notion of quality. In the view of the specificity of SSH perspective, the Volkswagen Foundation issued a document *What is Intellectual Quality in the Humanities? Some Guidelines* (Volkswagen Foundation, 2014). The guidelines define seven criteria which are recommended to help evaluate quality in humanities research: scholarly solidity; intellectual significance; critical stance; perspectival suppleness; originality; personal voice; and relevance. Although the set includes the same general terms ‘significance’, ‘originality’ and ‘relevance’ as other guidelines, the context of four additional terms and accompanying explanations gives them an SSH-specific meaning. A good example is ‘intellectual significance’:

“Intellectual significance: Is the problem significant and can the researcher explain why, preferably in terms that make sense also to scholars outside that specialty? Novelty is neither necessary nor sufficient to prove significance: some significant research returns to problems that have preoccupied scholars in a discipline since its inception; novelty for its own sake degenerates into eccentricity. There are many possible dimensions of significance, but almost all of them point beyond the problem at hand: a truly significant problem promises insights that others can build on, the more broadly the better. In some fields, the others in question may be the general public as well as other scholars. Because of both their subject matter and their interpretative standpoint, the humanities can and do change minds about meaning and values (e.g. in works of history, theater [sic] productions, or museum exhibitions).” (Volkswagen Foundation, 2014, p. 2).

This elaborated description gives the criterion of significance a very different meaning than that usually adopted in STEM.

Research modalities

Substantial attention is given in guidelines and recommendations to challenges posed by a group of criteria which have been recently more and more in use: interdisciplinarity, internationalisation and societal impact. It is felt that they are not intrinsic to research but refer rather to modalities of performing research (Ochsner et al., 2013).

It is pointed out in several cases that these criteria are not clearly defined for the purpose of a given evaluation and it is not always stated how they should be applied. In some evaluation protocols, information is lacking whether a given criterion should be directly applied, whether, for example, interdisciplinary research should be valued higher than research equally good in other respects but lacking interdisciplinary character. It is stressed that evaluation organisers should be clear what they want to achieve with these criteria and apply an appropriate procedure (ESF, 2011; ESF, 2015; European Research Council [ERC], 2019; Hornung et al., 2016; House of Commons Science and Technology Committee, 2011).

For example, in case of interdisciplinary research, the situation is clear whenever separate panels with separate budgets are established for proposals claiming interdisciplinarity, but the task of the panel may be less well defined if interdisciplinarity is one of the additional criteria, but no specific score is attributed to it. The case of ERC is interesting as interdisciplinarity is a part of its definition of excellence and is encouraged from the beginning. However, the peer review procedure of interdisciplinary proposals changed after 2011 from having separate panels and funding for interdisciplinary projects to evaluation by regular panels (Science Europe, 2019, p. 12); this is called “mainstreaming of interdisciplinarity” (ERC, 2019, p. 6).

The application of the criterion of internationalisation shows specificities of SSH research. In STEM disciplines, it is widely accepted that publishing outputs in international channels, that is in English, is a sign of quality of research. This is, however, not that evident for research in social sciences and humanities where understanding the internationality criterion as equal with “publishing in English” can lead to unintended consequences especially when combined with other factors. In case of the evaluation of Portuguese research units (ESF, 2015), the combination of exclusively international panel membership and a significant part of research outputs published in Portuguese language led to certain tensions in social sciences and humanities panels (Deem, 2016; ESF, 2015, pp. 25–26).

In case of these criteria, the question of selecting appropriate peers who will have individually or as a panel all necessary expertise to be able to judge them is raised (ESF, 2011; ESF, 2015; Science Europe, 2019).

The criterion of societal impact is still quite new and its position and role in the evaluation varies from country to country and from evaluation to evaluation. In the recommendations for a comparative evaluation of the humanities disciplines, the German Wissenschaftsrat indicated that this criterion, which is one of three main evaluation criteria, is difficult to operationalise and should be applied with caution (Wissenschaftsrat, 2010). It is not at the centre of consideration in any other of the documents.

Fairness, transparency and objectivity

Taking into consideration the role and significance of peer review in the evaluation process, the issues of fairness, transparency and objectivity of peer review and related ethical and

integrity concerns have become crucial. They are addressed in all documents, albeit the focus is different.

There is a number of recommendations regarding procedural and technical solutions. It is suggested that standardising procedures and forms (documentation) improves the quality of peer assessments (ESF, 2011, p.36). Also, ex-ante and/or ex-post publication of procedures, rules, criteria and documentation contributes to transparency (ERC, 2019; ESF, 2015; Hornung et al., 2016; Research Information Network [RIN], 2015; Wissenschaftsrat, 2010). Sometimes, monitoring and control of peer review processes is recommended; this may include monitoring of panel work by independent observers (European Research Council [ERC], 2017, p. 12) or, more often, strict guidelines for and the assessment of written reviews (Bonaccorsi, 2018, p. 89; ERC, 2017, p. 11; ERC, 2019, p. 7; ESF, 2011, p. 36; ESF, 2015, p. 10; RIN, 2015, p. 3)

An interesting challenge to achieving fairness specific for SSH is what Bonaccorsi calls “The Issue of Epistemic Pluralism” or in other words concurrent existence of different schools of thought which may directly influence peers’ opinions. He describes a number of practical solutions proposed by different disciplinary panels. He concludes that the approach taken by ANVUR that is rotation of panels and panel members and publication of the names of referees’ ex-post has been considered by the research community as appropriate (Bonaccorsi, 2018, p. 89–95; see also for the German Forschungsrating: Hornung et al., 2016).

Another way to make the process more open and to enable interaction between applicants and evaluators is creating a possibility of response to review or so-called rebuttal and providing systematic feedback to applicants in form of evaluation reports (ERC, 2019; ESF, 2011; ESF, 2015).

In case of reviewing of journal articles, introducing more openness in the process, e.g. open review procedures supported by new technologies, is recommended to ensure transparency and fairness (Research Information Network [RIN], 2010, p. 13; House of Commons Science and Technology Committee, 2011, p. 67).

Ethics and integrity

A lot of attention is recently paid to issues of ethics and integrity, which are perceived as key in ensuring trust in peer review and also in its quality and reliability. Thus, procedures to identify and eliminate conflicts of interest as well as ensuring the integrity of peers are central to analysed guidelines and recommendations (Académie des Sciences, Leopoldina, Royal Society, 2017; ESF, 2011; ESF, 2015; Institut de France, 2011). In addition to the general framework and a number of procedural recommendations, major funding organisations develop detailed instructions regarding ethical behaviour and the identification of conflicts of interest and require signing of quite detailed Codes of Conduct and Conflict of Interest declarations (ESF, 2011, p. 13–15; ESF, 2015, p. 25; ERC, 2017, pp. 31–35; ERC, 2019, pp. 15–19).

There are also organisational ways helping to avoid unfairness or bias e.g. already mentioned periodical rotation of evaluators (Académie des Sciences, Leopoldina, Royal Society, 2017, p. 3; Bonaccorsi, 2018, p. 89–95; ESF, 2011, p. 25).

Professionalisation of review process

One way to improve peer review and to address the challenges is increasing professionalism of reviewers and of the whole process. Together with the introduction of standardised procedures and forms, the professionalisation should help to achieve objective and comparable reviews. Practically all funders relying on review panels, but also publishers and other institutions performing evaluations, are therefore recommending providing systematic briefings, guidelines, training sessions, workshops and webinars for (future) evaluators (ERC, 2017, p. 15; ESF, 2011, p. 19, p. 24; ESF, 2015, p. 28; House of Commons Science and Technology Committee, 2011, p. 118; RIN, 2010, p. 9). Also technologically supported processes (on-line forms etc.) standardise and professionalise the peer review work (ESF, 2011; ESF, 2015, p. 28)

Costs and time

Peer review has been often criticised for being time-consuming both for administrative staff and for evaluators, who are usually also researchers, and thus generating high costs and delays in research performance and this also an important consideration in the guidelines (Académie des Sciences, Leopoldina, Royal Society, 2017; ESF, 2011; Hornung et al., 2016; House of Commons Science and Technology Committee, 2011; Institut de France, 2011; RIN, 2010, 2015).

A related challenge for peer review organisers is the recruitment of appropriate reviewers and the phenomenon of reviewer fatigue. These issues have been addressed in the analysed documentation by specific practical recommendations, progressively implemented. The main recommendations aiming at improving efficiency are:

- To limit evaluation by peers only to necessary instances and to adapt its complexity to the purpose and object of evaluation (Académie des Sciences, Leopoldina, Royal Society, 2017; ESF, 2011, p. 13; RIN, 2010, p. 12; Wissenschaftsrat, 2010, p. 24; or to introduce innovations simplifying the process (House of Commons Science and Technology Committee, 2011, p. 88).
- To limit the length and contents of applications for funding, jobs but also in institutional evaluation e.g. to limit the number of pages, to limit the number of publications per researchers or institution (Académie des Sciences, Leopoldina, Royal Society, 2017; ESF, 2015, p. 29; Wissenschaftsrat, 2010, p. 26).
- To provide more administrative support for reviewers simplifying their tasks and saving their time mainly through different forms of guidance, briefing and training but also through support staff and technological solutions. ERC, 2017; ERC, 2019, p. 10; ESF, 2011, p. 27; ESF, 2015, p. 12; House of Commons Science and Technology Committee, 2011, p. 118; RIN, 2015, p. 3, 23).
- To widen the pool of researchers from which reviewers are recruited searching for new untapped groups like “non-European experts” and “experts from emerging regions” (ESF, 2011, p. 24), “those from countries which are not traditional scientific leaders” and early-career researchers (House of Commons Science and Technology Committee, 2011, p. 44). In the Research Information Network report *Scholarly communication and peer review*, the authors note growing interest among editors and publishers, however mainly from STEM disciplines, in platforms promoting

and supporting peer review but also facilitating search for new referees like Publons (RIN, 2015, p. 24).

The issue of reward and remuneration of peer review work in view of pressures on researchers has been widely discussed and pros and cons are considered (ESF, 2011; RIN, 2015). Proposals range from scholarly credit and recognition to remuneration, depending on the time and effort involved. *The European peer review guide* points out to differences between funding agencies in this area: “Some organisations pay their reviewers (both external and panel) to conduct assessments while others do not” (ESF, 2011, p. 27). The Research Information Network report *Scholarly communication and peer review* describes practices – other than remuneration – of incentivising researchers to undertake journal reviewing, mainly focusing on recognition not only of quantity but also quality of their work (RIN, 2015, p. 23)

Conclusions

In most cases, the analysed documentation refers to evaluation covering all research domains. It shows that not only concepts of evaluation and peer review in research, but also new ideas and developments, come mainly from STEM disciplines whereas the SSH disciplines have been involved in shaping them only with a delay. Consequently, many aspects of peer review still need adaptation to specific characteristics and requirements of research in the SSH. Therefore, the SSH community needs to develop its own research on and practice of peer review.

Studies included in this report address a number of issues related to the practical application of peer review in SSH disciplines. They show that there is a need for a detailed and systematic research on how peer review is understood and implemented in the SSH context, to identify weak points, distortions of the process and necessary improvements to create protocols adapted to SSH specificities. To achieve this, systematic observation of peer review practices in different countries and different evaluation contexts is a necessary next step.

References

References with an asterisk (*) indicate recommendations and guidelines analysed in this chapter

*Académie des Sciences, Leopoldina, & Royal Society. (2017). *Statement by three national academies (Académie des Sciences, Leopoldina and Royal Society) on good practice in the evaluation of researchers and research programmes*. <https://www.academie-sciences.fr/pdf/rapport/avis111217.pdf>

*Bonaccorsi, A. (2018). Peer review in social sciences and humanities. Addressing the interpretation of quality criteria. In A. Bonaccorsi (Ed.), *The evaluation of research in social sciences and humanities* (pp. 71–101). Springer International Publishing. https://doi.org/10.1007/978-3-319-68554-0_4

Capaccioni, A., & Spina, G. (2018). Guidelines for peer review. A survey of international practices. In A. Bonaccorsi (Ed.), *The evaluation of research in social sciences and humanities* (pp. 55–69). Springer International Publishing, https://doi.org/10.1007/978-3-319-68554-0_3

- Deem, R. (2016). Recent research evaluations in the UK and Portugal: Methodologies, processes, controversies and consequences. In C. Sarrico, P. Teixeira, A. Magalhães, A. Veiga, M. J. Rosa, & T. Carvalho (Eds.), *Global challenges, national initiatives, and institutional responses* (pp. 159–186). SensePublishers. https://doi.org/10.1007/978-94-6300-675-0_9
- *European Research Council (ERC). (2017). *ERC rules for submission & evaluation. Commission Decision C(2017)4750. Version 3.0, 14 July 2017.* https://ec.europa.eu/research/participants/data/ref/h2020/sgl/erc/h2020-erc-se-rules-amended2_en.pdf
- *European Research Council (ERC). (2019). *Frontier Research Grants. Guide for peer reviewers applicable to the ERC Starting, Consolidator & Advanced Grants (ERC Work Programme 2019).* https://ec.europa.eu/research/participants/data/ref/h2020/other/experts_manual/h2020_peer-review_erc-stg-cog-adg-2019_en.pdf
- *European Science Foundation. (2011). *European peer review guide: Integrating policies and practices into coherent procedures.* http://archives.esf.org/fileadmin/Public_documents/Publications/European_Peer_Review_Guide_01.pdf
- *European Science Foundation. (2015). *ESF report on the Foundation for Science and Technology (FCT) evaluation of R&D Units 2013–2014.* European Science Foundation. https://www.fct.pt/apoios/unidades/avaliacoes/2013/docs/ESF_FCT_report_Final.pdf
- *Hornung, A., Khlavna, V., & Korte, B. (2016). Research rating Anglistik/Amerikanistik of the German Council of Science and Humanities. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research assessment in the humanities. Towards criteria and procedures* (pp. 219–233). Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_17
- *House of Commons Science and Technology Committee. (2011). *Peer review in scientific publications. HC 856.* House of Commons. <https://publications.parliament.uk/pa/cm201012/cmselect/cmsctech/856/856.pdf>
- *Institut de France, Académie des Sciences. (2011). *Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs.* <https://www.academie-sciences.fr/pdf/rapport/avis170111.pdf>
- *Institut de France, Académie des Sciences. (2014). *Les nouveaux enjeux de l'édition scientifique.* http://www.academie-sciences.fr/pdf/rapport/rads_241014.pdf
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2013). Four types of research in the humanities: Setting the stage for research quality criteria in the humanities. *Research Evaluation*, 22(2), 79–92. <https://doi.org/10.1093/reseval/rvs039>
- *Research Information Network (RIN). (2010). *Peer review: A guide for researchers.* www.rin.ac.uk/peer-review-guide
- *Research Information Network (RIN). (2015). *Scholarly communication and peer review: The current landscape and future trends. A report commissioned by the Wellcome*

Trust. <https://wellcome.ac.uk/sites/default/files/scholarly-communication-and-peer-review-mar15.pdf>

*Science Europe. (June, 2019). *Symposium report: Interdisciplinarity.* <http://scieur.org/interdisc-symp>

*The Volkswagen Foundation. (2014). *What is intellectual quality in the humanities? Some guidelines.* <https://www.volkswagenstiftung.de/en/news-press/publications/what-is-intellectual-quality-in-the-humanities-some-guidelines>

*Wissenschaftsrat. (2010). *Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften* [Recommendations for comparative research assessment in the humanities]. Wissenschaftsrat. <https://www.wissenschaftsrat.de/download/archiv/10039-10.html> (for an English summary see Hornung et al., 2016).

Ambiguity in identification of scholarly peer-reviewed publications

By Janne Pölönen, Tim C. E. Engels & Raf Guns

It is almost impossible to imagine a research evaluation or funding procedure that would not take into consideration publications, in which researchers seek to demonstrate new findings and applications of their research to other experts in the field. Pre-publication peer review originates from the sciences, where it has been established as a precondition of contributions to scientific knowledge (Baldwin, 2018), and is common also in the social sciences and humanities (SSH). As Sivertsen and Larsen (2012) point out, “it has become generally accepted in the SSH during the last decades that publications presenting new results from research should be peer reviewed”. Nowadays, the distinction between peer-reviewed scholarly publications and those intended for disseminating knowledge beyond academia plays a role in most expert and metrics-based evaluation and funding systems. In this chapter, we present and discuss problems related to identification of peer-reviewed outputs, and the possible implications for research evaluation and funding systems. The current research literature and examples mostly concern university funding and evaluation systems in the Western and Northern European countries, with a specific focus on Belgium and Finland. However, the issues are by no means limited to just these geographic or evaluation contexts.

When peer review is employed as baseline criterion for research outputs to be considered, for example, in a performance-based research funding system (PRFS) or research evaluation process, it is important to recognize that peer review practices differ across fields, and across journal, conference and book publishing (British Academy, 2007; Verleysen & Engels, 2013). Differences may concern the number of referees (one or more), their degree of anonymity vis-à-vis the authors (double-blind, single-blind or open identity), and their relation to the publication channel (editors, editorial board, reading committee, or external). Researchers used to journal peer review in their field may find it difficult to recognize peer review as practiced in journals of another field, or in book publications. While perhaps most researchers identify peer-reviewed publications as those that have actually undergone a certain type of recognizable review process before publication, some may still consider that any substantive contribution to knowledge merits to count as “peer-reviewed” output, whether or not the procedure behind the publication technically counts as peer-review. In all, a certain degree of ambiguity is present when deciding whether a publication channel applies peer review or whether a specific article, chapter or book has undergone pre-publication peer review (Csiszar, 2017; Dahler-Larsen, 2019).

The starting point of the identification of peer-reviewed publications usually is whether the publication channel (e.g. a journal, or a book series or a publisher) has a distinct procedure in place for applying pre-publication peer review. Several European PRFSs rely on the indexation of journals in Web of Science or Scopus as evidence of peer review (Zacharewicz et al., 2018). Other PRFSs, however, also include outputs from publication channels that are not indexed in the major international citation databases, in order to take into account book publications and journal output in a variety of languages. This is the case, for example, in Denmark, Finland, Flanders (Belgium) and Norway, where panels of experts in the field determine the peer review status of journals and book publishers (Sivertsen, 2017, 2018; Aagaard, 2018; Pölönen, 2018; Engels & Guns, 2018). Several studies point

out that even experts in the field may disagree whether a given journal (Nederhof & Zwaan, 1991; Burnhill & Tubby-Hille, 2003; Verleysen & Engels, 2015) or book publisher (Verleysen et al., 2014; Mañana-Rodríguez & Pölönen, 2018) applies peer review and is scholarly or not. Our analysis shows that 9.5 % of the 4505 SSH journals/series included in the national authority lists supporting PRFSs in Finland and Flanders have been evaluated differently by experts as being peer-reviewed or not (Pölönen et al., 2017; Pölönen et al., 2020).

Another source of ambiguity in identifying peer-reviewed outputs is that many journal issues and edited volumes that apply peer review also include items that are not peer-reviewed. Editorials, opinions, comments, discussions, book reviews, and abstracts are typical examples. Moreover, book publishers of peer-reviewed monographs and edited volumes often also publish textbooks, *libri amicorum*, and other types of books that mostly do not undergo peer review. If a PRFS or an evaluation procedure is based on comparison of peer-reviewed outputs, a mechanism ought to be in place for separating them from non-peer-reviewed articles and books. It is an open question, and one dependent on the context and aim of evaluation, how to take non-peer-reviewed outputs into consideration. Most PRFSs, for example, exclude non-peer-reviewed outputs. In the Finnish PRFS they are taken into account with lesser weight than peer-reviewed outputs in channels approved by expert-panels. In evaluation procedures, however, non-peer-reviewed outputs are often taken into account because (among other reasons) they illustrate activity beyond that which is relevant for peers only.

Instead of authority lists of peer-reviewed journals and book publishers, PRFSs and evaluation systems may also rely on researchers' self-reports to determine the peer review status of outputs. Given the ambiguity present in identification of peer-reviewed publication channels, the presence of also non peer-reviewed outputs in such channels, as well as differences in pre-publication peer review practices across fields and publications types, it is not surprising that identification of peer-reviewed outputs based on self-reports also suffer from a certain degree of ambiguity. Research has indicated this in some evaluation contexts, such as the Dutch research evaluation system (Kaltenbrunner & de Rijcke, 2016), and the Swedish universities internal evaluation models (Hammarfelt et al., 2016). Our analysis of 3,596 SSH outputs published in 2011–2015 with authors from more than one Finnish university shows that in 8% of the cases, co-authors of the same article or monograph differed in their assessment of whether it is peer-reviewed or not. The Finnish PRFS relies on both authority list of publication channels and self-reports to determine the value of outputs to universities in terms of annual core-funding, so it has been possible for us also to compare these two methods. Our analysis shows that 16% of 32,427 self-reported peer-reviewed SSH outputs were published in channels that have not been approved to be peer-reviewed by the experts. Overall, the grey zone of peer review appears to be larger in the humanities than the social sciences, and more common among book publications than journal articles and in the national than other language publications. (Pölönen et al. 2017; Pölönen et al., 2020).

Ambiguity in identification of peer-reviewed publications has implications for the PRFS and research evaluation criteria, the publication information systems supporting them, as well as individual researchers. The distinction between peer-reviewed and non-peer-reviewed outputs, and those addressing scholarly and non-scholarly audiences, is not always

clear-cut, especially in the SSH. PRFSs typically define peer review technically, focusing on the existence of a recognizable pre-publication procedure. This definition, based on self-reporting or authority lists, may not encompass all outputs valued by the researchers themselves as original knowledge contributions. From the PRFS perspective, this may not be a problem, as institutional core funding is based on the entire output of universities. In a research evaluation procedure, however, – especially at individual level – it can be of great consequence if a valued research output is not recognized because of the technical PRFS criteria. Also, the information systems supporting the PRFS with publication data often employ the PRFS criteria for peer-reviewed outputs. If these information systems are needed to support wider evaluation and communication purposes, they should be sufficiently inclusive, flexible and structured to include all outputs that researchers consider relevant contributions to research and dissemination, even if they may not be taken into account in the PRFS. The ambiguity in identifying peer-reviewed publications concerns also the self-reported lists of publications, by which individual researchers typically present their research output to various evaluations. This also has research integrity implications, as “misrepresenting research achievements” is one of the unacceptable practices indicated in The European Code of Conduct for Research Integrity (All European Academies [ALLEA], 2017).

References

- Aagaard, K. (2018). Performance-based research funding in Denmark: The adoption and translation of the Norwegian model. *Journal of Data and Information Science*, 3(4), 20–30. <https://doi.org/10.2478/jdis-2018-0018>
- All European Academies (ALLEA). (2017). *The European code of conduct for research integrity*. All European Academies.
- Baldwin, M. (2018). Scientific autonomy, public accountability, and the rise of “peer review” in the cold war United States. *Isis*, 109(3), 538–558. <https://doi.org/10.1086/700070>
- Burnhill, P.M., & Tubby-Hille, M. E. (2003). On Measuring the Relation between Social Science Research Activity and Research Publication. *Research Evaluation*, 4(3), 130–152. <https://doi.org/10.1093/rev/4.3.130>
- British Academy. (2007). *Peer Review: the challenges for the humanities and social sciences. A British Academy Report*. <https://www.thebritishacademy.ac.uk/publications/peer-review-challenges-humanities-and-social-sciences>
- Csiszar, A. (2017). How lives became lists and scientific papers became data: cataloguing authorship during the nineteenth century. *British Journal of History of Science*, 50(1), 23–60. <https://doi.org/10.1017/S0007087417000012>
- Dahler-Larsen, P. (2019). *Quality: From Plato to performance*. Palgrave Macmillan. <https://doi.org/10.1007/978-3-030-10392-7>
- Engels, T. C. E., & Guns, R. (2018). The Flemish performance-based research funding system: A unique variant of the Norwegian model. *Journal of Data and Information Science*, 3(4): 45–60. <https://doi.org/10.2478/jdis-2018-0020>

- Hammarfelt, B., Nelhans, G., Eklund, P., & Åström, F. (2016). The heterogeneous landscape of bibliometric indicators. Evaluating models for allocating resources at Swedish universities. *Research Evaluation*, 25(3), 292–305. <https://doi.org/10.1093/rev/rvv040>
- Kaltenbrunner, W., & de Rijcke, S. (2016). Quantifying ‘output’ for evaluation: Administrative knowledge politics and changing epistemic cultures in Dutch law faculties. *Science and Public Policy*, 44(2), 1–10. <https://doi.org/10.1093/scipol/scw064>
- Mañana-Rodríguez, J., & Pölönen, J. (2018). Scholarly book publishers’ ratings and lists in Finland and Spain: Comparison and assessment of the evaluative potential of merged lists. *Aslib Journal of Information Management*, 70(6), 643–659. <https://doi.org/10.1108/AJIM-05-2018-0111>
- Nederhof, A. J., & Zwaan, R. A. (1991). Quality judgements of journals as indicators of research performance in the humanities and the social and behavioral sciences. *Journal of the American Society for Information Science*, 42(5), 332–340. [https://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5%3C332::AID-ASI%3E3.0.CO;2-8](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5%3C332::AID-ASI%3E3.0.CO;2-8)
- Pölönen, J. (2018). Applications of, and experiences with, the Norwegian model in Finland. *Journal of Data and Information Science*, 3(4), 31–44. <https://doi.org/10.2478/jdis-2018-0019>
- Pölönen, J., Engels, T.C.E., & Guns, R. (2020). Ambiguity in identification of peer-reviewed publications in the Finish and Flemish performance-based research funding systems. *Science and Public Policy*, 47(1), 1–15. <https://doi.org/10.1093/scipol/scz041>
- Pölönen, J., Engels, T.C.E., Guns, R., & Verleysen, F.T. (2017). Is my publication peer reviewed? A comparison of top-down and bottom-up identification of peer review in the framework of the Finnish and Flemish performance-based research funding systems. In *Conference abstracts: Science, Technology and Innovation indicators STI 2017. Open indicators: innovation, participation and actor-based STI indicators Paris 2017, 6–8 September 2017*.
- Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications*, 3, 17078. <https://doi.org/10.1057/palcomms.2017.78>
- Sivertsen, G. (2018). The Norwegian Model in Norway. *Journal of Data and Information Science*, 3(4), 3–19. <https://doi.org/10.2478/jdis-2018-0017>
- Sivertsen, G., & Larsen, B. (2012). Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: an empirical analysis of the potential. *Scientometrics*, 91(2), 567–575. <https://doi.org/10.1007/s11192-011-0615-3>
- Verleysen, F. T., & Engels, T. C. E. (2013). A label for peer-reviewed books. *Journal of the American Society for Information Science and Technology*, 64(2), 428–430. <https://doi.org/10.1002/asi.22836>

- Verleysen, F. T., & Engels, T. C. E. (2015). *ERIH Plus in 2014: Stand van zaken en tijdschriftselectie vergeleken met het VABB-SHW*. ECOOM.
- Verleysen, F. T., Ghesquière, P., & Engels, T. C. E. (2014). The objectives, design and selection process of the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW). In W. Blockmans, L. Engwall & D. Weaire (Eds.), *Bibliometrics: use and abuse in the review of research performance* (pp. 115–125). Portland Press.
- Zacharewicz, T., Lepori, B., Reale, E., & Jonkers, K. (2018). Performance-based research funding in EU Member States – a comparative assessment. *Science and Public Policy*, 46(1), 105–115. <https://doi.org/10.1093/scipol/scy041>

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).

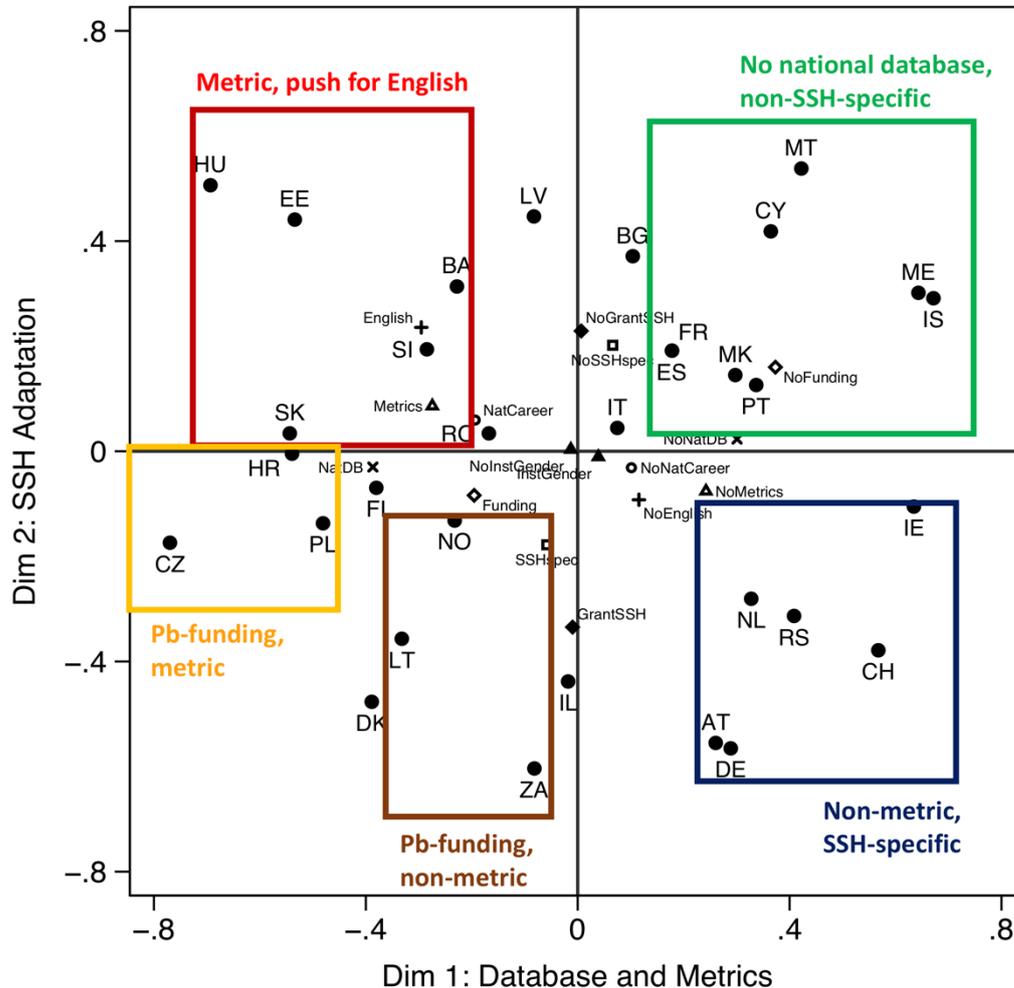


Figure 1. Map of National Research Evaluation Systems from Multiple Correspondence Analysis

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.

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

- 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., Soès, 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>

Practices of peer review in the SSH I: A systematic review of peer review criteria

By Sven E. Hug, Marek Hołowiecki, Lai Ma, Mirjam Aeschbach & Michael Ochsner

Introduction

Criteria are an essential component of any procedure for judging merit. This is widely acknowledged in the literature on peer review. Yet pertinent literature reviews and compendia do not mention or only briefly discuss peer review criteria. To address this research gap, a systematic review of studies on peer review criteria has been conducted. The review focused on the most fundamental question in any evaluation: what criteria are employed in the evaluation? The systematic review was restricted to the two most common forms of peer review: the assessment of manuscripts submitted to journals and the assessment of grant applications. The objectives of the review were (a) to identify studies that develop or derive criteria inductively,⁸ (b) to determine how many of these studies focus on the social sciences and humanities, and (c) to provide a taxonomy of criteria. In the following, preliminary findings on objectives (a) and (b) will be reported. Methodological details and final results will be published in scholarly journals (Hug & Aeschbach, 2019; Hug et al., in preparation).

Preliminary findings

Twelve studies on grant review criteria and twice as many on manuscript review criteria were identified (see Table 1). While the first inductive study on manuscript criteria (i.e. Bonjean & Hillum, 1978) dates back to the time when modern peer review emerged (see Baldwin, 2017, 2018; Moxham & Fyfe, 2018), the first study on funding criteria was only carried out in the 1990s (i.e. Hartmann, 1990). Most studies have examined criteria in the medical and health sciences and the social sciences. Studies on other fields are scarce and there are no studies on manuscript criteria in the natural sciences and in engineering and technology. A possible explanation for the latter could be the fact that all studies on manuscript criteria were done by “insiders” (i.e. researchers examined the criteria employed in a journal of their own field). Since qualitative-inductive approaches are not in the (standard) repertoire of researchers in the natural sciences and in engineering and technology, it is unlikely that criteria are inductively studied in these fields. The systematic review showed that manuscript criteria are mainly examined with data from actual reviews and comments. In contrast, data collection methods such as interviews, surveys and the Delphi method are as important as actual reviews in studies on grant criteria.

⁸ While an inductive approach generates criteria from empirical data, a deductive approach employs theoretically determined or otherwise predefined criteria. The very first studies on peer review criteria employed a deductive approach (e.g. Chase, 1970; Frantz, 1968). The systematic review, however, did not focus on theoretically derived or otherwise predefined criteria but on inductively and empirically established criteria, which were, for example, based on quality conceptions of scholars or on actual comments of reviewers.

Table 1. Characteristics of studies that develop or derive peer review criteria inductively (preliminary data).

	Studies on manuscript review criteria	Studies on grant review criteria
Total number of studies included in the review	24	12
Publication year of studies		
First study	1978	1990
Latest study	2018	2018
Median	2004	2007
Number of studies analysing criteria in the		
Natural sciences	–	2
Engineering and technology	–	2
Medical and health sciences	8	8
Agricultural sciences	1	–
Social sciences	14	4
Humanities	2	3
Data collection		
Interview, survey, Delphi method, etc.	5	7
Actual reviews and comments	19	5
Number of criteria per study		
Minimum	8	7
Maximum	223	66
Mean	44	26
Median	19.5	21

Studies on manuscript criteria on average report more criteria than studies on grant criteria (44 and 26, respectively). In particular, while the study that reports the most grant criteria (Pollitt et al., 1996) lists 66 criteria, there are six studies on manuscripts that list more criteria. For example, Campion (1993) lists no less than 223 criteria for reviewing research articles in applied psychology. A possible reason for this difference could be the strong improvement focus of the manuscript review process, which could promote more detailed comments of reviewers or prompt authors of studies on manuscript criteria to perform more fine-grained analyses. If, however, one ignores the studies that report a large number of criteria (i.e. those larger than the median), a similar pattern emerges: 50% of the manuscript and grant studies report 8 to 19 and 7 to 21 criteria, respectively.

Preliminary conclusions

Although there are tens of thousands of publications on peer review (see Batagelj et al., 2017) and although criteria are an essential component of any evaluation process, there are only very few studies that focus on criteria peers actually use or prefer. In particular, 24 inductive studies on manuscript review criteria and 12 inductive studies on grant review criteria were identified in the systematic review. With respect to research fields, the systematic review showed that most studies analysed criteria in the medical and health sciences and in the social sciences. These findings suggest that there is a need for more studies on peer review criteria in general and more studies on the natural sciences and humanities in particular. In addition, future studies should develop a comparative perspective to improve the understanding of the commonalities and peculiarities of the evaluation cultures of different fields and disciplines. From a practical standpoint, studies on peer review

criteria are relevant as they contribute to increasing the transparency of peer review processes and they support early career researchers in learning the basics of peer assessment.

References

References with one asterisk (*) indicate studies on manuscript peer review, references with two asterisks (**) indicate studies on grant peer review.

- **Abdoul, H., Perrey, C., Amiel, P., Tubach, F., Gottot, S., Durand-Zaleski, I., & Alberti, C. (2012). Peer review of grant applications: criteria used and qualitative study of reviewer practices. *PLoS ONE*, 7(9), 1–15. <https://doi.org/10.1371/journal.pone.0046054>
- *Albers, C. A., Floyd, R. G., Fuhrmann, M. J., & Martinez, R. S. (2011). Publication criteria and recommended areas of improvement within school psychology journals as reported by editors, journal board members, and manuscript authors. *Journal of School Psychology*, 49(6), 669–689. <https://doi.org/10.1016/j.jsp.2011.10.002>
- *Bakanic, V., McPhail, C., & Simon, R. J. (1989). Mixed messages: referees' comments on the manuscripts they review. *The Sociological Quarterly*, 30(4), 639–654. <https://doi.org/10.1111/j.1533-8525.1989.tb01540.x>
- Baldwin, M. (2017). In referees we trust? *Physics Today*, 70(2), 44–49. <https://doi.org/10.1063/PT.3.3463>
- Baldwin, M. (2018). Scientific autonomy, public accountability, and the rise of “peer review” in the Cold War United States. *Isis*, 109(3), 538–558. <https://doi.org/10.1086/700070>
- Batagelj, V., Ferligoj, A., & Squazzoni, F. (2017). The emergence of a field: a network analysis of research on peer review. *Scientometrics*, 113(1), 503–532. <https://doi.org/10.1007/s11192-017-2522-8>
- *Becker, B. (1991). The quality and credibility of research reviews: What the editors say. *Personality and Social Psychology Bulletin*, 17(3), 267–272. <https://doi.org/10.1177/0146167291173006>
- *Bonjean, C. M., & Hullum, J. (1978). Reasons for journal rejection. An analysis of 600 manuscripts. *PS: Political Science & Politics*, 11(4), 480–483. <https://doi.org/10.1017/S1049096500004844>
- *Bordage, G. (2001). Reasons reviewers reject and accept manuscripts: The strengths and weaknesses in medical education reports. *Academic medicine*, 76(9), 889–896. <https://doi.org/10.1097/00001888-200109000-00010>
- *Campion, M. A. (1993). Article review checklist. A criterion checklist for reviewing research articles in applied-psychology. *Personnel Psychology*, 46(3), 705–706. <https://doi.org/10.1111/j.1744-6570.1993.tb00896.x>
- Chase, J. M. (1970). Normative criteria for scientific publication. *American Sociologist*, 5(3), 262–265.
- *Chong, C. (1994). Editorial: We aim to publish – not to reject. *Canadian Journal of Plant Science*, 74(4). <https://doi.org/10.4141/cjps94-120>

- *Coniam, D. (2011). Systematising System: one reviewer's analysis of the review process. *System*, 39(4), 539–553. <https://doi.org/10.1016/j.system.2011.10.018>
- *Daft, R. L. (1995). Why I recommended that your manuscript be rejected and what you can do about it. *Publishing in the Organizational Sciences*, 164–183. <https://doi.org/10.4135/9781452240466.n14>
- *Day, F. C., Schriger, D. L., Todd, C., & Wears, R. L. (2002). The use of dedicated methodology and statistical reviewers for peer review: A content analysis of comments to authors made by methodology and regular reviewers. *Annals of Emergency Medicine*, 40(3), 329–333. <https://doi.org/10.1067/mem.2002.127326>
- *Dickersin, K., Ssemanda, E., Mansell, C., & Rennie, D. (2007). What do the JAMA editors say when they discuss manuscripts that they are considering for publication? Developing a schema for classifying the content of editorial discussion. *BMC Medical Research Methodology*, 7, Article 44. <https://doi.org/10.1186/1471-2288-7-44>
- *Ezeala, C., Nweke, I., & Ezeala, M. (2013). Common errors in manuscripts submitted to medical science journals. *Annals of medical and health sciences research*, 3(3), 376–379. <https://doi.org/10.4103/2141-9248.117957>
- *Fiske, D. W., & Fogg, L. (1990). But the reviewers are making different criticisms of my paper!: diversity and uniqueness in reviewer comments. *American Psychologist*, 45(5), 591–598. <https://doi.org/10.1037/0003-066X.45.5.591>
- Frantz, T. T. (1968). Criteria for publishable manuscripts. *Personnel and Guidance Journal*, 47(4), 384–386. <https://doi.org/10.1002/j.2164-4918.1968.tb02944.x>
- *Gogolin, I. (2011). *European Educational Research Quality Indicators (EERQI). Project final report*. http://eerqi.eu/sites/default/files/Final_Report.pdf
- **Guetzkow, J., Lamont, M., & Mallard, G., (2004). What is originality in the humanities and the social sciences? *American Sociological Review*, 69(2), 190–212. <https://doi.org/10.1177%2F000312240406900203>
- Hartmann, I. (1990). *Begutachtung in der Forschungsförderung: die Argumente der Gutachter in der Deutschen Forschungsgemeinschaft*. RG Fischer.
- **Hartmann, I., & Neidhardt, F. (1990). Peer review at the Deutsche Forschungsgemeinschaft. *Scientometrics*, 19(5–6), 419–425. <https://doi.org/10.1007/BF02020704>
- *Hesterman, C. M., Szperka, C. L., & Turner, D. P. (2018). Reasons for manuscript rejection after peer review from the journal headache. *Headache: The Journal of Head and Face Pain*. <https://doi.org/10.1111/head.13343>
- *Hewings, M. (2004). An ‘important contribution’ or ‘tiresome reading’? A study of evaluation in peer reviews of journal article submissions. *Journal of Applied Linguistics*, 1(3), 247–274. <https://doi.org/10.1558/jal.v1i3.247>

- Hug, S. E., & Aeschbach, M. (2019). Criteria for assessing grant applications. A systematic review. *Palgrave Communications*, 6, Article 37. <https://doi.org/10.1057/s41599-020-0412-9>
- Hug, S. E., Holowiecki, M., Ma, L., Aeschbach, M., & Ochsner, M. (in preparation). *Criteria for reviewing journal manuscripts. A systematic review.*
- **Lahtinen, E., Koskinen-Ollonqvist, P., Rouvinen-Wilenius, P., Tuominen, P., & Mittelmark, M. B. (2005). The development of quality criteria for research: A Finnish approach. *Health Promotion International*, 20(3), 306–315. <https://doi.org/10.1093/heapro/dai008>
- **Lamont, M. (2009). *How professors think: inside the curious world of academic judgment.* Harvard University Press.
- *Leung, D., Law, R., Kucukusta, D., & Guillet, B. D. (2014). How to review journal manuscripts: A lesson learnt from the world's excellent reviewers. *Tourism Management Perspectives*, 10, 46–56. <https://doi.org/10.1016/j.tmp.2014.01.003>
- *McKercher, B., Law, R., Weber, K., Song, H., & Hsu, C. (2007). Why referees reject manuscripts. *Journal of Hospitality and Tourism Research*, 31(4), 455–470. <https://doi.org/10.1177/1096348007302355>
- Moxham, N., & Fyfe, A. (2018). The Royal Society and the prehistory of peer review, 1665-1965. *Historical Journal*, 61(4), 863–889. <https://doi.org/10.1017/S0018246X17000334>
- **Pier, E. L., Brauer, M., Filut, A., Kaatz, A., Raclaw, J., Nathan, M. J., Ford, C. E., & Carnes, M. (2018). Low agreement among reviewers evaluating the same NIH grant applications. *Proceedings of the National Academy of Sciences of the United States of America*, 115(12), 2952–2957. <https://doi.org/10.1073/pnas.1714379115>
- **Pollitt, F. A., Notgrass, C. M., & Windle, C. (1996). Mental health services research. Peer review of rural research grant applications. *Administration and Policy in Mental Health and Mental Health Services Research*, 24(2), 173–180. <https://doi.org/10.1007/BF02042489>
- *Ramulu, V. G., Levine, R. B., Hebert, R. S., & Wright, S. M. (2005). Development of a case report review instrument. *International Journal of Clinical Practice*, 59(4), 457–461. <https://doi.org/10.1111/j.1368-5031.2005.00319.x>
- **Reinhart, M. (2010). Peer review practices: A content analysis of external reviews in science funding. *Research Evaluation*, 19(5), 317–431. <https://doi.org/10.3152/095820210X12809191250843>
- **Schmitt, J., Petzold, T., Nellesen-Martens, G., & Pfaff, H. (2015). Priorisierung und Konsentierung von Begutachtungs-, Förder- und Evaluationskriterien für Projekte aus dem Innovationsfonds: Eine multiperspektivische Delphi-Studie. [Prioritization and consentation of criteria for the appraisal, funding and evaluation of projects from the German Innovation Fund: a multi-perspective Delphi study]. *Gesundheitswesen*, 77(8–9), 570–579. <https://doi.org/10.1055/s-0035-1555898>

- *Smith, M. U., Wandersee, J. H., & Cummins, C. L. (1993). What's wrong with this manuscript? An analysis of the reasons for rejection given by Journal of Research in Science Teaching reviewers. *Journal of Research in Science Teaching*, 30(2), 209–211. <https://doi.org/10.1002/tea.3660300207>
- **Thomas, J. P., & Lawrence, T. S. (1991). Common deficiencies of NIDRR research applications. *American Journal of Physical Medicine & Rehabilitation*, 70(1), 161–164. <https://doi.org/10.1097/00002060-199004000-00005>
- *Turcotte, C., Drolet, P., & Girard, M. (2004). Study design, originality and overall consistency influence acceptance or rejection of manuscripts submitted to the Journal. *Canadian Journal of Anaesthesia-Journal Canadien d'anesthésie*, 51(6), 549–556. <https://doi.org/10.1007/BF03018396>
- **van Arensbergen, P., van der Weijden, I., & van den Besselaar, P. (2014). Different views on scholarly talent: What are the talents we are looking for in science? *Research Evaluation*, 23(4), 273–284. <https://doi.org/10.1093/reseval/rvu015>
- *van Lent, M., IntHout, J., & Out, H. J. (2015). Peer review comments on drug trials submitted to medical journals differ depending on sponsorship, results and acceptance: A retrospective cohort study. *BMJ Open*, 5(9), e007961. <https://doi.org/10.1136/bmjopen-2015-007961>
- **Whaley, A. L., Rodriguez, R., & Alexander, L. A. (2006). Development of a rating form to evaluate grant applications to the Hogg Foundation for Mental Health. *Evaluation Review*, 30(1), 3–26. <https://doi.org/10.1177/0193841X05275586>
- *Yarbrough, C., & Whitaker, J. A. (2009). Analysis of reviewer comments about quantitative manuscripts accepted by the Journal of Research in Music Education. *Journal of Research in Music Education*, 56(4), 287–292. <https://doi.org/10.1177/0022429408329448>
- *Zaruba, K. E., Toma, J. D., & Stark, J. S. (1996). Criteria used for qualitative research in the refereeing process. *Review of Higher Education*, 19(4), 435–460. <https://doi.org/10.1353/rhe.1996.0017>

Practices of peer review in the SSH II: Peer review and other manuscript selection processes for books in the SSH

By Elea Giménez-Toledo & Jorge Mañana-Rodríguez

Introduction

The relevance of books in the communication of research results in social sciences and humanities (SSH) has been shown in numerous occasions. In terms of market, the centrality of books in SSH is evident. Just as an example, in the Spanish market (5th publishing potential worldwide), 81% of scholarly books are within the SSH (Giménez-Toledo, 2017).

The selection of the original manuscripts which will eventually arrive at the market in book format is based on a combination of criteria of different nature. Among them, there is, undoubtedly, the scientific quality and rigor. The scientific value of a manuscript, its originality and its contribution to new knowledge are variables which directly intervene in the decision to publish or not a scholarly book. But those are not the only ones. Issues such as the potential number of readers, the timeliness of the topic, the reputation of the author, the editorial cost of the book or the expected sales are highly relevant in decision making.

While in the case of scientific journals, the reports of external reviewers (peer review) are determining in the decision to publish or not an article, in scholarly publishing more generally, there are more issues to consider and the procedures for decision making are different. Among other things, the different types of publishers schedule the composition of their catalogues differently. This has been shown in the studies carried out concerning Spanish and Latin American publishers (Giménez-Toledo & Córdoba-Restrepo, 2018). University Presses use peer review both in journals and scholarly books. A clear influence regarding this feature is the fact that the starting point of university presses is the academia and, therefore, the standards followed are those pertaining to the well-established tradition in the scientific community. Other contributing factors are the stability and attachment of the university press to the university, the objective of university presses with regards to the transmission of knowledge in different fields (the catalogues of university presses tend to be highly multidisciplinary), that it has to be self-sustaining but not necessarily provide benefit, and that it is managed by university lecturers who temporarily assume the responsibilities attached to the editorial project. It can be said there is an institutional editorial project but sometimes it does not have the means for engaging the very challenging questions of marketing scholarly books.

Imprints belonging to large multinational publishing companies tend to follow scholarly orthodoxy; manuscript selection processes are based on peer review. The author can perceive that and gets the information through the publishers' guidelines. Nevertheless, the assessment of the scientific quality of the manuscript is not the only criteria taken into account, but the sales estimation as well as its profitability are also important.

These circumstances are very different from those of small and medium-sized commercial or private publishers. Those businesses should create highly selective and competitive catalogues for a market saturated with titles. The heads of such publishers tend to be editors specialized in a field and tend to create a specialized catalogue with a specific focus, directed towards a certain type of reader. In order to achieve that, they should carry on a very important task of identification of authors and topics as well as surround themselves with

close experts who would help the editor with the creation of the catalogue within the topics, authors and focus which are distinctive for the publishing project. Also, the selection of peer reviewers is an important task. The power in the peer review process appears to be held by the reviewer. Reviewers are very hard to find. The stronger the reputation of an expert, the more difficult it is to get them engaged in peer review. In this type of companies, the consultation to trustworthy experts, the establishment of internal reading committees and an active search of original manuscripts are common procedures for the decision making on the titles which would be part of the catalogue. The decisions taken are critical: a set of wrong decisions can lead to bankruptcy.

In sum: peer review is one of the manuscript selection processes but not the only one in the creation of catalogues of scholarly books. The existence of other procedures does not imply lack of quality or control of the scientific rigor but responds to the specific features of the publishing companies or institutions. For this purpose and despite the fact that peer review is a selective criterion in the evaluation of publishers (Giménez-Toledo et al., 2017), it is necessary to consider the existing differences between journals and scholarly books.

Some data concerning the manuscript selection procedures in scholarly publishers

The following section provides general data for Spanish and Latin American publishers concerning the existing information available at the SPI section on manuscript selection processes.

Using a survey methodology, an interactive chart, publicly available, has been progressively completed through the incorporation of the information provided by the publishers on their manuscript selection processes (MSP).

It is relevant to mention that the data reflected corresponds to the responses of publishers who have explicitly authorized its public availability at the SPI Information system. The chart contains the names of the publishers, followed by six different manuscript selection processes:

- Reading report (from the series team)
- Reading report (from the publisher)
- Director of series
- Director of publishing house/university press (Decision/opinion of)
- Experts external to the series
- External reviewers (with regards to the publishing house)

As shown in Figure 2, for each publisher, it is possible to identify which MSP they use and, conversely, for each MSP the user can delimit the publishers which have declared to use it (the chart allows the ordering of all the elements according to that of any of the variables). Finally, the last column reflects the existence of publicly available information on the MSP used by the publisher in their websites. This information has been checked for each case individually

Sistemas de revisión de originales

Procedimientos de selección de originales en editoriales académicas españolas y latinoamericanas

Las editoriales académicas que constan en el gráfico muestran qué procedimientos de selección de originales siguen entre los propuestos (seis primeras columnas). Esta información ha sido recabada mediante encuesta. La última columna indica si la editorial ofrece alguna información en su sitio web sobre sus procedimientos de selección de originales.

Editorial	Comité de lectura de la colección	Comité de lectura de la editorial	Directora de la colección	Directora de la editorial	Expertos externos a la colección	Expertos externos a la editorial	Información sobre sistema de revisión de originales en su web
ADAPT University Press							
AECID	○	○			○	○	
Academia del Hispanismo			○			○	○
Aconcagua Libros	○	○	○				
Aduana Vieja		○	○	○		○	
Agencia Estatal Boletín Oficial del Estado							○
Albatros Ediciones			○		○	○	
Alianza Editorial			○	○		○	
Almud, Ediciones de Castilla-La Mancha		○			○	○	
Andavira			○	○	○	○	
Anthropos		○	○	○	○	○	
Antoni Bosch Editor, S.A.							○
Antonio Carlos Novas	○	○				○	
Archibald editores	○	○	○	○	○	○	

Figure 2. Snippet of the upper section of the chart reflecting MSP at SPI.

Currently, there are Spanish (175) and Latin American (34) publishers in the chart, the total number of publishers being 209. Also, there is a wide diversity of publisher types: from highly specialized publishers belonging to large publishing groups to university presses, including also small, independent publishers and publication services of several public and private institutions.

With regards to one of the main segmentation variables, the condition of University Press (UP, total number: 63) or other type of publisher (total number: 146), some differences can be observed in the percentage of use of each MSP: As can be seen in Figure 3, external reviewers as well as experts external to the series are more frequent in the case of UPs, while the opinion of the director of the publishing house or the director of the series are clearly more prominent in the case of non-UP publishers.

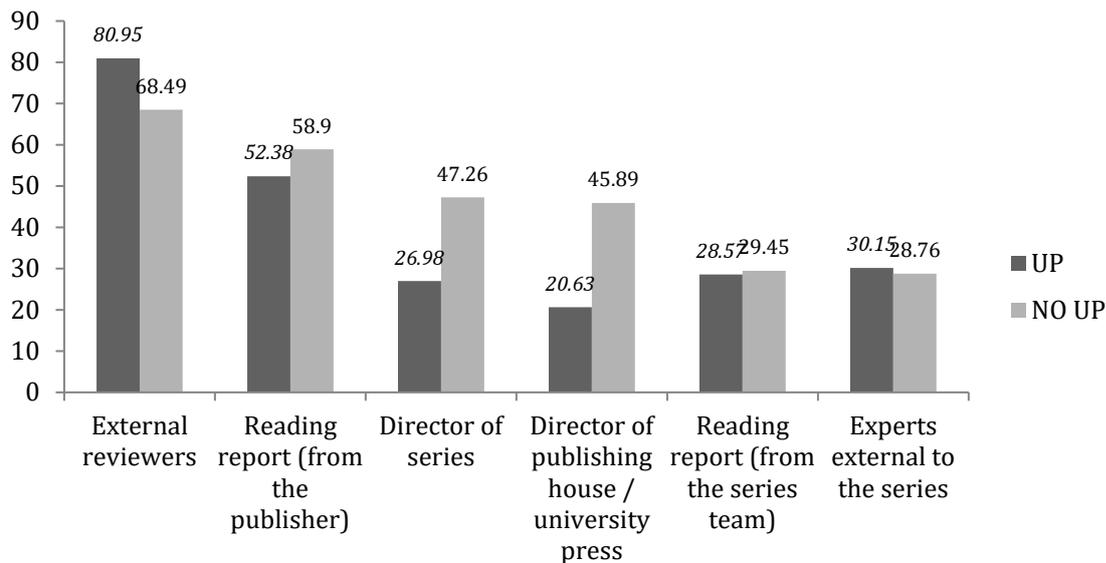


Figure 3. Percentages of each MSP by publisher type (UP or other types of publishers).

This can be interpreted in several ways, but the use of external reviewers seems consistent with the scholarly tradition, and thus more likely in publishers closely attached to a higher

education institution while the decisions taken by the head of the publisher are probably more consistent with entrepreneurial or business practices.

Another relevant difference, shown in Figure 4, between UP and non-UP with regards to MSPs is the availability of information on the MSP used by the publisher on their websites. Diversity in the procedures, approaches and strategies is a constant in the study of peer review in scholarly books.

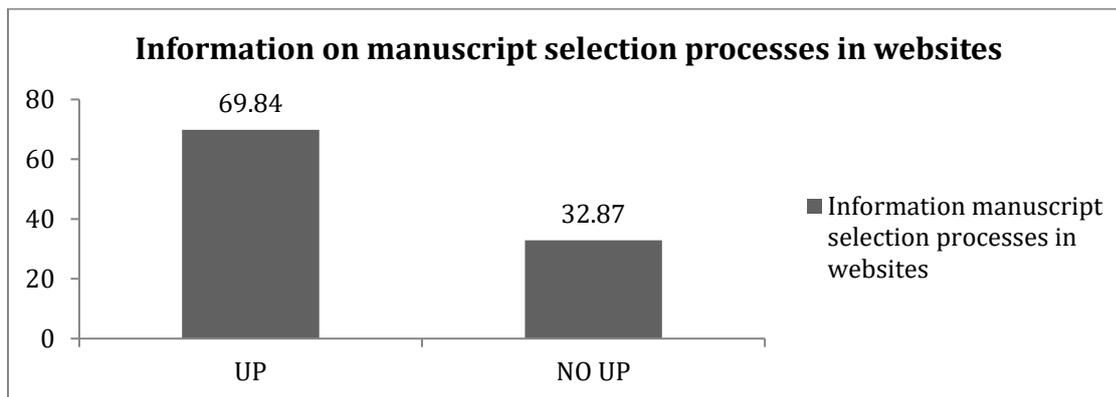


Figure 4. Public availability of information on MSP on UP and non-UP publishers' websites.

References

- Giménez-Toledo, E. (Ed.) (2017). *La edición académica española. Indicadores y características*. Madrid: Federación de Gremios de Editores de España. Informe. <http://ilia.cchs.csic.es/SPI/spi-fgee/docs/EAEV1.pdf>
- Giménez-Toledo, E., & Córdoba-Restrepo, J. F. (Eds.). (2018). *Edición académica y difusión. Libro abierto en Iberoamérica*. Universidad del Rosario. <https://doi.org/10.12804/th9789587841671>
- Giménez-Toledo, E., Sivertsen, G., & Mañana-Rodríguez, J. (2017). Peer review as a delineation criterion in data sources for the assessment and measurement of scholarly book publishing in social sciences and humanities. In *Proceedings of the 16th International Conference on Scientometrics and Informetrics ISSI*, 16.-20. October at Wuhan University (pp. 118–124). Wuhan University.

Practices of peer review in the SSH III: peer review in the legal domain. Three parallel case studies in Italy, Spain and Croatia

By Ginevra Peruginelli, Elías Sanz-Casado & Jadranka Stojanovski

The Context

Evaluating the scientific quality of legal publications is a central debate in legal academia at the international level (Flückiger & Tanquerel, 2015; Peruginelli & Faro, 2018). Research evaluation in legal science is a delicate and complex process due to the fact that legal disciplines are not monolithic: there are profound differences between the various branches of law, for which the tools of communication are very different. Legal scholarship is both the science of law and one of the authoritative and influencing sources of that law. This is why there is a strict correlation between legal science and legal practice (Gutwirth, 2010). For historical, epistemological and economic reasons, in legal sciences the peer review it is not widespread because the communities of jurists are self-referential, little cohesive and very fragmented (Peruginelli, Faro & Agnoloni, 2018). Until recently, the same concept of independent quality assessment was not covered. Law journals always have an editorial board that has the task of accepting or not accepting contributions. In most cases, however, it does not make use of external and independent referees. Furthermore, most law journals do not indicate the criteria/indicators for assessing the publication of contributions (van Gestel & Vranken, 2011).

In this delicate context, three national surveys have been carried out (one in Italy, one in Spain and one in Croatia), all of which provided a picture of peer review procedures for the evaluation of scholarship included in legal periodicals.

For each country, a brief explanation about the survey and the methodology used to collect data is provided. The surveys conducted have yielded interesting results that are presented in a comprehensive manner, followed by some reflections on the presented results.

Italy

The Italian survey was conducted by the Institute of Legal Information Theory and Techniques of the National Research Council (ITTIG-CNR) from September 2017 to March 2018. The survey aimed to provide an overview of the situation of the review process of the highly rated law journals (Class A) published in Italy (102 journals).

The method used for the survey was the material observation of the evaluation procedures to assess contributions declared by the law journals. The information has been retrieved by the last two issues of the analysed periodicals and the identification of the information on peer review has not always been immediate and simple. In most cases the information on evaluation procedures were found at the beginning or at the end of each contribution and in some cases, the journal did not show any information. It is to be noted that few journals (6) did not present any reference to an evaluation process and that certain contributions were not subject to evaluation, as they are written by scholars of prestige and clear fame, which is a very vague expression used in some journals. In this case, only the editor is responsible for the publication quality. We omitted such journals from the analysis and selected the highly rated law journals published in Italy that present indications on the evaluation procedures. Therefore, we investigate 96 journals.

Some results

The survey has investigated the presence of internal and/or external referees. Internal referees are usually chosen among members of the scientific committee or editorial board, while the external referees are recruited among professors of specific branches of law and scholars of clear fame and prestige. The majority of journals (36 periodicals) does not indicate whether the referees are external or internal.

It is important to notice also that 23 s out of 96 journals give importance to a preliminary phase, which is preparatory to the actual evaluation activity by one or more referees. After being accepted, the article can follow the next step of the proper evaluation process. Otherwise, the contribution is sent back to the author with the indications of the reasons for rejection or with suggestions for minor or major changes to be made in order for the paper to be accepted. Furthermore, only 10 journals explicitly provide information on the existence and conservation of the evaluation report on the contribution, and its availability to the management committee, the editorial staff or the publisher.

Regarding the number of referees, only 18 journals state that the assessment is carried out by a single referee. About twice as many journals (34 journals) report the implementation of the more than one referees procedure. 18 journals say that the evaluation is conducted by “more than two referees or by a committee of peers”; other journals (22), although referring to the evaluation procedure and the assignment of evaluation to referees, do not explicitly expressly indicate the number of referees. Finally, only four journals specifically present the indication of the number of referees (one or two) based on the position of the article in a specific section of the journal and/or on a specific topic

The survey continues showing that 48 periodicals, while referring to the presence of an anonymous evaluation procedure, do not provide any indication of the type of anonymity (single-blind, double-blind or open peer review) of the reviewers and authors. Only two journals report a mix between single and double-blind peer review depending on the type of contribution to be assessed.

Finally, the survey examined whether the periodicals say something on quality criteria by which reviewers decide to accept or not a contribution for publication. Only 16 journals refer to qualitative evaluation criteria such as relevance of the subject, correctness of the methodological approach, adequacy of the essential bibliography, order and clarity, originality of the contribution, practical utility of the ideas expressed, adequacy of the documentation.

The majority of the journals does not provide any information on the criteria to be followed by the reviewers. Most of the current law journals and publishers have not made it very clear what they expect from authors in terms of methodological accountability (e.g. to what extent should author explain their research design?) and what is expected from referees in terms of feedback and response to criticism from scholars who have submitted contributions for peer review.

Reflections

The described Italian survey on law journals provides a framework, which empirically confirms the state of the Italian debate on the evaluation of scientific legal outputs. The results gave an overview of the landscape of the peer review process adopted or better to say indicated in the Italian law journals. This clarification is quite important because these numbers are the result of the available information gathered from single issues of each

periodical. It is to be noticed that on the basis of the collected data, criteria or indicators do not yet seem to exist. There is not a well-defined peer-review culture in the Italian law journals' context. Legal publishers and academics mostly fear that the introduction of peer review might bring about time constraints and bureaucracy. Perhaps it is safe to expect that the introduction of effective peer review will develop gradually since more and more scholars apparently focus on publishing for an international audience in English. Of course, these few brief considerations are placed in a cultural context of constant change, with the consequent need for continuous repositioning of this study, future investigations and experiments in the field of evaluation of legal science.

Spain

The number of Spanish law journals considered in this analysis is 44. These journals have been selected because they have been accredited with the Quality Seal of the Spanish Foundation for Science and Technology (FECYT), which is granted to Spanish journals of any area of knowledge that have overcome 14 quantitative and qualitative indicators. Some of these indicators are: opening of the editorial board, external peer review, internationality of the advisory board, number of databases in which the journals are indexed, and impact and visibility of the journal in the popular databases (WOSCC and SCOPUS).

Some results

The analysis of the indicators obtained from these 44 journals showed that, regarding the type of evaluation, 34 journals (or 77%) evaluated the manuscripts by referees external to the journal, while in five journals (11%), the evaluation was carried out by internal and external referees at the same time. Only one of these journals evaluated manuscripts by internal referees only (2%). Four journals (9%) did not provide any information on the review systems they use to evaluate the manuscripts they publish.

Regarding the number of referees participating in the evaluation processes, in most of the journals, two referees evaluate each manuscript (23 journals or 52%). Furthermore, of these 23 journals, seven send the manuscript to a third referee in case the two reviews are very different (16% of the journals). Only three journals send the manuscript to only one referee for its evaluation, which is a very low share (7 of the cases). Only one of the analysed journals (2%) sends its manuscripts to three or more referees for its evaluation. However, we must highlight the high percentage of Spanish law journals that do not mention the number of referees that evaluate their manuscripts (23% or 10 journals).

Another aspect analysed in the evaluation process of Spanish law journals was the type of review. The results showed that most of the journals (68% or 30 journals) used the double-blind system. Only four 4 journals used a single-blind procedure (9%). A high percentage of journals (23% or 10n journals) did not mention any system for the revision of manuscripts.

Reflections

The analysis of the indicators obtained about the review processes of Spanish law journals accredited by the quality processes of FECYT shows that a high percentage of the journals (77%) carry out the evaluation of the manuscripts through external referees. This shows the important change that is taking place in the processes of quality and editorial transparency of Spanish law journals. However, it is still worrying that 9% of journals do not offer

any information about this review process, or 11% carry the review process out in an endogenous manner.

These data are consistent with those obtained from the indicator of the number of referees participating in the evaluation processes, since 68% of the journals send the manuscript to at least two referees for its evaluation, and 16% send it to a third referee in case of disagreement. However, a high percentage of journals (23%) does not mention the number of referees participating in the manuscript evaluation processes.

If we take into account that the analysed journals of law are those that have a higher quality in their editorial systems, we still have to do an important work in improving the transparency of and commitment to the manuscript revision processes by national journals in order they can play an important role in the dissemination of scientific knowledge.

Regarding the type of review, the results come to strengthen those obtained with the previous indicators since the double-blind system is the most used by the journals studied (68%). Nevertheless, it is necessary to emphasise the high percentage of journals that do not mention the type of evaluation (23%). This result reaffirms the aforementioned about the effort that still needs to be made with many editors to update and report on the evaluation systems they perform.

Croatia

All Croatian law journals are included in the central open access repository of Croatian scientific and research journals HRČAK⁹. HRČAK is an excellent tool for the promotion of the journal editorial policies, offering a uniform and transparent approach and good ground for the exchange of practices and improvements of the scholarly communication system. HRČAK includes 474 open access journals from all disciplines, of which 50 claim to cover law. Among these 50 journals, 11 cover law only, and other journals also include other disciplines, mainly from social sciences. We have excluded from the further analysis interdisciplinary journals, which cover a broad range of disciplines, and law articles represent only a small portion of published articles. We also excluded 5 law journals because of the absence of the information on peer review. Obviously, the peer review process is not considered as an inevitable phase of scholarly publishing by all law journals. Inactive journals which did not publish a new issue during past 13 months were excluded too. Finally, 32 active peer-reviewed journals covering law and published by higher education research institutions and professional societies were selected for the analysis.

Every journal included in the HRČAK repository is described by a rich set of highly structured metadata, including information on peer review. Metadata are provided by journal editors who have their administrators' accounts at HRČAK system. Peer review is described by:

1. scope: internal/editorial reviewers, external reviewers;
2. reviewers' provenance: national (Croatian) reviewers, international reviewers;
3. number of reviewers: one, two, three or more reviewers;
4. openness/blindness of the peer review process: single-blind, double-blind, not blind, open peer review;

⁹ <https://hrcak.srce.hr/?lang=en>

5. type of papers which undergo peer review process: scientific, professional, all papers.

The data on 32 law journals were collected during February 2019.

Some results

Only one journal employs internal/editorial peer review. All other journal editors send their manuscripts to the external reviewers. In this research, we consider internal/editorial peer review as review done by editors, and it should not be mixed with triage of the manuscripts where editors are deciding upon immediate rejection of the submission or sending it out for peer review. During the reviewer selection process, editors can choose a reviewer from Croatia (national) or abroad (international).

Among 32 selected law journals, only one publishes articles only in English. All other journals publish besides Croatian also in other languages: English, French, German, Italian, Bosnian, Serbian etc. For editors of Croatian journals, it is difficult to find suitable referees willing to do a review for a small and relatively unknown journal. This challenge is even more significant for manuscripts written in Croatian language understood only by Croatian peers coming from a small research community. As a result of all challenges concerning the selection of referees, manuscripts from 19 law journals are sent to the reviewers coming from Croatia, 4 journals select reviewers mainly coming from abroad, while 9 journals select national and international reviewers equally, usually one national and one international per submission.

Regarding the number of referees, the majority of manuscripts are reviewed by two referees (28 journals). Only 3 journals provide only one referee per manuscript and only one provides three or more reviewers per manuscript. The provided number of reviewers refers to the editorial practice that is mostly applied. For journals employing two referees per manuscript, if two reviewers disagree the editor sends the manuscript to a third reviewer and does whatever he or she advises.

Describing blindness/openness of the peer review process, editors of HRČAK journals have four options at disposal, and can select one of them: single-blind – where reviewers' identities are kept hidden from authors; double-blind – where identities of both the authors and reviewers are kept hidden; not blind – where the identities of the reviewer and authors are known to each other during the peer review process, but the reviewer's identities are kept hidden after publishing; and open peer review – where authors see the identity of their reviewers and vice versa, reviewers' reports are publicly available, and the reviewer's identity could be disclosed after publishing.

According to the data, 26 law journals employ double-blind peer review, 4 journals use single-blind peer review and 2 journals not blind peer review. By applying a double-blind peer review, editors want to achieve a fairer and more objective assessment of the manuscript. If the authors' identity is unknown to the reviewer, it will prevent the reviewer from forming any bias (e.g. gender bias or standing within research community bias). To what extent double-blind peer review is really blind, and to what extent open peer review could be a solution for resolving disadvantages in the assessment within a small scientific community, is to be discussed. Currently, no Croatian journal does employ an open peer review process (Hebrang-Grgić & Stojanovski, 2017).

Concerning the types of paper which undergo a peer review process, there is a variety of practices among Croatian law journals. 7 journals send to referees only scientific papers, 14 journals send scientific and professional papers, and 11 journals send all papers to be reviewed. Most of the Croatian journals make a distinction between original scientific papers, which must present original results of a scientifically conceived and conducted research, and professional papers, which bring knowledge on already published scientific results and do not bring new knowledge to the profession. It is recommended for professional papers to undergo peer review too, but some journals do not follow this practice.

Reflections

Law journals in Croatia are in open access and well described through the common HRČAK platform. According to the analysis of the data on peer review provided by the editors of the 32 Croatian law journals listed in the HRČAK repository of open access journals, law journals employ mostly two external peer reviewers coming from Croatia, the peer review process is mostly double-blind and most of the law journals send scientific and professional articles for peer review.

We believe that the quality and transparency of the peer review process, and consequently the quality of scholarly communication in Croatia, could be improved by the introduction of open peer review. Open peer review will include publicly available reviews for all scientific and professional papers, more reviewers from abroad (not necessarily with disclosed identities), public comments and provide a platform for the academic discussion, strongly needed in Croatia.

In addition, we would like to see more journals providing the data underlying the scientific findings they publish. When an author fields a survey or experiment over a period of time, the results are usually presented in the condensed and/or reduced form by charts, images and tables. If reviewers want to check how authors produced the results, they will want to examine the data that underlies the results. This is especially important since most of the provided charts and images are only available as bitmap images. To perform the peer review effectively, it is crucial that peer reviewers are given unlimited access to the data underlying the research they are reviewing. Unfortunately, most journals do not provide access to this material during the peer review process, making it almost impossible to perform a reliable peer review function. In the future, we would like to have all data and related metadata underlying the findings reported in a submitted manuscript deposited in an appropriate repository and available for the reviewers. Journals should mandate that research data be shared at least with reviewers.

Conclusions

These three surveys are significant despite the fact that the number of law journals under investigation is not high. It is, however, interesting to note that the three different countries have more or less a similar approach to peer review, with a more rigorous approach in Croatia and Spain where most of the law journals used the double-blind peer review procedure composed of external referees. On the other side, it is interesting to note that in Italy, there is a lack of homogeneity in the types of peer review used and in the number of referees. Further analysis could be carried out in this direction, also from the side of the authorship, so to investigate the perception of the authors.

This short text is just a picture of the state of the art of peer review procedure used in the law domain in Italy, Spain and Croatia and of the sharp change with respect to the past that goes towards the direction of a more rigorous and transparent peer review process in the law domain. This change gives the opportunity to the social peer review to play an important role. In the digital world and, above all, in the social web, the revision mechanism becomes open, dialogic, less self-referential (not only academics). In this case, the quality assessment is carried out exclusively ex-post through the social tools (blog, wiki etc.). These evaluation tools seem to be suitable for the characteristics of legal research, which has its strong point in dialectics, in communication and in opening the debate towards the outside, so to offer a possible alternative to the ex-ante qualitative review system, maybe never really consolidated in the legal science.

References

- Flückiger, A., & Tanquerel, T. (Eds.). (2015). *L'évaluation de la recherche en droit / Assessing research in law. Enjeux et méthodes / Stakes and methods*. Bruylant.
- Gutwirth, S. (2010). The evaluation of legal science. The V.I.R.-model for integral quality assessment of research in law: What next? In H. Eisendrath, J. P. van Bendegem (Eds.), *It takes two to do science. the puzzling interactions between science and society* (pp. 69–79). ASP.
- Hebrang Grgić, I., & Stojanovski, J. (2017). Peer review in Croatian open access scholarly and professional journals: A cross-disciplinary survey. In *Eighth International Congress on Peer Review and Scientific Publication: Enhancing the quality and credibility of science*. Chicago. <https://peerreviewcongress.org/prc17-0364>
- Peruginelli, G., & Faro, S. (2018). Research quality evaluation: The case of legal studies. In A. Bonaccorsi (Ed.), *The evaluation of research in social sciences and humanities* (pp. 103–129). Springer International Publishing. https://doi.org/10.1007/978-3-319-68554-0_5
- Peruginelli, G., Faro, S., Agnoloni, T. (2018). A survey on legal research monograph evaluation in Italy. In A. Bonaccorsi (Ed.), *The evaluation of research in social sciences and humanities* (pp. 211–230). Springer International Publishing. https://doi.org/10.1007/978-3-319-68554-0_9
- van Gestel, R., & Vranken, J. (2011). Assessing legal research: Sense and nonsense of peer review versus bibliometrics and the need for a European approach. *German Law Journal*, 12(3), 901–929. <https://doi.org/10.1017/S2071832200017144>

PART IV: Current Challenges for Peer Reviewing: Towards More Open and Gender-Sensitive Peer Reviewing Practices in the SSH

A gender and geopolitical perspective on peer review

By Karolina Lendák-Kabók & Michael Ochsner

Introduction

Gender biases in academic work have received a great deal of scholarly attention recently. The great majority of the extant research focuses on academic women's achievements; authors highlight gender differences in success to obtain a permanent position (Dubois-Shaik & Fuselier, 2017; Morley, 1999, Waaijer et al., 2016), or to complete high quantity and quality of publications (Kretschmer et al., 2012). Peer review plays an important role in the discussion of gender differences in academia. Academic progression and research in large are connected tightly to peer reviewing and even though peer reviewing promises to adhere only to academic quality (Roberts & Shambrook, 2012, p. 33), several biases in peer review have been identified in the literature but have also been equally questioned (for an overview, see Lee et al., 2013).

Many of these biases are relevant for gender. For example, Roberts and Shambrook (2012, p. 34) state that peer review is often seen as controlled by "elitists" or "gatekeepers" whose influence can be deemed as arrogant power mongering. According to the EU Commission's report "She Figures 2015", women publish fewer papers as corresponding authors (but in journals of similar prestige) than men and the gender gap in the funding success rate is decreasing but women's success rates are still lower than men's. Furthermore, the percentage of publications with a gender dimension remains low (with the highest score being 6.2% in the social sciences, see European Commission, 2016, p. 149). Both issues are often seen to be related to journal editorial policies and gender bias during the review process (GENDERACTION, 2019, p. 20). Helmer et al. suggest that women are underrepresented in the peer review process, and that editors of both genders operate with substantial same-gender preference (homophily) when appointing reviewers (Helmer et al., 2017, p. 1). Consequently, according to Budden et al., a double-blind peer review process can significantly increase the publication of female first-authored papers. The authors therefore suggest that this practice should be introduced widely (Budden et al, 2008, p. 4). This is considered important because research grants are also decided using peer review procedures and receiving grants or not can decide upon careers. Some studies show that men have on average statistically significantly greater odds of approval than women applying for grants (Bornmann et al., 2007, p. 234; Wennerås & Wold, 1997). Other studies suggest a gender bias in academic recruitment (van den Besselaar & Sandström, 2015; van den Brink et al., 2006). However, as the overview on the empirical literature on bias in peer review by Lee et al. (2013) shows, there is quite some counter evidence. Following the findings by Wennerås and Wold (1997), many studies could not replicate a gender bias in grant rewards (Bornmann & Daniel, 2006; Friesen, 1998; Mutz et al., 2012). The replication at the same institution as the original study found even a gender bias in the other direction (Sandström & Hällsten, 2008). Borsuk et al. (2009) showed in their experimental study that changing the gender of authors does not influence the judgement by peers. Also, they did not find support

for the gender-homophily thesis but rather that female post-docs are the most critical reviewers. Similarly, Husu and Cheveigné (2010), when discussing gatekeeping of excellence in research funding, state that increasing the proportion of women among gatekeepers of research funding does not necessarily or automatically lead to higher success rates for women applicants, nevertheless the a more equal representation among gatekeepers on women's participation in research may have a more indirect positive impact, not least by providing opportunities for women to become integrated in important networks (Husu & Cheveigné, 2010, p. 43). Van den Besselaar and Sandström (2015) did not find a difference in citation impact between papers by women and men but found a difference in performance, i.e. the quantity of papers produced.

Lee et al. (2013, p. 8) conclude that there is not much evidence for gender bias in peer review, however, other biases might apply. Meta-analyses show that controlled for different factors, such as discipline, seniority, reputation of the institution, no gender bias persists (Marsh et al., 2009; Mutz et al., 2012). Similarly, controlling for research stage, Ley and Hamilton (2008) do not find a gender difference in funding success rates. At the same time, it is still obvious that women are underrepresented at almost all stages of academic careers, the higher the scarcer women are. This points to the conclusion that while there is no evidence for a direct discrimination of women in peer review, there seems to be an interaction between different biases: Less money is granted to SSH disciplines, researchers affiliated to high reputation institutions or such with higher ranks receive better evaluations in peer review. At the same time, women are more likely to be active in SSH disciplines, are more likely to be at lower levels in the academic career and work at institutions with lower reputation. Ceci and Williams (2011) therefore argue women are not being denied grants and journal publications because of their sex, but rather due primarily to factors surrounding family formation and childbearing, gendered expectations, lifestyle choices, and career preferences. These factors might well reflect self-selection but also discrimination – yet not strictly within the academic realm but rather within society at large. Hence, they argue for shifting the focus from investigating bias in selection processes to studying social processes that pull women into inferior positions. This is even more important as the type of studies presented above has several theoretical and methodological shortcomings. Here, we point to four shortcomings: first, the studies start with the assumption that disagreement between peers is normatively not desirable (as notes Lee et al., 2013), second, most of them do only look at funding or publishing rates but not at performance (as criticises van den Besselaar & Sandström, 2015), third – and related to that –, outcomes of non-funded or non-published research cannot be compared to funded or published research (as pointed out by Mutz et al. 2015), fourth, the focus is on the reviewers, editors or funders but not on the researchers (as noted by Rowley & Scaffi, 2017). Lee et al. (2013) argue that while impartiality of peer review is seen as important in ensuring both consistency and meritocracy in the evaluation process, such expectations on peer review might be questionable and ask whether impartiality should be upheld as an ideal for peer review altogether, as peer reviewing, editorship and evaluation serve as a social function in negotiating and improving academic quality (Lee et al., 2013, p. 13). This, then, asks for further investigations in how peer review is seen by scholars and how they react on and interact with peer review and whether there are gender differences in the perceptions of and reactions to being reviewed.

Rowley and Sbaffi (2018) report on survey research on scholars' attitudes towards peer review. They find that, in general, gender was not seen as a source of bias, but region and seniority were seen as potential sources of bias (Rowley & Sbaffi, 2018, p. 652). However, their methodology is highly questionable ("The confidence interval (at a 95% confidence level) for any one question is 1.18", Rowley and Sbaffi, 2018, p. 647; note that means and standard deviations, and consequently standard errors, were widely different across variables). What has rarely been studied before is the scholars' opinion on peer review from a gender perspective, which could be the key in resolving female academics' lower scientific production and could lead towards a better inclusion. Of high interest in this context is how young scholars perceive the process as, first, bias is shown to be small at the early career stage and it is at this stage when scholars will decide on whether they want to stay or leave academia, both of which is important regarding later gender bias (van den Besselaar & Sandström, 2015). Thus, in the following, we will present first results of a Europe-wide project on Early Career Investigator's experiences with peer review, at the beginning of their career, struggling to enter to academia and secure a place in a very competitive system. On this path they are encountering the peer review process, in various forms, but mostly as the ones who are submitting papers in peer-reviewed journals or applying for various grants and being exposed to potential biases in this process. The following presentation of the findings is to be considered as exploratory as we hope to raise interest towards this topic for further research investigations.

Methods and Sample

In March 2017, ENRESSH's Special Interest Group on Early Career Investigators (SIG ECI¹⁰), agreed to conduct a qualitative research in seventeen European countries, namely Belgium, Bosnia & Herzegovina, Croatia, Cyprus, Finland, France, Latvia, Lithuania, Montenegro, the Netherlands, Poland, Portugal, Serbia, Slovakia, Slovenia, Switzerland and Malta. In each country, semi-structured interviews were conducted. The interview grid was developed by the SIG members, with the aim of analysing the narratives of young researchers about their early careers, more specifically, how certain "moments of evaluation" (such as doctoral exams, dissertation defence, post-doc recruitment, academic job application processes, research activities and peer review) played a role in their career development. Each of the interviewers conducted up to four interviews from his or her home country. The interviews were conducted in native and/or state languages of the interviewees to avoid selection bias through language and they were later translated into English language.

For this short contribution, 48 interviews were analysed. There were 28 female respondents (58%) and 20 male respondents (42%). Not all interviews contained data on peer review, as the interviewers let the respondents talk freely, therefore, some of the questions were unintentionally skipped, respectively were not important enough for the respondent to

¹⁰ George Afxentiou, José Gabriel Andrade, Katya De Giovanni, Stefan De Jong, Gemma Derrick, Rita Faria, Haris Gekic, Bradley Good, Emanuel Kulczycki, Karolina Lendák-Kabók, Stephanie Mignot, Reetta Muhonen, Filippo Nereo, Martin Nuhliček, Michael Ochsner, Antun Plenković, Stevo Popovic, Jolanta Sinkuniene, Julia Stukalina, Mimi Urbanc and Marc Vanholsbeek.

emerge during the interview, or simply the interviewers were more focused on other questions from the interview grid and skipped the question about peer review.

Preliminary findings

In the following, we present preliminary results from analysing the interviews regarding gender and geopolitical differences in the perception of peer review. These are two important topics as gender bias in peer review is a strongly debated issue, while geopolitical differences are a confirmed bias of peer review (Lee et al., 2013, pp. 6–8). Both can be studied with our sample as there is sufficient variance between gender and geopolitical location, specifically Eastern and Western Europe.

The interviewees see peer review predominantly as a suitable way of advancing in the scientific career and improving research or publications. Yet, both male and female respondents mentioned several negative aspects. Given the vivid scholarly discussion on bias in peer review, we will start with presenting the negative aspects and end with the positive aspects.

Negative aspects of peer review

First, we present a general criticism from a geopolitical context towards peer review, which was mentioned by both men and women from the Eastern European countries.¹¹ Respondents in Eastern European countries report two different, if not conflictive, perspectives on geopolitical biases: On the one hand, the respondents (both male and female) were complaining about nepotism and local networks which are dominant in either small countries (like Slovenia), where the scholars are mainly familiar with each other's work and peer review practices, or in Bosnia and Herzegovina and Croatia where they are criticising the national system and emphasizing that there is a need for “Western system”, which is presumably more fair and objective. On the other hand, some respondents criticised the international peer review system or editorial practices as well, which they suspect being biased against Eastern European researchers.

In Western European countries, the discrepancy between local and international peer review focused on topics or language rather than evaluation bias. The hurdle between passing local peer review vs international peer review was attributed to differences in relevance of topics at the local and international level, leading to the problem of career advancement if one focuses on local relevance or if one publishes in local languages rather than in English. Contrary to the respondents in Eastern Europe, the respondents in Western Europe interpret this as a perverse effect of science policy favouring international research in English rather than local research, while Eastern Europeans feel a discrimination even if they publish in English on internationally relevant topics.

Regarding gender-specific views on peer review, we will start with the male respondents who are more critical towards peer review than female respondents are. On the one hand,

¹¹ Eastern and Central European countries are the ones geographically and geopolitically considered as Eastern Europe; more specifically, they are the ones who accessed the EU with the 2004 enlargement or later (Croatia, Latvia, Lithuania, Poland, Slovakia, and Slovenia) and the ones who are considered to be part of Western Balkans and are not EU members, i.e. Bosnia & Herzegovina, Montenegro, and Serbia. Western European countries are the ones geographically and geopolitically considered as Western Europe (Switzerland) and/or as “the old member states” of the EU, i.e. Belgium, Finland, France, the Netherlands, Portugal.

male respondents believe that peer review can be biased and some of them have a strongly negative viewpoint, adding that the expertise of the peer reviewers is questionable, even to the point that they are not accepting the reviewer's comments. Men are emphasizing that good relations are needed to get the paper reviewed, concluding that this form of evaluation may even damage the paper's quality. Among the arguments against peer review, other male respondents point out that network is of great importance, as well as that the quality of reviewing is very variable and depends on the journal's editor. Finally, according to some of the respondents, the process slows down the development of new ideas. It is noteworthy to mention that all of the male respondents already had some experience with peer review. In a general sense, male respondents were much more critical towards peer review, from the ones who answered the question about peer review five reported more positive aspects of peer review and thirteen more negative aspects about it.

Female respondents share the opinion with their male counterparts, that peer review is often influenced by interpersonal relationships (between senior researchers). They agree with their male peers that the comments are sometimes irrelevant and misleading. Female respondents question the slowness and the long process of peer review, that requires a good command of English. The main gender difference, however, lies in the effect reviews have on the ECIs. Peer review seems to affect the confidence of young female researchers negatively, while this was not the case of male interviewees. For instance, some of them were very surprised to get criticised for something she invested a lot of effort in and some of the female respondents expressed their disappointment with the process. The second gender difference concerns a more positive stance toward peer review, i.e. eleven female respondents had a generally positive attitude towards peer review whereas twelve mentioned more negative aspects. Furthermore, some female respondents never had any experience with peer-review, which is important to note, as peer review gives credibility to scientific production.

Positive aspects of peer review

Both male and female respondents reported also positive experiences with peer review. As stated by some of the respondents, they perceive peer review, all things considered, still better than any other evaluation procedure.

Despite being conscious about the reported negative aspects, many male respondents point out that the process is of great importance for the improvement of a paper, and that one can learn a lot along the way. The process is very often fair and makes young researchers progress faster by having constructive comments to rely on.

Similarly, female ECIs point out that peer review is not always perfect, but very often fair, useful and reasonable. They often find the reviewer's comments very useful. In order to make the process more objective, one of the respondents suggested an internationalisation of the peer review process, pointing out that collegial peer review should be replaced by fairer peer review.

Conclusions

Peer review seems to be an important moment of evaluation for both male and female respondents. The majority of them believes that this sort of evaluation is the most appropriate way to improve the quality of the paper or to evaluate grant applications. However, both men and women perceive the process as dependent on interpersonal relations in some

contexts, dominated by local networks, sometimes even biased and unfair, and slowing down the process of publication. There was a clear geopolitical difference when both men and women from Eastern Europe were criticising national peer review and idealising the “Western style” of peer-review. At the same time, they were mentioning that the level of English was acting as a discriminatory factor, while scholars from Western non-English speaking countries mentioned language bias not as a discriminatory factor but rather an effect of science policy favouring topics relevant in the Anglo-Saxon context. It is interesting to point out, that neither male nor female respondents mentioned a gender bias in peer review, they were solely concentrating on perceptions of and responses to the peer review process.

While the ECI scholars did not mention gender aspects of peer review, their reflections and reactions to peer review nevertheless revealed important gender differences in how peer review influences them and their research practice. Female respondents seem to have less experience with the process whereas all the male interviewees had previous experience with it. Moreover, female scholars are affected to a greater extent by the comments than male researchers, who show more confidence and sometimes disagreement with the reviewers. Male ECIs seem to develop their standpoints and build their own profile in the sometimes conflict-ridden interaction with reviews, whereas female ECIs seem to try to follow the suggestions of the reviewers and even question themselves if they disagree.

We thus find empirical evidence for the hypotheses formulated by Ceci and Williams (2011) arguing that the reasons for gender differences in academic publication and positions might lie outside the peer review process. However, our results suggest that they do indeed not necessarily lie in discriminatory practices by reviewers but nevertheless are not completely outside of the peer review process: rather, the way female researchers react to – and maybe interact with – reviews or moments of evaluation might lead to different success rates in academic careers. Furthermore, discrimination appears on the level of topics as well, be it through gender or region-specific ways of approaching research questions. This might lead to seemingly self-discriminatory behaviour, but the roots lie deeper in society, for example in socialisation, gender norms or geopolitical hierarchy. For example, measures for helping women reconcile family and work in academia might not help increasing the share of women in higher positions as it might have adverse effects: If it is easier for women to reconcile, it is the women taking responsibility for caring for the children as men do not have the same options. At the same time, a notion of “quota female professor” can emerge if policies are favouring women. Besides helping women entering the work sphere, it seems to be important to support (or push) men to take more responsibility at home. Besides the conclusions in line with Ceci and Williams (2011), our results suggest, however, that it is also important to not mask gendered or localised ways of prioritising topics or reactions to evaluations behind general societal developments. Such gender and geopolitical issues need to be addressed in evaluation practices and in research on evaluation.

In sum, our preliminary results report on geopolitical differences in peer review and a gendered perception of and response to peer review. We therefore argue that instead of focusing almost exclusively on analysing impartiality of peer review that comes with methodological but also conceptual problems, as shown in the introduction (see also Lee et al., 2013), and besides monitoring relative representation of women or other discriminated groups in certain positions, more research should investigate how researchers at different

stages of their career react to selection mechanisms and how peer review can be enhanced to improve research and to help building an academic identity and academic careers. On the one hand, academics should be better trained to provide useful reviews and to interpret reviews in a constructive way. On the other hand, one should de-mystify peer review as an impartial, objective quality assessment and rather acknowledge its gatekeeping and social functions that need to be actively and critically negotiated between different actors in academia. Furthermore, research should investigate peer review of interdisciplinary research where disciplinary differences between research and reviewing practices can lead to similar ways of (self-)discrimination, for example of SSH researchers. Awareness of the social functions of peer review and their active negotiation will be an important issue to be addressed in the future for a better inclusion of Eastern European and female scholars of different disciplines, and especially the combination of the three, in the European Research Area because also forms of self-discrimination and discriminations outside evaluation practices can lead to inefficient selection processes.

References

- Bornmann, L., & Daniel, H.-D. (2006). Potential sources of bias in research fellowship assessments. Effects of university prestige and field of study on approval and rejection of fellowship applications. *Research Evaluation*, 15(3), 209–219. <https://doi.org/10.3152/147154406781775850>
- Bornmann, L., Mutz, R., & Daniel, H.-D., (2007). Gender differences in grant review: A meta-analysis. *Journal of Informetrics*, 1(3), 226–238. <https://doi.org/10.1016/j.joi.2007.03.001>
- Borsuk, R. M., Aarssen, L. W., Budden, A. E., Koricheva, J., Leimu, R., Tregenza, T., & Lortie, C. J. (2009). To name or not to name: The effect of changing author gender on peer review. *BioScience*, 59(11), 985–989. <https://doi.org/10.1525/bio.2009.59.11.10>
- Budden, A.E., Tregenza, T., Aarssen, L. W., Koricheva, J., Leimu, R., & Lortie, C. J., (2008). Double-blind review favours increased representation of female authors. *Trends in Ecology and Evolution*, 23(1), 4-6. <https://doi.org/10.1016/j.tree.2007.07.008>
- Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, 108(8), 3157–3162. <http://doi.org/10.1073/pnas.1014871108>
- Dubois-Shaik, F., & Fusulier, B. (2017). Understanding gender inequality and the role of the work/family interface in contemporary academia: An introduction. *European Educational Research Journal*, 16(2–3), 99–105. <https://doi.org/10.1177/1474904117701143>
- European Commission. (2016). *She Figures 2015. Gender in Research and Innovation*. European Union. https://ec.europa.eu/research/swafs/pdf/pub_gender_equality/she_figures_2015-final.pdf

- Friesen, H. G. (1998). Equal opportunities in Canada. *Nature*, 391, 326. <https://doi.org/10.1038/34768>
- GENDERACTION. (2019). *Report on strategic advice for enhancing the gender dimension of open science and innovation policy*. Institute of Sociology of the Czech Academy of Sciences. https://genderaction.eu/wp-content/uploads/2019/04/GENDERACTION_Report-5.1_D11_OS0I.pdf
- Helmer, M., Schottdorf, M., Neef, A., & Battaglia, D. (2017). Gender bias in scholarly peer Review. *eLife*, 6, e21718. <https://doi.org/10.7554/eLife.21718>
- Husu, L., & de Cheveigné, S. (2010). Gender and gatekeeping of excellence in research funding: European perspectives. In B. Riegraf, B. Aulenbacher, E. Kirsch-Auwärter & U. Müller (Eds), *Gender change in academia* (pp. 43–59). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92501-1_4
- Kretschner, H., Pudovkin, A., & Stegmann J. (2012). Research evaluation. Part II: Gender effects of evaluation: Are men more productive and more cited than women? *Scientometrics*, 93, 17–30. <https://doi.org/10.1007/s11192-012-0658-0>
- Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1), 2–17. <http://doi.org/10.1002/asi.22784>
- Ley, T. J., & Hamilton, B. H. (2008). The gender gap in NIH grant applications. *Science*, 322(5907), 1472–1474. <https://doi.org/10.1126/science.1165878>
- Morley, L. (1999). *Organizing feminisms: The micropolitics of the academy*. New York: Palgrave Macmillan. <https://doi.org/10.1057/9780333984239>
- Marsh, H. W., Bornmann, L., Mutz, R., Daniel, H.-D., & O'Mara, A. (2009). Gender effects in the peer reviews of grant proposals: A comprehensive meta-analysis comparing traditional and multilevel approaches. *Review of Educational Research*, 79(3), 1290–1326. <https://doi.org/10.3102%2F0034654309334143>
- Mutz, R., Bornmann, L., & Daniel, H.-D. (2012). Does gender matter in grant peer review? *Zeitschrift für Psychologie*, 220(2), 121–129. <https://doi.org/10.1027/2151-2604/a000103>
- Mutz, R., Bornmann, L., & Daniel, H.-D. (2015). Testing for the fairness and predictive validity of research funding decisions: A multilevel multiple imputation for missing data approach using ex-ante and ex-post peer evaluation data from the Austrian Science Fund. *Journal of the Association for Information Science and Technology*, 66(11), 2321–2339. <https://doi.org/10.1002/asi.23315>
- Nielsen, M. W. (2015). Gender inequality and research performance: moving beyond individual-meritocratic explanations of academic advancement. *Studies in Higher Education*, 41(11), 2044–2060. <https://doi.org/10.1080/03075079.2015.1007945>
- Roberts, T. J., & Shambrook, J. (2012). Academic excellence: A commentary and reflections on the inherent value of peer review. *Journal of Research Administration*, 43(1), 33–38.

- Rowley, J., & Sbaffi, J., (2018). Academics' attitudes towards peer review in scholarly journals and the effect of role and discipline. *Journal of Information Science*, 44(5), 644–657. <http://doi.org/10.1177/0165551517740821>
- Sandström, U., & Hallsten, M. (2008). Persistent nepotism in peer-review. *Scientometrics*, 74(2), 175–189. <https://doi.org/10.1007/s11192-008-0211-3>
- van den Besselaar, P., & Sandström, U. (2015). Gender differences in research performance and its impact on careers: a longitudinal case study. *Scientometrics*, 106(1), 143–162. <http://doi.org/10.1007/s11192-015-1775-3>
- van den Brink, M., Brouns, M., & Waslander, S. (2006). Does excellence have a gender? A national research study on recruitment and gender. *Employee Relations*, 28(6), 523–539. <https://doi.org/10.1108/01425450610704470>
- Waaijer, C. J., Sonneveld, H., Buitendijk, S. E., van Bochove, C. A., & van der Weijden, I. C. (2016). The role of gender in the employment, career perception and research performance of recent PhD graduates from Dutch universities. *PloS ONE*, 11(10), e0164784. <https://doi.org/10.1371/journal.pone.0164784>
- Wennerås, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, 387, 341–343. <https://doi.org/10.1038/387341a0>

Peer review in the context of the new modes of knowledge production, dissemination and evaluation

By Marc Vanholsbeeck

Introduction

There has been a tendency since the 1980s to open the knowledge production, dissemination and evaluation to stakeholders outside of the circles of the disciplinary peers who traditionally produce fundamental research. Such ‘exoterization’ of research – in its etymologic meaning of going out of the more ‘esoteric’ communities of science – goes far beyond the linearity of knowledge transfer and marketization of research outputs, since it includes social innovation – which can take non-commercial forms – and relates to the notion of a knowledge society rather than to the narrower concept of economy of knowledge (Vanholsbeeck, 2017).

In some cases, such as in recently institutionalized disciplines in social sciences and humanities (SSH) like communication studies, moving beyond disciplinary frontiers has been the consequence of the need of educating future professionals in emerging fields. In other cases, such as in the development of gender studies in the 1970s, interdisciplinarity has been stimulated by the raise of the feminism movement, with societal claims that went beyond the confines of any discipline. More generally though, the opening of research has been supported by policy makers, desiring to make research more responsive to real world problems, rather than focusing on problematics enunciated in strictly disciplinary terms.

At European level, the European Research Area (ERA) – which aims at providing free circulation for researchers, scientific knowledge and technology (article 179 of the Treaty on the Functioning of the European Union, 2007) – resulted in the promotion of programmatic ideas through which policy makers support the exoterization of research. Worth mentioning here are the European “strategic research” of the 1980s and the more recent support of EU policy makers for the “co-creation” of solutions to societal challenges by researchers from diverse disciplines, together with stakeholders from the industry and/or citizens. Policy makers have indeed been prompt to endorse concepts directly or indirectly based on the notion of co-creation, such as the “Mode 2 of Knowledge Production” (Gibbons et al., 1994) – according to which multidisciplinary teams work together for short periods of time on real world problems –, the “Responsible Research and Innovation” (European Commission, 2013), the notion of “societal challenges” which constitutes one of the three pillars of the current Horizon 2020 Framework Programme (2014-2020) as well as, to a lesser extent, “citizen science” (Irwin, 1995). In the context of the preparation of the next European research and innovation framework programme “Horizon Europe” (2021-2027), a key notion is that of “missions” focusing on problem-specific societal challenges and the interaction of several public and private actors to solve them (Kattel & Mazzucato, 2018). The European Open Science agenda¹² – including open access to publications, open research data and citizen science – and the impact related policies – the so-called “impact agenda” of the European Commission – also align with and foster this tendency to exoterization.

¹² https://ec.europa.eu/research/openscience/pdf/draft_european_open_science_agenda.pdf#view=fit&page-mode=none

SSH integration to European impact-driven interdisciplinary research

In this perspective, SSH research has been under pressure to “integrate” – or to be “embedded” into – European funded research and to contribute to the resolution of societal challenges. According to a recent monitoring report from the Commission (European Commission, 2018), it still appears that current SSH integration in Horizon 2020 is not satisfactory. In 2016, 70 out of 239 projects funded under the SSH flagged topics had no SSH partners (29%), while some disciplines are practically not involved such as history (2%) and anthropology/ethnology (1%). The quality of SSH integration is also highly uneven across Horizon 2020.

Furthermore, it has been noted that Evaluation Summary Reports (ESR) in Horizon 2020 often suffer from a lack of SSH expertise (Universiteit Gent, 2018). This not only constitutes an obstacle towards any true SSH integration in framework programmes, but more generally it highlights a new challenge for peer reviewing. There is indeed a need for funders to engage SSH researchers from diverse disciplinary backgrounds into peer reviewing of impact driven research projects. This also supposes that researchers should be provided with the right skills, such as learning to debate and argue with peers from other disciplinary horizons and, in some cases, with peers without knowledge of SSH specific epistemologies and methods. They should also be properly trained to assess impact.

While peer review of research projects that go beyond the frontiers of traditional or disciplinary knowledge is challenging per se (Lamont, 2009; Luukkonen, 2012), a particular issue for SSH scholars engaged in interdisciplinary panels together with STEM (science, technology, engineering and mathematics) colleagues is that the latter are not always ready to allow them the same level of legitimacy of expertise. SSH peer reviewers may thus face a lack of “legitimate locus of interpretation” (Collins & Evans, 2007), from which to exercise their expertise. The notion of legitimate locus of interpretation relates to the (social) “location”, in terms of communities and expertise, from which legitimate knowledge claims and judgements of those knowledge claims can be made. Thus, in interdisciplinary panels, SSH peer reviewers may soon be faced with the diverse levels of prestige that are attached to the various disciplines inside academia rather than with the equalitarian perspective that policy makers’ discourses on interdisciplinary partnerships most often take for granted. This issue about legitimate interpretation concerns also SSH scholars’ relations with non-academic actors potentially engaged in the review. Indeed, as Lewis (2018) argues, “within the physical, chemical and biological sciences, the legitimate locus of interpretation usually lies well inside the community of producers, as only those with specialist expertise are deemed sufficiently equipped to make valid judgements. By contrast, the locus of legitimate interpretation in the social sciences is much more diffuse.”

Finally, policy makers and funders should avoid referring to the notions of “SSH expert” and “SSH expertise” in the context of interdisciplinary evaluation. Such concepts indeed contribute to the reification of the SSH as a coherent epistemological and methodological entity, while epistemologies and methods vary a lot from one discipline to the other. Hence SSH disciplinary expertise engaged in interdisciplinary peer reviewing should rather reflect as far as possible the diversity of the individual SSH disciplines relevant to the project to assess. If it is not feasible to engage reviewers from the disciplines that are concerned, then experts from adjacent disciplines should be privileged. Pools of experts in interdisciplinary

evaluation could also be created, with indication of their thematic rather than strictly disciplinary fields of competence.

Perception of senior sociologists regarding peer reviewing

While the context into which research is produced, disseminated and assessed is subject to the abovementioned changes and spurs peer reviewers to adapt to more interdisciplinary and impact driven research, literature shows that researchers at all stages of the academic career (including early career investigators) are generally satisfied with the traditional peer review system, which they find mostly valuable (Johnson et al., 2018; Mulligan et al., 2013; Nicholas et al., 2015; Publons 2018; Sense About Science, 2009; Thelwall et al., 2013; Ware, 2008; Ware and Monkman 2008). According to a survey reported by Ware (2008), a large majority of scholars (93 per cent of the respondents) even consider peer reviewing as necessary. Another large-scale survey (Nicholas et al., 2015) indicates that peer review is still considered as the most important indicator for assessing the trustiness of a publication and has an important role in improving the quality of articles. According to the same study, researchers are not likely to cite non-peer-reviewed material although some – in particular younger ones – are willing to use it. Peer review and Impact Factor also remain defining criteria in the choice of a journal where to publish. Nevertheless, researchers also recognize flaws in peer reviewing relating to unfairness, bias and delays, and have concerns in regard to the burdens put on reviewers and the general effectiveness of the process (Johnson et al., 2018).

In the context of a subgroup of the WG1¹³, a study of senior SSH researchers' attitudes towards research evaluation has been carried out. The subgroup members conducted in 2018 and 2019 semi-structured interviews with 16 sociologists, male and female, having earned their PhD for at least eight years and active in eight European countries (Belgium, Croatia, Cyprus, Finland, Iceland, Lithuania, Poland and Slovenia). They interviewed them in native and/or state languages about their perceived roles in the definition, dissemination and implementation of the quality criteria and rationales used in evaluation situations. In this report, only the respondents' perception of peer review will be considered.

Respondents' perception of peer review relates firstly to the changes they see in research evaluation which, according to them, gives now privilege to international publications and the use of quantitative indicators of performance. Hence some interviewees consider that quantitative elements in evaluation have become more important than qualitative peer reviewing, while others deem that, in their experience, performance indicators have rather been integrated into a more informed and hybrid peer reviewing process.

In any case, peer review is mostly perceived as a very important – or even essential – system of quality control in the production of sociological knowledge. Still the attitudes of the respondents remain ambivalent: even if they perceive clear assets linked to the prepublication traditional peer reviewing system, some researchers also consider the downsides of peer review, or even some threats that may be linked to it.

On the one hand, peer review is considered by many as a top-quality criterion for selecting scholarly journals. It would make a clear-cut distinction between quality publications and

¹³ Theodosia Demetriou, Agne Girkontaite, Andreja Istenic Starcic, Ville Keiski, Emanuel Kulczycki, Elena Papanastasiou, Janne Polonen, Hulda Proppe, Marc Vanholsbeeck and Maja Vehovec

– in the words of one male Icelandic sociologist – the other “small things” of lesser value that a sociologist produces. It also leads several respondents to distinguish between academics who are skilled in publishing in international peer-reviewed journals and the less educated others. Essential skills relate here to (English) language, but also to how to manage peer review, as a reviewer and as a reviewed author.

Peer review is deemed as being particularly well adapted to individual level evaluation. In case of article peer reviewing, it fulfils important epistemic functions, helping the authors to improve their paper, and the reviewers to be updated on the last developments in certain fields. In small countries, involving international peer reviewers is considered as helping to overcome the problems inherent to small academic communities in which “everybody knows everybody” (Slovenian female sociologist). When international peer reviewers are involved by national funders, it can also encourage the internationalization of the SSH research national production as well: “If you want to be funded you should publish internationally and this makes sense, actually a lot of sense to start publishing internationally and to increase chances to be funded” (Croatian male sociologist).

On the other hand, peer review is far from being perceived as a flawless process and appears to several respondents as potentially biased. Indeed, some interviewees mention that peer reviewers may be unfair, or not willing to provide a balanced review. One respondent, an Icelandic female scholar, even considers that peer reviewers in the social sciences are particularly harsh and critical, in comparison to other disciplines: “Maybe in particular researchers within the social sciences who are educated in critical thought... to have a critical view... that there is something wrong with us if we don’t find anything, to take proposers down, that is.” There is also a perceived lack of formalized guidelines on how to conduct peer review. An overreliance on international peer review processes may have as an unwanted consequence that local contexts of application and societally relevant impact are not taken enough into account by international peer reviewers.

A Belgian male sociologist goes as far as linking peer review to a form of censorship to which one complies because one needs to, but which one would rather bypass otherwise: “We sometimes have the impression that when we write for a scholarly journal, there is still some form of censorship somewhere, i.e. you get published only if you are in line with the journal. I do not like to rewrite something where I think I’m right, because someone tells me, unless I obviously agree with it.”

According to some respondents, “predatory” open access journals that ask article processing charges for publishing papers have also cut back on the scholars’ confidence in peer reviewing, because of the low level of peer reviewing such journals maintain (if any). In some cases, predatory journals manage to attract researchers though, because of the pressure to publish internationally.

Finally, some respondents consider that attention to the peer review status of a publication has to be balanced with the local impact that some non-peer-reviewed types of publication may have, such as national journals or publications targeting non-academic audiences: “For me it is also important to reach out to the wider community by publishing in maybe perhaps lower impact factor journal papers that would reach a lot more people through more publicly oriented journals that would be read by actually the practitioners of whatever research output you produced from your own research.” (English language female professor from Cyprus)

Discussion

The interviews we have conducted in the context of a subgroup of WG1 and the balanced interviewees' perceptions of peer review remind us that, beyond the challenges discussed above in relation to the needed adaptation of peer review to more interdisciplinary and impact driven research, there is a distinction to be made between peer review as a scholarly process aimed at vetting research results and peer review as a privileged indicator of the highest level of achievement in research production. While peer review definitely matters in SSH, considering it – in its current stage – as the exclusive label of scientific quality and relevance, like in the STEM, may lead us to ignore a large and significant amount of SSH research outputs. In the SSH disciplines, there is indeed a wider diversity of valuable types of research publications that goes beyond the article published in (international) peer-reviewed journals, like for example books or non-scholarly press (Hicks, 2004).

Furthermore, on an epistemological level, SSH knowledge progresses mostly through a heterogeneity of simultaneous quasi-paradigms, rather than as a succession of consensual paradigms on which most scientists would agree (Bonaccorsi, 2018). It has also been shown that the peer reviewers' definition of originality is much broader in the SSH than in the STEM, taking into account the new character of findings, theories, approaches and methods as well (Guetzkow et al., 2004). This should on the one hand be duly considered while integrating the perspective of SSH research into the development of peer review processes that are better adapted to interdisciplinary and impact driven research. On the other hand, it should also further discourage policy makers and evaluators to focus solely on peer review, taken as the defining label of excellence, while discussing quality issues within the SSH.

References

- Bonaccorsi, A. (2018). Towards an epistemic approach to evaluation in SSH. In A. Bonaccorsi (Ed.), *The evaluation of research in social sciences and humanities* (pp. 1–29). Springer. https://doi.org/10.1007/978-3-319-68554-0_1
- Collins, H.M., & Evans, R. (2007). *Rethinking Expertise*. University of Chicago Press.
- European Commission. (2013). *Options for strengthening Responsible Research and Innovation. Report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation*. European Commission. <https://doi.org/10.2777/46253>
- European Commission. (2018). *Integration of social sciences and humanities in Horizon 2020: Participants, budget and disciplines. 3rd monitoring report on SSH flagged projects funded in 2016 under the Societal Challenges and Industrial Leadership priorities*. Publications Office of the European Union. <https://doi.org/10.2777/756427>
- 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.

- Guetzkow, J., Lamont, M., & Mallard, G., (2004). What is originality in the humanities and the social sciences? *American Sociological Review*, 69(2), 190–212. <https://doi.org/10.1177%2F000312240406900203>
- Hicks, D. (2004). The four literatures of social science. H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 473–496). Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-2755-9_22
- Irwin, A. (1995). *Citizen science: A study of people, expertise and sustainable development*. Routledge.
- Johnson, R., Watkinson, A., & Mabe, M. (2018). The 5th edition of *The STM Report: An overview of scientific and scholarly publishing*. https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf
- Kattel, R., & Mazzucato, M. (2018). Mission-oriented innovation policy and dynamic capabilities in the public sector. *Industrial and Corporate Change*, 27(5), 787–801. <https://doi.org/10.1093/icc/dty032>
- Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Harvard University Press
- Lewis, J. (2018). The locus of legitimate interpretation in social science and what it means for public understanding. *Communication at Science in Public 2018 Intersecting Science 2018 Conference, Cardiff, 17–19 December 2018*. https://docs.wix-static.com/ugd/48bd50_66b801501d204ed3a123b84b947ac823.pdf
- Luukkonen, T. (2012). Conservatism and risk-taking in peer review: Emerging ERC practices. *Research Evaluation*, 21(1), 48–60. <https://doi.org/10.1093/reseval/rvs001>
- Mulligan, A., Hall, L., & Raphael, E. (2013). Peer review in a changing world: an international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology*, 64(1), 132–161. <https://doi.org/10.1002/asi.22798>
- Nicholas, D., Watkinson, A., Jamali, H.R., Herman, E., Tenopir, C., Volentine, R., Allard, S., & Levine, K. (2015). Peer review: still king in the digital age. *Learned Publishing*, 28(1), 15–21. <https://doi.org/10.1087/20150104>
- Publons. (2018). *Global State of Peer Review*. <https://publons.com/community/gspr>
- Sense About Science (2009). *Peer review survey 2009: Preliminary findings*. Sense About Science. <https://senseaboutscience.org/activities/peer-review-survey-2009/>
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS ONE*, 8(5), e64841. <https://doi.org/10.1371/journal.pone.0064841>
- Universiteit Gent. (2018). *Position Paper IE H2020. The 'SSH embedding Challenge'*. https://www.ugent.be/en/research/position-papers/ie-position-paper.pdf/at_download/file

- Vanholsbeeck, M. (2017). La notion de Science Ouverte dans l’Espace européen de la recherche: Entre tendances à l’«exotérisation» et à la «gestionnarisation» de la recherche scientifique. *Revue française des sciences de l’information et de la communication*, 11. <https://doi.org/10.4000/rfsic.3241>
- Ware, M., & Monkman, M. (2008). *Peer review in scholarly journals: Perspective of the scholarly community—an international study*. Publishing Research Consortium. <http://www.publishingresearch.net/PeerReview.htm>

Peer review in the context of Open Science: towards more open peer reviewing practices?

By Marc Vanholsbeeck

Introduction

The notion of “boundary object”, as coined by Star and Griesemer (1989), refers to concepts which are defined in a way that is flexible enough to be shared between diverse communities of practice, but also have definitions within them that are more community specific. As such, “boundary objects permit working relationships between communities while allowing local use and development of the concept.” (Moore, 2017)

Moore (2017) considers that Open Access to publications (OA) can be considered as a boundary object. Indeed, if the general definition of OA is quite consistent across various communities of practices – framing it as free access to scholarly publications via Internet –, “publishing means different things to different communities and individual approaches to OA are representative of this fact.” We argue that not only OA but the whole concept of Open Science, including the notion of open peer reviewing (OPR), can be considered as a boundary object too. As such, different stakeholders engage and promote Open Science, while they still adopt different definitions of it, and emphasize some of its dimensions rather than others.

From a policy makers’ perspective, the Open Science agenda of the European Commission sits at the crossroads of the European policy makers’ support towards ‘exoterization’ of research – i.e. the political will to open research outside of the ‘esoteric’ circles of the disciplinary peers who traditionally produce fundamental science (Vanholsbeeck, 2017) – and the “digital agenda” for Europe (European Commission, 2010) which promotes digital tools and media to produce and disseminate research. Open Science policies have been put in place first in regard to open access to publications (via 7th framework programme mandates since 2008, and with the current EU-supported PlanS¹⁴) and then to Open Research Data (via Horizon 2020 mandates since 2014, and through the current European Open Science Cloud initiative¹⁵). OA also constitutes the 5th priority of the European Research Area (European Commission, 2012) and had been announced as becoming a default principle for 2020 (Council of the European Union, 2016; European Commission, 2018).

On their side, although not rejecting Open Science as such, SSH communities have sometimes expressed worries in regard to Open Access (Chartron, 2014) and framed Open Science in a less enthusiastic way. On the one hand, there are concerns that a flipping to article processing charges and Gold Open Access – i.e.: publication in OA journals, rather than secondary deposit on OA archives (Green Open Access) – may mean the end of small scale publishers, not able to digitize their journals, or scholarly societies. Because the OA business model originally focused on journal articles, issues have also been raised regarding the sustainability of book publishing in an Open Science context, books being more important in SSH than in Science, Technology, Engineering and Mathematics (STEM) (Crossick, 2016). On a more epistemological level, some scholars in the humanities

¹⁴ <https://www.coalition-s.org/>

¹⁵ <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

expressed worries that the use of too liberal (Creative Commons) licenses may allow undesirable modifications to be brought to their work and argumentations, while concerns have also been raised in regard to what constitutes (open) data in the SSH (Eve, 2014).

Open Peer Review

Open Peer Review (OPR) has been mostly promoted by Open Access publishers and innovators (Ross-Hellauer, 2017) and been made possible by the more open, collaborative and social communication provided by the Web technologies (Tennant et al., 2017). As a dimension of Open Science, OPR shares its abovementioned nature of boundary object and does not have any unique definition. While keeping the common idea of increasing the levels of transparency and accountability of the publishing process, OPR may refer to open identities (of reviewers and/or authors), open publication of review reports together with the articles, open preprints, open final-version commenting, open platforms (‘decoupled review’) and ‘portable peer review’ (reviewer reports of rejected papers being made available to the editors of other journals), with or without possibilities of open interactions and/or participation of readers (Ross-Hellauer, 2017; Tennant et al., 2017). In some cases – like in particular in the so-called OA ‘mega-journals’ –, only the technical or scientific soundness of the article is peer-reviewed before publication, the article being opened to open peer review once published (Spezi et al., 2018).

Some downsides of OPR have been mentioned in the literature. First, researchers still show a high level of trust in the traditional – most often single-blind – peer review (Johnson et al., 2018) and tensions may appear between more conservative research communities and individual actors in favour of a rapid transformation of the system (Tennant, 2018). There is a lack of career incentives to engage scholars into more innovative peer reviewing practices (Tennant et al., 2017), while some of those practices are particularly time-consuming. There are other concerns, notably that revealing the name of the reviewer would lead to less frank reviews and biases as well as fear of reprisals either by criticized authors or even readers, reinforcing power relationships between reviewers and authors as well as detrimental strategic attitudes (Groves & Khan, 2010; Teixeira da Silva, 2019). It has also been argued that soundness-only peer review lowers the acceptance thresholds because the business model behind it encourages editors to accept an ever-increasing number of papers (Buriak, 2015). Furthermore, findings have demonstrated that the pretention to produce objective soundness-only peer review is not always matched by the reality. Indeed, some publishers still require reviewers to assess the “worthiness” of the articles and reviewers have sometimes the tendency to write soundness-only reports in the same way as for journals with conventional peer review (Spezi et al., 2018).

On a technical level, OPR – in particular under the form of decoupled and portable peer review – is supposed to help to improve already published articles, but the most commonly used PDF format does not allow any alteration of the published material. There is also a general lack of version control of articles and no sufficient interoperability between decoupled review reports (Tennant, 2018). Eventually, mega journals do not have been a success in the SSH and tend from now on to concentrate on the STEM (Spezi et al., 2018).

Significant assets of OPR are to be considered too. OPR makes the publication process faster and offers an “open dialogue between authors and readers” (Tennant et al., 2019). As such, OPR has quite a similar function to the long-standing tradition of SSH book

review, intended to stimulate broader discussions about a specific piece of work. Hence, through OPR, authors benefit from more feedback and criticism and have more opportunity to improve their work, while external researchers gain an insight into research and readers may explain the reasons of their disagreement with the authors (Tennant, 2017; Tennant et al., 2019). While some consider that OPR will favour more strategic attitudes from the reviewer's side, the publication of the peer review reports may as well provide a further incentive for the peer reviewer to adopt an ethic and responsible attitude towards the author (Groves & Khan, 2010). Anyhow, although there is a lack of quantitative research on OPR (Squazzoni et al., 2017), first findings indicate that open peer review does not compromise on the quality of the peer reviewing process, at least when referees are able to protect their anonymity (Bravo et al., 2019; Pöschl, 2012). Furthermore, OPR makes it feasible for any reader or potential author to directly assess the quality and the efficiency of the review system of any given journal, and potentially look at the value for money of the requested articles processing charges. In our experience, OPR would make for a better quality proxy than most if not all existing journal level metrics currently in usage.

Finally, we would like to raise here a major epistemological and philosophical argument in favour of open peer review. Feminist philosophers of science (Harding, 1986; Longino, 1990) have argued that gender diversity and social openness made science more likely to be trustworthy, while homogeneity of scholarly communities weakens it and makes it less objective. Naomi Oreskes (2019) has generalised this argument and argues that the social nature of science – and notably the peer reviewing as a process intended to vet scientific results – is together with scientists' expertise to study the (social) world the main reason why people should trust science, rather than trust in a particular scientific method. “The crucial element of these practices [of peer review and tenure]”, she writes, “is their social and institutional character, which works to ensure that the judgements and opinions of no one person dominate and therefore that the value preferences and biases of no one person are controlling. [...] The social character of science forms the basis of its approach to objectivity and therefore the grounds on which we can trust it.” (Oreskes, 2019, p. 58). In this perspective, any initiative that tends to broaden the social interactions of experts around a scholarly piece of work beyond two or three selected peers should be considered as making it sounder and more reliable. And as such, OPR, although still in an infant age and presenting significant shortcomings, should be promoted and further elaborated through dedicated guidelines in the SSH fields as well as in the STEM disciplines (Ross-Hellauer & Görögh, 2019; Schmidt et al., 2018).

References

- Bravo, G., Grimaldo, F., López-Iñesta, E., Mehmani, B., & Squazzoni, F. (2019). The effect of publishing peer review reports on referee behavior in five scholarly journals. *Nature Communications*, 10, 322. <https://doi.org/10.1038/s41467-018-08250-2>
- Buriak, J.M. (2015), Mega-journals and peer review: can quality and standards survive? *Chemistry of Materials*, 27(7), 2243–2243. <https://doi.org/10.1021/acs.chemmater.5b01142>
- Chartron, G. (2014). Open access et SHS: Controverses. *Revue européenne des sciences sociales*, 52(1), 37–63. <https://doi.org/10.4000/ress.2658>

- Council of the European Union. (2016). *Council conclusions on the transition towards an Open Science system*. <https://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>
- Crossick, G. (2016). Monographs and open access. *Insights*, 29(1), 14–19. <https://doi.org/10.1629/uksg.280>
- European Commission. (2010). *A digital agenda for Europe. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*. European Commission. <https://eur-lex.europa.eu/LexUriServ/LexUriS-erv.do?uri=COM:2010:0245:FIN:EN:PDF>
- European Commission. (2012). *A reinforced European Research Area partnership for excellence and growth. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*. European Commission. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=2122
- European Commission. (2018). Recommendation of the Commission on access to and preservation of scientific information. *Official Journal of the European Union*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018H0790&from=EN>
- Eve, M.P. (2014). *Open access and the humanities: Contexts, controversies and the future*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316161012>
- Groves, T., & Khan, K. (2010). Is open peer review the fairest system? *BMJ*, 341, c6425. <https://doi.org/10.1136/bmj.c6425>
- Harding, S.G. (1986). *The science question in feminism*. Cornell University Press.
- Johnson, R., Watkinson, A., & Mabe, M. (2018). The 5th edition of *The STM Report: An overview of scientific and scholarly publishing*. https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf
- Longino, H.E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press.
- Moore, S.A. (2017). A genealogy of open access: negotiations between openness and access to research. *Revue française des sciences de l'information et de la communication*, 11. <https://doi.org/10.4000/rfsic.3220>
- Oreskes, N. (2019). *Why trust science?* Princeton University Press.
- Pöschl, U. (2012). Multi-stage open peer review: scientific evaluation integrating the strengths of traditional peer review with the virtues of transparency and self-regulation. *Frontiers in Computational Neuroscience*, 6, 33. <https://doi.org/10.3389/fncom.2012.00033>
- Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research*, 6, 588. <https://doi.org/10.12688/f1000research.11369.2>

- Ross-Hellauer, T., & Görögh, E. (2019). Guidelines for open peer review implementation. *Research Integrity and Peer Review*, 4, 4. <https://doi.org/10.1186/s41073-019-0063-9>
- Schmidt, B., Ross-Hellauer, T., van Edig, X., & Moylan, E.C. (2018). Ten considerations for open peer review. *F1000Research*, 7, 969. <https://doi.org/10.12688/f1000research.15334.1>
- Spezi, V., Wakeling, S., Pinfield, S., Fry, J., Creaser, C., & Willett, P. (2018). Let the community decide? The vision and reality of soundness-only peer review in open-access mega-journals. *Journal of Documentation*, 74(1), 137–161. <https://doi.org/10.1108/JD-06-2017-0092>
- Squazzoni, F., Brezis, E., & Marušić, A. (2017). Scientometrics of peer review. *Scientometrics*, 113(1), 501–502. <https://doi.org/10.1007/s11192-017-2518-4>
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, ‘translations’ and boundary objects: Amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, 19(3), 387–420. <https://doi.org/10.1177/030631289019003001>
- Teixeira da Silva, J. A. (2019). Challenges to open peer review. *Online Information Review*, 43(2), 197–200. <https://doi.org/10.1108/OIR-04-2018-0139>
- Tennant, J. P. (2017). What are the barriers to post-publication peer review? *LSE Impact Blog*. <https://blogs.lse.ac.uk/impactofsocialsciences/2017/04/12/what-are-the-barriers-to-post-publication-peer-review/>
- Tennant, J. P. (2018). The state of the art in peer review. *FEMS Microbiology Letters*, 365(19), fny204. <https://doi.org/10.1093/femsle/fny204>
- Tennant, J. P., Crane, H., Crick, T., Davila, J., Enkhbayar, A., Havemann, J., Kramer, B., Martin, R., Masuzzo, P., Nobes, A., Rice, C., Rivera-López, B., Ross-Hellauer, T., Sattler, S., Thacker, P. D., & Vanholsbeeck, M. (2019). Ten hot topics around scholarly publishing. *Publications*, 7(2), 34. <https://doi.org/10.3390/publications7020034>
- Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., Elkhatab, Y., B. Collister, L., Pikas, C. K., Crick, T., Masuzzo, P., Caravaggi, A., Berg, D. R., Niemeyer, K. E., Ross-Hellauer, T., Mannheimer, S., Rigling, L., Katz, D. S., Greshake Tzovaras, B., ... Colomb, J. (2017). A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Research*, 6, 1151. <https://doi.org/10.12688/f1000research.12037.3>
- Vanholsbeeck, M. (2017). La notion de Science Ouverte dans l’Espace européen de la recherche: Entre tendances à l’«exotérisation» et à la «gestionnarisation» de la recherche scientifique. *Revue française des sciences de l’information et de la communication*, 11. <https://doi.org/10.4000/rfsic.3241>

PART V: Conclusion

By Michael Ochsner, Nina Kancewicz-Hoffman, Jon Holm & Marek Hołowicki

Peer review is here to stay, 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 the mismatch between dimensions of quality as defined by peers and standard bibliometric indicators. However, peer review is not without challenges and – as any other socially embedded activity – evolves constantly. Moreover, peer review in the SSH faces particular challenges. A few of which were mentioned in this report, 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; lack of linear progress and 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 and national languages; challenges associated with growing interdisciplinarity and the Open Science agenda relevant for recent developments both in research and evaluation. To this, the general challenges of peer review are added, such as the risk of gender bias, regional bias, conservative bias and the workload for all parties involved.

The chapters in this report reveal that peer review takes a central role in academic life. Even metric-based evaluations involve a peer review component because even pure bibliometric approaches are based on citations to articles published in peer-reviewed journals and thus are not independent of peer review. Evaluation takes place at different *levels* (supervise PhD students, evaluation of articles for publication, recruitment procedures, evaluating research units, disciplines or institutions), can take different *perspectives* (ex-ante and ex-post evaluation) and serve different *purposes* (formative evaluation and summative, performance-based evaluation). In all these contexts peer review can take on different *roles* in the evaluation process (assessing research outcome, defining the weighting of publication channels etc.). The review process itself also comes in different *forms* (double-blind review, expert panels, informed peer review, open review etc.) and can be of different *significance* (deciding on a career, a project or the continuation of a research unit).

This means that peer review fulfils different functions and that peer review practices not only need to acknowledge different disciplinary particularities but also their evaluative context. There is not a single best practice, and the ambiguity of peer review identified in the chapter by Pölönen et al. will remain as long as the different functions and contexts of peer review are not taken into account.

The current criticism of the subjective and sometimes political nature of the activity of reviewing could be based on a too narrow vision of the role of peer review as just selecting “the best” or separating “excellence” from the rest. Peer review often takes a gatekeeping role, which is not a bias but its actual function in certain situations: the political action of selecting between different excellences that are not directly comparable in any objective

way. For example, in the context of scarce resources, some topics might be considered more important than others even though research is excellent also in those other topics.

This report thus calls for a more detailed and systematic analysis of peer review and its advantages and disadvantages taking into account its role, form and significance as well as the level, perspective and purpose of the evaluation procedure in which it is used. 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. SSH disciplines are rooted in a local context yet operate on an international level just as other disciplines, with the difference that internationality manifests itself by the use of multiple languages rather than English as the only lingua franca. Multilingualism as a mode of knowledge generation and dissemination is central to SSH research but it is also important for research evaluation – leading to additional challenges for peer review in the SSH as peers need to be able to acknowledge the local embeddedness of research and possibly to read the national language. Furthermore, SSH disciplines are in direct contact with their stakeholders, such as the wider public, the economy or public administration. Thus, a further challenge is the peer review of societal impact of research. This issue needs further investigation, not only regarding what societal impact is and how it can be traced but also regarding who a peer is in such a context.

A further challenge is connected with the increasing importance of evaluation. The burden for reviewers increases continuously as more scientific publication channels change to a double-blind peer review procedure, institutional evaluation is implemented more systematically at several levels and the share of competitive project funding is increasing, all of which need peer reviewers' time and energy. Solutions have to be found how to decrease reviewer burden and how to spread the workload on several shoulders. This is particularly relevant for the SSH disciplines that often act in a local context in national languages and include small disciplinary communities.

This report suggests ways for the SSH disciplines to respond to these challenges building on established epistemic practices while increasing the scientific and societal relevance of these disciplines. We believe that 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. In these troubled times, 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.

Bibliography

ENRESSH publications

- Derrick, G. & Ross-Helauer, T. (forthcoming). *Decolonising the social sciences and humanities through peer review*.
- 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>
- 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>
- Giménez-Toledo, E., Sivertsen, G., & Mañana-Rodríguez, J. (2017). Peer review as a delineation criterion in data sources for the assessment and measurement of scholarly book publishing in social sciences and humanities. In *Proceedings of the 16th International Conference on Scientometrics and Informetrics ISSI*, 16.-20. October at Wuhan University (pp. 118–124). Wuhan University.
- Helsinki Initiative (2019). *Helsinki Initiative on Multilingualism in Scholarly Communication*. Federation of Finnish Learned Societies, Committee for Public Information, Finnish Association for Scholarly Publishing, Universities Norway & European Network for Research Evaluation in the Social Sciences and the Humanities. <https://www.helsinki-initiative.org/> <https://doi.org/10.6084/m9.figshare.7887059>
- Hug, S. E., Holowiecki, M., Ma, L., Aeschbach, M., & Ochsner, M. (in preparation). *Criteria for reviewing journal manuscripts. A systematic review*.
- Kulczycki, E., Engels, T. C. E., Pölönen, J., Bruun, K., Dušková, M., Guns, R., Nowotniak, R., Petr, M., Sivertsen, G., Istenič Starčič, A., & Zuccala, A. (2018). Publication patterns in the social sciences and humanities: Evidence from eight European countries. *Scientometrics*, 116(1), 463–486. <https://doi.org/10.1007/s11192-018-2711-0>
- Kulczycki, E., Guns, R., Pölönen, J., Engels, T. C. E., Rozkosz, E. A., Zuccala, A. A., Bruun, K., Eskola, O., Starčič, A. I., Petr, M., & Sivertsen, G. (2020). Multilingual publishing in the social sciences and humanities: A seven-country European study. *Journal of the Association for Information Science and Technology*, asi.24336. <https://doi.org/10.1002/asi.24336>
- Mañana-Rodríguez, J., & Pölönen, J. (2018). Scholarly book publishers' ratings and lists in Finland and Spain: Comparison and assessment of the evaluative potential of merged lists. *Aslib Journal of Information Management*, 70(6), 643–659. <https://doi.org/10.1108/AJIM-05-2018-0111>
- Ochsner, M., & Dokmanović, M. (2017). Quality criteria and research obstacles in the SSH in Macedonia. In *2nd international conference on research evaluation in the social*

- sciences and humanities*, *RESSH 2017* (pp. 69–71). https://www.uantwerpen.be/images/uantwerpen/container41447/files/A5-BOOK_RESSH2017_170717-interactive.pdf
- Ochsner, M., Galleron, I., & Ionescu, A. (2017). *List of projects on SSH scholars' notions of research quality in participating countries*. ENRESSH Report Version 1.4. https://enressh.eu/wp-content/uploads/2017/09/Report_Quality_Projects.pdf
- Ochsner, M., Hug, S. E., & Galleron, I. (2017). The future of research assessment in the humanities: bottom-up assessment procedures. *Palgrave Communications*, 3, 17020. <https://doi.org/10.1057/palcomms.2017.20>
- 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>
- Pölönen, J., Engels, T.C.E., & Guns, R. (2020). Ambiguity in identification of peer-reviewed publications in the Finish and Flemish performance-based research funding systems. *Science and Public Policy*, 47(1), 1–15. <https://doi.org/10.1093/scipol/scz041>
- Pölönen, J., Engels, T.C.E., Guns, R., & Verleysen, F.T. (2017). Is my publication peer reviewed? A comparison of top-down and bottom-up identification of peer review in the framework of the Finnish and Flemish performance-based research funding systems. In *Conference abstracts: Science, Technology and Innovation indicators STI 2017. Open indicators: innovation, participation and actor-based STI indicators Paris 2017, 6–8 September 2017*.
- Ross-Hellauer, A., & Derrick, G. (2019). *Decolonising the social sciences and humanities through peer review*. The Third Research Evaluation in the SSH Conference (RESSH 2019), Valencia.
- Sīle, L., Pölönen, J., Sivertsen, G., Guns, R., Engels, T. C. E., Arefiev, P., Dušková, M., Faurbæk, L., Holl, A., Kulczycki, E., Macan, B., Nelhans, G., Petr, M., Pisk, M., Soós, S., Stojanovski, J., Stone, A., Šušol, J. & Teitelbaum, R. (2018) Comprehensiveness of national bibliographic databases for social sciences and humanities: findings from a European survey. *Research Evaluation*, 27(4), 310–322. <https://doi.org/10.1093/reseval/rvy016>
- Wroblewska, M. N. (2019). *Impact evaluation in Norway and in the UK: A comparative study, based on REF 2014 and Humeval 2015-2017*. ENRESSH. <https://doi.org/10.3990/4.2666-0776.2019.01>

Full bibliography

- Aagaard, K. (2018). Performance-based research funding in Denmark: The adoption and translation of the Norwegian model. *Journal of Data and Information Science*, 3(4), 20–30. <https://doi.org/10.2478/jdis-2018-0018>
- Académie des Sciences, Leopoldina, & Royal Society. (2017). *Statement by three national academies (Académie des Sciences, Leopoldina and Royal Society) on good practice in the evaluation of researchers and research programmes*. <https://www.academie-sciences.fr/pdf/rapport/avis111217.pdf>
- Albers, C. A., Floyd, R. G., Fuhrmann, M. J., & Martinez, R. S. (2011). Publication criteria and recommended areas of improvement within school psychology journals as reported by editors, journal board members, and manuscript authors. *Journal of School Psychology*, 49(6), 669–689. <https://doi.org/10.1016/j.jsp.2011.10.002>
- Albert, M., Laberge, S., & McGuire, W. (2012). Criteria for assessing quality in academic research: The views of biomedical scientists, clinical scientists and social scientists. *Higher Education*, 64(5), 661–676. <https://doi.org/10.1007/s10734-012-9519-2>
- All European Academies (ALLEA). (2017). *The European code of conduct for research integrity*. All European Academies.
- Azar, O. H. (2004). Rejections and the importance of first response times. *International Journal of Social Economics*, 31(3), 259–274. <https://doi.org/10.1108/03068290410518247>
- Bakanic, V., McPhail, C., & Simon, R. J. (1989). Mixed messages: referees' comments on the manuscripts they review. *The Sociological Quarterly*, 30(4), 639–654. <https://doi.org/10.1111/j.1533-8525.1989.tb01540.x>
- Baldwin, M. (2017). In referees we trust? *Physics Today*, 70(2), 44–49. <https://doi.org/10.1063/PT.3.3463>
- Baldwin, M. (2018). Scientific autonomy, public accountability, and the rise of “peer review” in the cold war United States. *Isis*, 109(3), 538–558. <https://doi.org/10.1086/700070>
- Batagelj, V., Ferligoj, A., & Squazzoni, F. (2017). The emergence of a field: a network analysis of research on peer review. *Scientometrics*, 113(1), 503–532. <https://doi.org/10.1007/s11192-017-2522-8>
- Bayley, J. E., & Phipps, D. (2019). Building the concept of research impact literacy. *Evidence & Policy: A Journal of Research, Debate and Practice*, 15(4), 597–606. <https://doi.org/10.1332/174426417X15034894876108>
- Bazeley, P. (2010). Conceptualising research performance. *Studies in Higher Education*, 35(8), 889–903. <https://doi.org/10.1080/03075070903348404>
- Becker, B. (1991). The quality and credibility of research reviews: What the editors say. *Personality and Social Psychology Bulletin*, 17(3), 267–272. <https://doi.org/10.1177/0146167291173006>

- Benneworth, P. (2015). Tracing how arts and humanities research translates, circulates and consolidates in society. How have scholars been reacting to diverse impact and public value agendas? *Arts and Humanities in Higher Education*, 14(1), 45–60. <https://doi.org/10.1177/1474022214533888>
- Benneworth, P., Gulbrandsen, M., & Hazelkorn, E. (2016). *The Impact and Future of Arts and Humanities Research*. Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-40899-0>
- Benneworth, P., & Jongbloed, B. W. (2010). Who matters to universities? A stakeholder perspective on humanities, arts and social sciences valorisation. *Higher Education*, 59(5), 567–588. <https://doi.org/10.1007/s10734-009-9265-2>
- Björk, B.-C., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. *Journal of Informetrics*, 7(4), 914–923. <https://doi.org/10.1016/j.joi.2013.09.001>
- Bonaccorsi, A. (2018). Peer review in social sciences and humanities. Addressing the interpretation of quality criteria. In A. Bonaccorsi (Ed.), *The evaluation of research in social sciences and humanities* (pp. 71–101). Springer International Publishing. https://doi.org/10.1007/978-3-319-68554-0_4
- Bonaccorsi, A. (2018). Towards an epistemic approach to evaluation in SSH. In A. Bonaccorsi (Ed.), *The evaluation of research in social sciences and humanities* (pp. 1–29). Springer. https://doi.org/10.1007/978-3-319-68554-0_1
- Bonjean, C. M., & Hullum, J. (1978). Reasons for journal rejection. An analysis of 600 manuscripts. *PS: Political Science & Politics*, 11(4), 480–483. <https://doi.org/10.1017/S1049096500004844>
- Bordage, G. (2001). Reasons reviewers reject and accept manuscripts: The strengths and weaknesses in medical education reports. *Academic medicine*, 76(9), 889–896. <https://doi.org/10.1097/00001888-200109000-00010>
- Bornmann, L., & Daniel, H.-D. (2006). Potential sources of bias in research fellowship assessments. Effects of university prestige and field of study on approval and rejection of fellowship applications. *Research Evaluation*, 15(3), 209–219. <https://doi.org/10.3152/147154406781775850>
- Bornmann, L. (2012). Measuring the societal impact of research: Research is less and less assessed on scientific impact alone – we should aim to quantify the increasingly important contributions of science to society. *EMBO Reports*, 13(8), 673–676. <https://doi.org/10.1038/embor.2012.99>
- 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., Mutz, R., & Daniel, H.-D., (2007). Gender differences in grant review: A meta-analysis. *Journal of Informetrics*, 1(3), 226–238. <https://doi.org/10.1016/j.joi.2007.03.001>

- Bornmann, L., & Daniel, H.-D. (2008a). The Effectiveness of the Peer Review Process: Inter-Referee Agreement and Predictive Validity of Manuscript Refereeing at *Angewandte Chemie*. *Angewandte Chemie International Edition*, 47(38), 7173–7178. <https://doi.org/10.1002/anie.200800513>
- Bornmann, L., & Daniel, H.-D. (2008b). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80. <https://doi.org/10.1108/00220410810844150>
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2008). How to detect indications of potential sources of bias in peer review: A generalized latent variable modeling approach exemplified by a gender study. *Journal of Informetrics*, 2(4), 280–287. <https://doi.org/10.1016/j.joi.2008.09.003>
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). A Reliability-Generalization Study of Journal Peer Reviews: A Multilevel Meta-Analysis of Inter-Rater Reliability and Its Determinants. *PLoS One*, 5(12), e14331. <https://doi.org/10.1371/journal.pone.0014331>
- Bornmann, L., Mutz, R., Marx, W., Schier, H., & Daniel, H.-D. (2011). A multilevel modelling approach to investigating the predictive validity of editorial decisions: do the editors of a high profile journal select manuscripts that are highly cited after publication? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(4), 857–879. <https://doi.org/10.1111/j.1467-985X.2011.00689.x>
- Borsuk, R. M., Aarssen, L. W., Budden, A. E., Koricheva, J., Leimu, R., Tregenza, T., & Lortie, C. J. (2009). To name or not to name: The effect of changing author gender on peer review. *BioScience*, 59(11), 985–989. <https://doi.org/10.1525/bio.2009.59.11.10>
- Burnhill, P.M., & Tubby-Hille, M. E. (2003). On Measuring the Relation between Social Science Research Activity and Research Publication. *Research Evaluation*, 4(3), 130–152. <https://doi.org/10.1093/rev/4.3.130>
- Bravo, G., Grimaldo, F., López-Iñesta, E., Mehmani, B., & Squazzoni, F. (2019). The effect of publishing peer review reports on referee behavior in five scholarly journals. *Nature Communications*, 10, 322. <https://doi.org/10.1038/s41467-018-08250-2>
- British Academy. (2007). *Peer Review: The challenges for the humanities and social sciences*. <https://www.thebritishacademy.ac.uk/publications/peer-review-challenges-humanities-and-social-sciences>
- Budden, A.E., Tregenza, T., Aarssen, L. W., Koricheva, J., Leimu, R., & Lortie, C. J., (2008). Double-blind review favours increased representation of female authors. *Trends in Ecology and Evolution*, 23(1), 4-6. <https://doi.org/10.1016/j.tree.2007.07.008>
- Buriak, J.M. (2015), Mega-journals and peer review: can quality and standards survive? *Chemistry of Materials*, 27(7), 2243–2243. <https://doi.org/10.1021/acs.chemmater.5b01142>

- Burrows, R. (2012). Living with the H-Index? Metric assemblages in the contemporary academy. *The Sociological Review*, 60(2), 355–372. <https://doi.org/10.1111/j.1467-954X.2012.02077.x>
- Butler, L. (2007). Assessing University Research: A Plea for a Balanced Approach. *Science and Public Policy*, 34(8), 565–574. <https://doi.org/10.3152/030234207X254404>
- Campion, M. A. (1993). Article review checklist. A criterion checklist for reviewing research articles in applied-psychology. *Personnel Psychology*, 46(3), 705–706. <https://doi.org/10.1111/j.1744-6570.1993.tb00896.x>
- Capaccioni, A., & Spina, G. (2018). Guidelines for peer review. A survey of international practices. In A. Bonaccorsi (Ed.), *The evaluation of research in social sciences and humanities* (pp. 55–69). Springer International Publishing. https://doi.org/10.1007/978-3-319-68554-0_3
- Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, 108(8), 3157–3162. <https://doi.org/10.1073/pnas.1014871108>
- Chartron, G. (2014). Open access et SHS: Controverses. *Revue européenne des sciences sociales*, 52(1), 37–63. <https://doi.org/10.4000/ress.2658>
- Chase, J. M. (1970). Normative criteria for scientific publication. *American Sociologist*, 5(3), 262–265.
- Chong, C. (1994). Editorial: We aim to publish – not to reject. *Canadian Journal of Plant Science*, 74(4). <https://doi.org/10.4141/cjps94-120>
- Collins, H.M., & Evans, R. (2007). *Rethinking Expertise*. University of Chicago Press.
- Coniam, D. (2011). Systematising System: one reviewer's analysis of the review process. *System*, 39(4), 539–553. <https://doi.org/10.1016/j.system.2011.10.018>
- Coryn, C. L. S., Hattie, J. A., Scriven, M., & Hartmann, D. J. (2007). Models and mechanisms for evaluating government-funded research: An international comparison. *American Journal of Evaluation*, 28(4), 437–457. <https://doi.org/10.1177/1098214007308290>
- Council of the European Union. (2016). *Council conclusions on the transition towards an Open Science system*. <https://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>
- Crossick, G. (2016). Monographs and open access. *Insights*, 29(1), 14–19. <https://doi.org/10.1629/uksg.280>
- Csiszar, A. (2017). How lives became lists and scientific papers became data: cataloguing authorship during the nineteenth century. *British Journal of History of Science*, 50(1), 23–60. <https://doi.org/10.1017/S0007087417000012>
- Daft, R. L. (1995). Why I recommended that your manuscript be rejected and what you can do about it. *Publishing in the Organizational Sciences*, 164–183. <https://doi.org/10.4135/9781452240466.n14>

- Dahler-Larsen, P. (2019). *Quality: From Plato to performance*. Palgrave Macmillan. <https://doi.org/10.1007/978-3-030-10392-7>
- Daniel, H.-D. (1993). *Guardians of science: Fairness and reliability of peer review*. VCH Verlagsgesellschaft. <https://doi.org/10.1002/3527602208>
- Day, F. C., Schriger, D. L., Todd, C., & Wears, R. L. (2002). The use of dedicated methodology and statistical reviewers for peer review: A content analysis of comments to authors made by methodology and regular reviewers. *Annals of Emergency Medicine*, 40(3), 329–333. <https://doi.org/10.1067/mem.2002.127326>
- Deem, R. (2016). Recent research evaluations in the UK and Portugal: Methodologies, processes, controversies and consequences. In C. Sarrico, P. Teixeira, A. Magalhães, A. Veiga, M. J. Rosa, & T. Carvalho (Eds.), *Global challenges, national initiatives, and institutional responses* (pp. 159–186). SensePublishers. https://doi.org/10.1007/978-94-6300-675-0_9
- de Jong, S. P. L., Smit, J., & van Drooge, L. (2016). Scientists' response to societal impact policies: A policy paradox. *Science and Public Policy*, 43(1), 102–114. <https://doi.org/10.1093/scipol/scv023>
- 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>
- Derrick, G. (2018). Impact from the evaluators' eye. In G. Derrick, *The evaluators' eye* (pp. 1–19). Springer International Publishing. https://doi.org/10.1007/978-3-319-63627-6_1
- Derrick, G. E., & Samuel, G. N. (2016). The evaluation scale: exploring decisions about societal impact in peer review panels. *Minerva*, 54(1), 75–97. <https://doi.org/10.1007/s11024-016-9290-0>
- Derrick, G., & Samuel, G. (2017). The future of societal impact assessment using peer review: pre-evaluation training, consensus building and inter-reviewer reliability. *Palgrave Communications*, 3, palcomms201740. <https://doi.org/10.1057/palcomms.2017.40>
- Derrick, G. E., & Samuel, G. S. (2018). Exploring the degree of delegated authority for the peer review of societal impact. *Science and Public Policy*, 45(5), 673–682. <https://doi.org/10.1093/scipol/scx091>
- Derrick, G. & Ross-Helauer, T. (forthcoming). *Decolonising the social sciences and humanities through peer review*.
- Derricourt, R. (2012). Peer review: Fetishes, fallacies, and perceptions. *Journal of Scholarly Publishing*, 43(2), 137–147. <https://doi.org/10.1353/scp.2012.0000>
- D'Este, P., Ramos-Vielba, I., Woolley, R., & Amara, N. (2018). How do researchers generate scientific and societal impacts? Toward an analytical and operational framework. *Science and Public Policy*, 45(6), 752–763. <https://doi.org/10.1093/scipol/scy023>

- Dickersin, K., Ssemanda, E., Mansell, C., & Rennie, D. (2007). What do the JAMA editors say when they discuss manuscripts that they are considering for publication? Developing a schema for classifying the content of editorial discussion. *BMC Medical Research Methodology*, 7, Article 44. <https://doi.org/10.1186/1471-2288-7-44>
- Donovan, C. (2007). The qualitative future of research evaluation. *Science and Public Policy*, 34(8), 585–597. <https://doi.org/10.3152/030234207X256538>
- DORA. (2012). San Francisco Declaration on Research Assessment. <https://sfdora.org/read/>
- Dubois-Shaik, F., & Fusulier, B. (2017). Understanding gender inequality and the role of the work/family interface in contemporary academia: An introduction. *European Educational Research Journal*, 16(2–3), 99–105. <https://doi.org/10.1177/1474904117701143>
- Dunlop, C. A. (2018). The political economy of politics and international studies impact: REF2014 case analysis. *British Politics*, 13(3), 270–294. <https://doi.org/10.1057/s41293-018-0084-x>
- Engels, T. C. E., & Guns, R. (2018). The Flemish performance-based research funding system: A unique variant of the Norwegian model. *Journal of Data and Information Science*, 3(4): 45–60. <https://doi.org/10.2478/jdis-2018-0020>
- Enserink, M. (2015). Dutch sexism study comes under fire. *Science*, 364(6446). <https://doi.org/10.1126/science.aad4618>
- ERIH PLUS. (2014). *European Reference Index for the Humanities and Social Sciences*. NSD. <https://dbh.nsd.uib.no/publiseringskanaler/erihplus/about/index>
- European Commission. (2010). *A digital agenda for Europe. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*. European Commission. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:EN:PDF>
- European Commission. (2010). *Assessing Europe's university-based research-expert group on assessment of university-based research*. Publications Office of the European Union. https://ec.europa.eu/research/science-society/document_library/pdf_06/assessing-europe-university-based-research_en.pdf
- European Commission. (2012). *A reinforced European Research Area partnership for excellence and growth. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*. European Commission. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=2122
- European Commission. (2013). *Options for strengthening Responsible Research and Innovation. Report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation*. European Commission. <https://doi.org/10.2777/46253>

- European Commission. (2016). *She Figures 2015. Gender in Research and Innovation*. European Union. https://ec.europa.eu/research/swafs/pdf/pub_gender_equality/she_figures_2015-final.pdf
- European Commission. (2018). *Integration of social sciences and humanities in Horizon 2020: Participants, budget and disciplines. 3rd monitoring report on SSH flagged projects funded in 2016 under the Societal Challenges and Industrial Leadership priorities*. Publications Office of the European Union. <https://doi.org/10.2777/756427>
- European Commission. (2018). Recommendation of the Commission on access to and preservation of scientific information. *Official Journal of the European Union*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018H0790&from=EN>
- European Research Council (ERC). (2017). *ERC rules for submission & evaluation. Commission Decision C(2017)4750. Version 3.0, 14 July 2017*. https://ec.europa.eu/research/participants/data/ref/h2020/sgl/erc/h2020-erc-se-rules-amended2_en.pdf
- European Research Council (ERC). (2019). *Frontier Research Grants. Guide for peer reviewers applicable to the ERC Starting, Consolidator & Advanced Grants (ERC Work Programme 2019)*. https://ec.europa.eu/research/participants/data/ref/h2020/other/experts_manual/h2020_peer-review_erc-stg-cog-adg-2019_en.pdf
- European Science Foundation. (2011). *European peer review guide: Integrating policies and practices into coherent procedures*. http://archives.esf.org/fileadmin/Public_documents/Publications/European_Peer_Review_Guide_01.pdf
- European Science Foundation. (2015). *ESF report on the Foundation for Science and Technology (FCT) evaluation of R&D Units 2013–2014*. European Science Foundation. https://www.fct.pt/apoios/unidades/avaliacoes/2013/docs/ESF_FCT_report_Final.pdf
- Eve, M.P. (2014). *Open access and the humanities: Contexts, controversies and the future*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316161012>
- Ezeala, C., Nweke, I., & Ezeala, M. (2013). Common errors in manuscripts submitted to medical science journals. *Annals of medical and health sciences research*, 3(3), 376–379. <https://doi.org/10.4103/2141-9248.117957>
- Fiske, D. W., & Fogg, L. (1990). But the reviewers are making different criticisms of my paper!: diversity and uniqueness in reviewer comments. *American Psychologist*, 45(5), 591–598. <https://doi.org/10.1037/0003-066X.45.5.591>
- Flückiger, A., & Tanquerel, T. (Eds.). (2015). *L'évaluation de la recherche en droit / Assessing research in law. Enjeux et méthodes / Stakes and methods*. Bruylant.
- Flyvbjerg, B. (2001). *Making social science matter: Why social inquiry fails and how it can succeed again*. Cambridge Univ. Press.

- Frantz, T. T. (1968). Criteria for publishable manuscripts. *Personnel and Guidance Journal*, 47(4), 384–386. <https://doi.org/10.1002/j.2164-4918.1968.tb02944.x>
- Friesen, H. G. (1998). Equal opportunities in Canada. *Nature*, 391, 326. <https://doi.org/10.1038/34768>
- 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>
- GENDERACTION. (2019). *Report on strategic advice for enhancing the gender dimension of open science and innovation policy*. Institute of Sociology of the Czech Academy of Sciences. https://genderaction.eu/wp-content/uploads/2019/04/GENDERACTION_Report-5.1_D11_OSOL.pdf
- 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>
- 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. (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. (Ed.) (2017). *La edición académica española. Indicadores y características*. Madrid: Federación de Gremios de Editores de España. Informe. <http://ilia.cchs.csic.es/SPI/spi-fgee/docs/EAEV1.pdf>
- Giménez-Toledo, E., & Córdoba-Restrepo, J. F. (Eds.). (2018). *Edición académica y difusión. Libro abierto en Iberoamérica*. Universidad del Rosario. <https://doi.org/10.12804/th9789587841671>
- 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>
- Giménez-Toledo, E., Sivertsen, G., & Mañana-Rodríguez, J. (2017). Peer review as a delineation criterion in data sources for the assessment and measurement of scholarly book publishing in social sciences and humanities. In *Proceedings of the 16th*

- International Conference on Scientometrics and Informetrics ISSI*, 16.-20. October at Wuhan University (pp. 118–124). Wuhan University.
- Giménez-Toledo, E., Tejada-Artigas, C., & Mañana-Rodríguez, J. (2013). Evaluation of scientific books' publishers in social sciences and humanities: Results of a survey. *Research Evaluation*, 22(1), 64–77. <https://doi.org/10.1093/reseval/rvs036>
- Gogolin, I. (2011). *European Educational Research Quality Indicators (EERQI). Project final report*. http://eerqi.eu/sites/default/files/Final_Report.pdf
- Gozlan, C. (2016). Les sciences humaines et sociales face aux standards d'évaluation de la qualité académique. *Sociologie*, 7(3), 261–21. <https://doi.org/10.3917/socio.073.0261>
- Greenhalgh, T., & Fahy, N. (2015). Research impact in the community-based health sciences: An analysis of 162 case studies from the 2014 UK Research Excellence Framework. *BMC Medicine*, 13(1). <https://doi.org/10.1186/s12916-015-0467-4>
- Groves, T., & Khan, K. (2010). Is open peer review the fairest system? *BMJ*, 341, c6425. <https://doi.org/10.1136/bmj.c6425>
- Guetzkow, J., Lamont, M., & Mallard, G., (2004). What is originality in the humanities and the social sciences? *American Sociological Review*, 69(2), 190–212. <https://doi.org/10.1177%2F000312240406900203>
- Gutwirth, S. (2010). The evaluation of legal science. The VI.I.R.-model for integral quality assessment of research in law: What next? In H. Eisendrath, J. P. van Bendegem (Eds.), *It takes two to do science. the puzzling interactions between science and society* (pp. 69–79). ASP.
- Hammarfelt, B. (2017). Four claims on research assessment and metric use in the humanities. *Bulletin of the Association for Information Science and Technology*, 43(5), 33–38. <https://doi.org/10.1002/bul2.2017.1720430508>
- Hammarfelt, B., Nelhans, G., Eklund, P., & Åström, F. (2016). The heterogeneous landscape of bibliometric indicators. Evaluating models for allocating resources at Swedish universities. *Research Evaluation*, 25(3), 292–305. <https://doi.org/10.1093/reseval/rvv040>
- Harding, S.G. (1986). *The science question in feminism*. Cornell University Press.
- Hartmann, I. (1990). *Begutachtung in der Forschungsförderung: die Argumente der Gutachter in der Deutschen Forschungsgemeinschaft*. RG Fischer.
- Hartmann, I., & Neidhardt, F. (1990). Peer review at the Deutsche Forschungsgemeinschaft. *Scientometrics*, 19(5–6), 419–425. <https://doi.org/10.1007/BF02020704>
- Hebrang Grgić, I., & Stojanovski, J. (2017). Peer review in Croatian open access scholarly and professional journals: A cross-disciplinary survey. In *Eighth International Congress on Peer Review and Scientific Publication: Enhancing the quality and credibility of science*. Chicago. <https://peerreviewcongress.org/prc17-0364>

- Heilbron, J., Boncourt, T., Schögler, R., & Sapiro, G. (2017). *European Social Sciences and Humanities (SSH) in a Global Context. Preliminary findings from the INTERCO-SSH Project*. <http://interco-ssh.eu/wp-content/uploads/2017/02/European-Social-Science-in-a-Global-Contextv2.pdf>
- Helmer, M., Schottdorf, M., Neef, A., & Battaglia, D. (2017). Gender bias in scholarly peer Review. *eLife*, 6, e21718. <https://doi.org/10.7554/eLife.21718>
- Helsinki Initiative (2019). *Helsinki Initiative on Multilingualism in Scholarly Communication*. Federation of Finnish Learned Societies, Committee for Public Information, Finnish Association for Scholarly Publishing, Universities Norway & European Network for Research Evaluation in the Social Sciences and the Humanities. <https://www.helsinki-initiative.org/> <https://doi.org/10.6084/m9.figshare.7887059>
- Hemlin, S. (1993). Scientific quality in the eyes of the scientist. A questionnaire study. *Scientometrics*, 27(1), 3–18. <https://doi.org/10.1007/BF02017752>
- Hesterman, C. M., Szperka, C. L., & Turner, D. P. (2018). Reasons for manuscript rejection after peer review from the journal headache. *Headache: The Journal of Head and Face Pain*. <https://doi.org/10.1111/head.13343>
- Hewings, M. (2004). An ‘important contribution’ or ‘tiresome reading’? A study of evaluation in peer reviews of journal article submissions. *Journal of Applied Linguistics*, 1(3), 247–274. <https://doi.org/10.1558/jal.v1i3.247>
- Hicks, D. (2004). The four literatures of social science. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 473–496). Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-2755-9_22
- 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>
- Hicks, D., Wouters, P. F., Waltman, L., de Rijcke, S., and Rafols, I. (2015) The Leiden Manifesto for research metrics: Use these 10 principles to guide research evaluation. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>
- Hornung, A., Khlavna, V., & Korte, B. (2016). Research rating Anglistik/Amerikanistik of the German Council of Science and Humanities. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research assessment in the humanities. Towards criteria and procedures* (pp. 219–233). Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_17
- House of Commons Science and Technology Committee. (2011). *Peer review in scientific publications*. HC 856. House of Commons. <https://publications.parliament.uk/pa/cm201012/cmselect/cmsctech/856/856.pdf>
- Hug, S. E., & Aeschbach, M. (2019). Criteria for assessing grant applications. A systematic review. *Palgrave Communications*, 6, Article 37. <https://doi.org/10.1057/s41599-020-0412-9>
- Hug, S. E., Holowiecki, M., Ma, L., Aeschbach, M., & Ochsner, M. (in prepration). *Criteria for reviewing journal manuscripts. A systematic review*.

- 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>
- Hug, S. E., Ochsner, M., & Daniel, H.-D. (2014). A framework to explore and develop criteria for assessing research quality in the humanities. *International Journal for Education Law and Policy*, 10(1), 64.
- Hughes, T., Webber, D., & O'Regan, N. (2019). Achieving wider impact in business and management: Analysing the case studies from REF 2014. *Studies in Higher Education*, 44(4), 628–642. <https://doi.org/10.1080/03075079.2017.1393059>
- Huisman, J., & Smits, J. (2017). Duration and quality of the peer review process: The author's perspective. *Scientometrics*, 113(1), 633–650. <https://doi.org/10.1007/s11192-017-2310-5>
- Husu, L., & de Cheveigné, S. (2010). Gender and gatekeeping of excellence in research funding: European perspectives. In B. Riegraf, B. Aulenbacher, E. Kirsch-Auwärter & U. Müller (Eds), *Gender change in academia* (pp. 43–59). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92501-1_4
- Institut de France, Académie des Sciences. (2011). *Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs*. <https://www.academie-sciences.fr/pdf/rapport/avis170111.pdf>
- Institut de France, Académie des Sciences. (2014). *Les nouveaux enjeux de l'édition scientifique*. http://www.academie-sciences.fr/pdf/rapport/rads_241014.pdf
- INTERCO-SSH. (2013). *International Collaboration in the SSH*. <http://interco-ssh.eu/en/>
- Irwin, A. (1995). *Citizen science: A study of people, expertise and sustainable development*. Routledge.
- Johnson, R., Watkinson, A., & Mabe, M. (2018). The 5th edition of *The STM Report: An overview of scientific and scholarly publishing*. https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf
- Johnston, R. (2008). On structuring subjective judgements: Originality, significance and rigour in RAE2008. *Higher Education Quarterly*, 62(1-2), 120–147. <https://doi.org/10.1111/j.1468-2273.2008.00378.x>
- Jonkers, K., & Zacharewicz, T. (2016). *Research Performance Based Funding Systems: a Comparative Assessment* (No. EUR 27837 EN). European Commission.
- Kaltenbrunner, W., & de Rijcke, S. (2016). Quantifying 'output' for evaluation: Administrative knowledge politics and changing epistemic cultures in Dutch law faculties. *Science and Public Policy*, 44(2), 1–10. <https://doi.org/10.1093/scipol/scw064>
- Kattel, R., & Mazzucato, M. (2018). Mission-oriented innovation policy and dynamic capabilities in the public sector. *Industrial and Corporate Change*, 27(5), 787–801. <https://doi.org/10.1093/icc/dty032>

- König, T. (2016). Peer review in the social sciences and humanities at the European level: The experiences of the European Research Council. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research assessment in the humanities. Towards criteria and procedures* (pp. 151–163). Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_12
- Korytkowski, P., & Kulczycki, E. (2019). *Dominant language of researchers across fields*. The Third Research Evaluation in the SSH Conference (RESSH 2019), Valencia. https://ressh2019.webs.upv.es/wp-content/uploads/2019/10/ressh_2019_paper_21.pdf
- Kretschner, H., Pudovkin, A., & Stegmann J. (2012). Research evaluation. Part II: Gender effects of evaluation: Are men more productive and more cited than women? *Scientometrics*, 93, 17–30. <https://doi.org/10.1007/s11192-012-0658-0>
- Krull, W., & Tepperwien, A. (2016). The four “T”s: Quality indicators for the humanities. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research Assessment in the Humanities. Towards criteria and procedures* (pp. 165–179). Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_13
- Kulczycki, E., Engels, T. C. E., Pölönen, J., Bruun, K., Dušková, M., Guns, R., Nowotniak, R., Petr, M., Sivertsen, G., Istenič Starčič, A., & Zuccala, A. (2018). Publication patterns in the social sciences and humanities: Evidence from eight European countries. *Scientometrics*, 116(1), 463–486. <https://doi.org/10.1007/s11192-018-2711-0>
- Kulczycki, E., Guns, R., Pölönen, J., Engels, T. C. E., Rozkosz, E. A., Zuccala, A. A., Bruun, K., Eskola, O., Starčič, A. I., Petr, M., & Sivertsen, G. (2020). Multilingual publishing in the social sciences and humanities: A seven-country European study. *Journal of the Association for Information Science and Technology*, asi.24336. <https://doi.org/10.1002/asi.24336>
- Lack, E. (2008). Einleitung – Das Zauberwort “Standards”. In E. Lack & C. Marksches (Eds.), *What the hell is quality? Qualitätsstandards in den Geisteswissenschaften* (pp. 9–34). Campus.
- Lahtinen, E., Koskinen-Ollonqvist, P., Rouvinen-Wilenius, P., Tuominen, P., & Mittelmark, M. B. (2005). The development of quality criteria for research: A Finnish approach. *Health Promotion International*, 20(3), 306–315. <https://doi.org/10.1093/heapro/dai008>
- 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., Bloch, C. W., & Sivertsen, G. (2015). Options and limitations in measuring the impact of research grants – evidence from Denmark and Norway. *Research Evaluation*, 24(3), 256–270. <https://doi.org/10.1093/reseval/rvv012>

- Langfeldt, L., & Scordato, L. (2016). *Efficiency and flexibility in research funding. A comparative study of funding instruments and review criteria*. NIFU Report No. 2016:9. Oslo: NIFU.
- Lawrence, P. A. (2002). Rank injustice. *Nature*, 415(6874), 835–836. <https://doi.org/10.1038/415835a>
- Leão, D., Angelaki, M., Bertino, A., Dumouchel, S., & Vidal, F. (2018). *Operas multilingualism white paper*. <https://doi.org/10.5281/ZENODO.1324026>
- Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1), 2–17. <https://doi.org/10.1002/asi.22784>
- 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>
- Leung, D., Law, R., Kucukusta, D., & Guillet, B. D. (2014). How to review journal manuscripts: A lesson learnt from the world's excellent reviewers. *Tourism Management Perspectives*, 10, 46–56. <https://doi.org/10.1016/j.tmp.2014.01.003>
- Lewis, J. (2018). The locus of legitimate interpretation in social science and what it means for public understanding. *Communication at Science in Public 2018 Intersecting Science 2018 Conference, Cardiff, 17–19 December 2018*. https://docs.wix-static.com/ugd/48bd50_66b801501d204ed3a123b84b947ac823.pdf
- Ley, T. J., & Hamilton, B. H. (2008). The gender gap in NIH grant applications. *Science*, 322(5907), 1472–1474. <https://doi.org/10.1126/science.1165878>
- Longino, H.E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press.
- Lotriet, C. J. (2012). Reviewing the review process: Identifying sources of delay. *Australasian Medical Journal*, 5(1), 26–29. <https://doi.org/10.4066/AMJ.2012.1165>
- Luukkonen, T. (2012). Conservatism and risk-taking in peer review: Emerging ERC practices. *Research Evaluation*, 21(1), 48–60. <https://doi.org/10.1093/reseval/rvs001>
- MacRoberts, M. H., & MacRoberts, B. R. (2018). The mismeasure of science: Citation analysis. *Journal of the Association for Information Science and Technology*, 69(3), 474–482. <https://doi.org/10.1002/asi.23970>
- Mallard, G., Lamont, M., & Guetzkow, J. (2009). Fairness as appropriateness: Negotiating epistemological differences in peer review. *Science, Technology, & Human Values*, 34(5), 573–606. <https://doi.org/10.1177/0162243908329381>

- Mañana-Rodríguez, J., & Pölönen, J. (2018). Scholarly book publishers' ratings and lists in Finland and Spain: Comparison and assessment of the evaluative potential of merged lists. *Aslib Journal of Information Management*, 70(6), 643–659. <https://doi.org/10.1108/AJIM-05-2018-0111>
- Marsh, H. W., Bornmann, L., Mutz, R., Daniel, H.-D., & O'Mara, A. (2009). Gender effects in the peer reviews of grant proposals: A comprehensive meta-analysis comparing traditional and multilevel approaches. *Review of Educational Research*, 79(3), 1290–1326. <https://doi.org/10.3102%2F0034654309334143>
- McKercher, B., Law, R., Weber, K., Song, H., & Hsu, C. (2007). Why referees reject manuscripts. *Journal of Hospitality and Tourism Research*, 31(4), 455–470. <https://doi.org/10.1177/1096348007302355>
- Molinié, A., & Bodenhausen, G. (2010). Bibliometrics as weapons of mass citation / La bibliométrie comme arme de citation massive. *CHIMIA International Journal for Chemistry*, 64(1), 78–89. <https://doi.org/10.2533/chimia.2010.78>
- Moore, S.A. (2017). A genealogy of open access: negotiations between openness and access to research. *Revue française des sciences de l'information et de la communication*, 11. <https://doi.org/10.4000/rfsic.3220>
- Moxham, N., & Fyfe, A. (2018). The Royal Society and the prehistory of peer review, 1665-1965. *Historical Journal*, 61(4), 863–889. <https://doi.org/10.1017/S0018246X17000334>
- Mulligan, A., Hall, L., & Raphael, E. (2013). Peer review in a changing world: An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology*, 64(1), 132–161. <https://doi.org/10.1002/asi.22798>
- Mutz, R., Bornmann, L., & Daniel, H.-D. (2012). Does gender matter in grant peer review? *Zeitschrift für Psychologie*, 220(2), 121–129. <https://doi.org/10.1027/2151-2604/a000103>
- Mutz, R. D., Bornmann, L., & Daniel, H.-D. (2015). Testing for the fairness and predictive validity of research funding decisions: A multilevel multiple imputation for missing data approach using ex-ante and ex-post peer evaluation data from the Austrian Science Fund. *Journal of the Association for Information Science and Technology*, 66(11), 2321–2339. <https://doi.org/10.1002/asi.23315>
- Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, 66(1), 81–100. <https://doi.org/10.1007/s11192-006-0007-2>
- Nederhof, A. J., & Zwaan, R. A. (1991). Quality judgements of journals as indicators of research performance in the humanities and the social and behavioral sciences. *Journal of the American Society for Information Science*, 42(5), 332–340. [https://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5%3C332::AID-ASI3%3E3.0.CO;2-8](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5%3C332::AID-ASI3%3E3.0.CO;2-8)

- Nicholas, D., Watkinson, A., Jamali, H.R., Herman, E., Tenopir, C., Volentine, R., Allard, S., & Levine, K. (2015). Peer review: still king in the digital age. *Learned Publishing*, 28(1), 15–21. <https://doi.org/10.1087/20150104>
- Nielsen, M. W. (2015). Gender inequality and research performance: moving beyond individual-meritocratic explanations of academic advancement. *Studies in Higher Education*, 41(11), 2044–2060. <https://doi.org/10.1080/03075079.2015.1007945>
- Nutley, S. M., Walter, I., & Davies, H. T. O. (2007). *Using evidence* (1st ed.). Bristol University Press. <https://doi.org/10.2307/j.ctt9qgw1>
- Ochsner, M., & Dokmanović, M. (2017). Quality criteria and research obstacles in the SSH in Macedonia. In *2nd international conference on research evaluation in the social sciences and humanities, RESSH 2017* (pp. 69–71). https://www.uantwerpen.be/imagages/uantwerpen/container41447/files/A5-BOOK_RESSH2017_170717-interactive.pdf
- Ochsner, M., Galleron, I., & Ionescu, A. (2017). *List of projects on SSH scholars' notions of research quality in participating countries*. ENRESSH Report Version 1.4. https://enressh.eu/wp-content/uploads/2017/09/Report_Quality_Projects.pdf
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2012). Indicators for research quality in the humanities: Opportunities and limitations. *Bibliometrie - Praxis und Forschung*, 1, 4. <https://doi.org/10.5283/bpf.157>
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2013). Four types of research in the humanities: Setting the stage for research quality criteria in the humanities. *Research Evaluation*, 22(2), 79–92. <https://doi.org/10.1093/reseval/rvs039>
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2014). Setting the stage for the assessment of research quality in the humanities. Consolidating the results of four empirical studies. *Zeitschrift für Erziehungswissenschaft*, 17(6), 111–132. <https://doi.org/10.1007/s11618-014-0576-4>
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (Eds.). (2016). *Research assessment in the humanities. Towards criteria and procedures*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-29016-4>
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2017). Assessment Criteria for Early Career Researcher's Proposals in the Humanities. In *Proceedings of the 16th International Conference on Scientometrics and Informetrics ISSI*, 16.-20. October at Wuhan University (pp. 105–111). Wuhan University.
- Ochsner, M., Hug, S. E., & Galleron, I. (2017). The future of research assessment in the humanities: bottom-up assessment procedures. *Palgrave Communications*, 3, 17020. <https://doi.org/10.1057/palcomms.2017.20>
- 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>

- OPERAS. (2017). Open Access in the ERA through scholarly communication. <https://operas.hypotheses.org/>
- Oreskes, N. (2019). *Why trust science?* Princeton University Press.
- Ossenblok, T. L. B., Engels, T. C. E., & Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science: A comparison of publication patterns and incentive structures in Flanders and Norway (2005-9). *Research Evaluation*, 21(4), 280–290. <https://doi.org/10.1093/reseval/rvs019>
- Ossenblok, Truyken L.B., Verleysen, F. T., & Engels, T. C. E. (2014). Coauthorship of journal articles and book chapters in the social sciences and humanities (2000-2010): Coauthorship of journal articles and book chapters in the social sciences and humanities (2000-2010). *Journal of the Association for Information Science and Technology*, 65(5), 882–897. <https://doi.org/10.1002/asi.23015>
- Peruginelli, G., & Faro, S. (2018). Research quality evaluation: The case of legal studies. In A. Bonaccorsi (Ed.), *The evaluation of research in social sciences and humanities* (pp. 103–129). Springer International Publishing. https://doi.org/10.1007/978-3-319-68554-0_5
- Peruginelli, G., Faro, S., Agnoloni, T. (2018). A survey on legal research monograph evaluation in Italy. In A. Bonaccorsi (Ed.), *The evaluation of research in social sciences and humanities* (pp. 211–230). Springer International Publishing. https://doi.org/10.1007/978-3-319-68554-0_9
- Pier, E. L., Brauer, M., Filut, A., Kaatz, A., Raclaw, J., Nathan, M. J., Ford, C. E., & Carnes, M. (2018). Low agreement among reviewers evaluating the same NIH grant applications. *Proceedings of the National Academy of Sciences of the United States of America*, 115(12), 2952–2957. <https://doi.org/10.1073/pnas.1714379115>
- Phipps, D., Cummins, J., Pepler, D. J., Craig, W., & Cardinal, S. (2016). The Co-produced Pathway to Impact Describes Knowledge Mobilization Processes. *Journal of Community Engagement and Scholarship*, 9(1), Article 5. <https://digitalcommons.northgeorgia.edu/jces/vol9/iss1/5>
- Polanyi, M. (1967). *The tacit dimension*. Routledge & Kegan Paul.
- Pollitt, F. A., Notgrass, C. M., & Windle, C. (1996). Mental health services research. Peer review of rural research grant applications. *Administration and Policy in Mental Health and Mental Health Services Research*, 24(2), 173–180. <https://doi.org/10.1007/BF02042489>
- Pölönen, J. (2018). Applications of, and experiences with, the Norwegian model in Finland. *Journal of Data and Information Science*, 3(4), 31–44. <https://doi.org/10.2478/jdis-2018-0019>
- Pölönen, J., Engels, T.C.E., & Guns, R. (2020). Ambiguity in identification of peer-reviewed publications in the Finish and Flemish performance-based research funding systems. *Science and Public Policy*, 47(1), 1–15. <https://doi.org/10.1093/scipol/scz041>

- Pölönen, J., Engels, T.C.E., Guns, R., & Verleysen, F.T. (2017). Is my publication peer reviewed? A comparison of top-down and bottom-up identification of peer review in the framework of the Finnish and Flemish performance-based research funding systems. In *Conference abstracts: Science, Technology and Innovation indicators STI 2017. Open indicators: innovation, participation and actor-based STI indicators Paris 2017, 6–8 September 2017*.
- Pöschl, U. (2012). Multi-stage open peer review: scientific evaluation integrating the strengths of traditional peer review with the virtues of transparency and self-regulation. *Frontiers in Computational Neuroscience*, 6, 33. <https://doi.org/10.3389/fncom.2012.00033>
- Publons. (2018). *Global State of Peer Review*. <https://publons.com/community/gspr>
- Ramulu, V. G., Levine, R. B., Hebert, R. S., & Wright, S. M. (2005). Development of a case report review instrument. *International Journal of Clinical Practice*, 59(4), 457–461. <https://doi.org/10.1111/j.1368-5031.2005.00319.x>
- 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>
- Reinhart, M. (2010). Peer review practices: A content analysis of external reviews in science funding. *Research Evaluation*, 19(5), 317–431. <https://doi.org/10.3152/095820210X12809191250843>
- Research Information Network (RIN). (2010). *Peer review: A guide for researchers*. www.rin.ac.uk/peer-review-guide
- Research Information Network (RIN). (2015). *Scholarly communication and peer review: The current landscape and future trends. A report commissioned by the Wellcome Trust*. <https://wellcome.ac.uk/sites/default/files/scholarly-communication-and-peer-review-mar15.pdf>
- Roberts, T. J., & Shambrook, J. (2012). Academic excellence: A commentary and reflections on the inherent value of peer review. *Journal of Research Administration*, 43(1), 33–38.
- Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research*, 6, 588. <https://doi.org/10.12688/f1000research.11369.2>
- Ross-Hellauer, T., Deppe, A., & Schmidt, B. (2017). Survey on open peer review: Attitudes and experience amongst editors, authors and reviewers. *PLOS ONE*, 12(12), e0189311. <https://doi.org/10.1371/journal.pone.0189311>
- Ross-Hellauer, A., & Derrick, G. (2019). *Decolonising the social sciences and humanities through peer review*. The Third Research Evaluation in the SSH Conference (RESSH 2019), Valencia.

- Ross-Hellauer, T., & Görögh, E. (2019). Guidelines for open peer review implementation. *Research Integrity and Peer Review*, 4, 4. <https://doi.org/10.1186/s41073-019-0063-9>
- Rowley, J., & Sbaffi, L. (2018). Academics' attitudes towards peer review in scholarly journals and the effect of role and discipline. *Journal of Information Science*, 44(5), 644–657. <https://doi.org/10.1177/0165551517740821>
- Sandström, U., & Hallsten, M. (2008). Persistent nepotism in peer-review. *Scientometrics*, 74(2), 175–189. <https://doi.org/10.1007/s11192-008-0211-3>
- Sapiro, G., Seiler-Juilleret, H. (2016). Disseminating the social sciences and humanities. European Policy Brief, September 2016. INTERCO-SSH. https://ec.europa.eu/research/social-sciences/pdf/policy_briefs/interco-ssh_pb_092016.pdf
- Schmidt, B., Ross-Hellauer, T., van Edig, X., & Moylan, E.C. (2018). Ten considerations for open peer review. *F1000Research*, 7, 969. <https://doi.org/10.12688/f1000research.15334.1>
- Schmitt, J., Petzold, T., Nellessen-Martens, G., & Pfaff, H. (2015). Priorisierung und Konsentierung von Begutachtungs-, Förder- und Evaluationskriterien für Projekte aus dem Innovationsfonds: Eine multiperspektivische Delphi-Studie. [Prioritization and consentation of criteria for the appraisal, funding and evaluation of projects from the German Innovation Fund: a multi-perspective Delphi study]. *Gesundheitswesen*, 77(8–9), 570–579. <https://doi.org/10.1055/s-0035-1555898>
- Science Europe. (June, 2019). *Symposium report: Interdisciplinarity*. <http://scieur.org/interdisc-symp>
- Sense About Science (2009). *Peer review survey 2009: Preliminary findings*. Sense About Science. <https://senseaboutscience.org/activities/peer-review-survey-2009/>
- Shimp, C. (2004). Scientific peer review: A case study from local and global analyses. *Journal of the Experimental Analysis of Behavior*, 82(1), 103–116. <https://doi.org/10.1901/jeab.2004.82-103>
- Sîle, L., Pölönen, J., Sivertsen, G., Guns, R., Engels, T. C. E., Arefiev, P., Dušková, M., Faurbæk, L., Holl, A., Kulczycki, E., Macan, B., Nelhans, G., Petr, M., Pisk, M., Soós, S., Stojanovski, J., Stone, A., Šušol, J. & Teitelbaum, R. (2018) Comprehensiveness of national bibliographic databases for social sciences and humanities: findings from a European survey. *Research Evaluation*, 27(4), 310–322. <https://doi.org/10.1093/reseval/rvy016>
- Sivertsen, G. (2016). Patterns of internationalization and criteria for research assessment in the social sciences and humanities. *Scientometrics*, 107(2), 357–368. <https://doi.org/10.1007/s11192-016-1845-1>
- Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications*, 3, 17078. <https://doi.org/10.1057/palcomms.2017.78>

- Sivertsen, G. (2018). Balanced multilingualism in research. In *L'évaluation de la recherche scientifique : Enjeux, méthodes et instruments, Actes de la colloque International*, Rabat, 6–7 Decembre 2017 (pp. 88–102). <http://colloque.csefrs.ma/wp-content/uploads/2019/01/Actes-du-colloque-2017-6-11-2018.pdf>
- Sivertsen, G. (2018). The Norwegian Model in Norway. *Journal of Data and Information Science*, 3(4), 3–19. <https://doi.org/10.2478/jdis-2018-0017>
- Sivertsen, G., & Larsen, B. (2012). Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: an empirical analysis of the potential. *Scientometrics*, 91(2), 567–575. <https://doi.org/10.1007/s11192-011-0615-3>
- Smith, M. U., Wandersee, J. H., & Cummins, C. L. (1993). What's wrong with this manuscript? An analysis of the reasons for rejection given by Journal of Research in Science Teaching reviewers. *Journal of Research in Science Teaching*, 30(2), 209–211. <https://doi.org/10.1002/tea.3660300207>
- Sousa, S. B., & Brennan, J. L. (2014). The UK Research Excellence Framework and the transformation of research production. In C. Musselin and P. Teixeira (Eds.), *Reforming higher education* (pp. 65–80). Springer. https://doi.org/10.1007/978-94-007-7028-7_4
- Spezi, V., Wakeling, S., Pinfield, S., Fry, J., Creaser, C., & Willett, P. (2018). Let the community decide? The vision and reality of soundness-only peer review in open-access mega-journals. *Journal of Documentation*, 74(1), 137–161. <https://doi.org/10.1108/JD-06-2017-0092>
- Squazzoni, F., Brezis, E., & Marušić, A. (2017). Scientometrics of peer review. *Scientometrics*, 113(1), 501–502. <https://doi.org/10.1007/s11192-017-2518-4>
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, 19(3), 387–420. <https://doi.org/10.1177/030631289019003001>
- Tamblyn, R., Girard, N., Qian, C. J., & Hanley, J. (2018). Assessment of potential bias in research grant peer review in Canada. *Canadian Medical Association Journal*, 190(16), E489–E499. <https://doi.org/10.1503/cmaj.170901>
- Teixeira da Silva, J. A. (2019). Challenges to open peer review. *Online Information Review*, 43(2), 197–200. <https://doi.org/10.1108/OIR-04-2018-0139>
- Tennant, J. P. (2017). What are the barriers to post-publication peer review? *LSE Impact Blog*. <https://blogs.lse.ac.uk/impactofsocialsciences/2017/04/12/what-are-the-barriers-to-post-publication-peer-review/>
- Tennant, J. P. (2018). The state of the art in peer review. *FEMS Microbiology Letters*, 365(19), fny204. <https://doi.org/10.1093/femsle/fny204>
- Tennant, J. P., Crane, H., Crick, T., Davila, J., Enkhbayar, A., Havemann, J., Kramer, B., Martin, R., Masuzzo, P., Nobes, A., Rice, C., Rivera-López, B., Ross-Hellauer, T.,

- Sattler, S., Thacker, P. D., & Vanholsbeeck, M. (2019). Ten hot topics around scholarly publishing. *Publications*, 7(2), 34. <https://doi.org/10.3390/publications7020034>
- Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., Elkhatib, Y., B. Collister, L., Pikas, C. K., Crick, T., Masuzzo, P., Caravaggi, A., Berg, D. R., Niemeyer, K. E., Ross-Hellauer, T., Mannheimer, S., Rigling, L., Katz, D. S., Greshake Tzovaras, B., ... Colomb, J. (2017). A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Research*, 6, 1151. <https://doi.org/10.12688/f1000research.12037.3>
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS ONE*, 8(5), e64841. <https://doi.org/10.1371/journal.pone.0064841>
- Thomas, J. P., & Lawrence, T. S. (1991). Common deficiencies of NIDRR research applications. *American Journal of Physical Medicine & Rehabilitation*, 70(1), 161–164. <https://doi.org/10.1097/00002060-199004000-00005>
- Thorngate, W., Dawes, R. M., & Foddy, M. (2009). *Judging merit*. Psychology Press.
- Turcotte, C., Drolet, P., & Girard, M. (2004). Study design, originality and overall consistency influence acceptance or rejection of manuscripts submitted to the Journal. *Canadian Journal of Anaesthesia-Journal Canadien d'anesthésie*, 51(6), 549–556. <https://doi.org/10.1007/BF03018396>
- Tversky, A., & Kahnemann, D. (1974). Judgement under uncertainty: heuristics and biases. *Science*, 185, 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- van Arensbergen, P., van der Weijden, I., & van den Besselaar, P. (2014). Different views on scholarly talent: What are the talents we are looking for in science? *Research Evaluation*, 23(4), 273–284. <https://doi.org/10.1093/reseval/rvu015>
- van den Akker, W. (2016). Yes we should; Research assessment in the humanities. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research assessment in the humanities. Towards criteria and procedures* (pp. 23–29). Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_3
- van den Besselaar, P., & Sandström, U. (2015). Gender differences in research performance and its impact on careers: a longitudinal case study. *Scientometrics*, 106(1), 143–162. <http://doi.org/10.1007/s11192-015-1775-3>
- van den Brink, M., Brouns, M., & Waslander, S. (2006). Does excellence have a gender? A national research study on recruitment and gender. *Employee Relations*, 28(6), 523–539. <https://doi.org/10.1108/01425450610704470>
- van Gestel, R., & Vranken, J. (2011). Assessing legal research: Sense and nonsense of peer review versus bibliometrics and the need for a European approach. *German Law Journal*, 12(3), 901–929. <https://doi.org/10.1017/S2071832200017144>
- Vanholsbeeck, M. (2017). La notion de Science Ouverte dans l’Espace européen de la recherche: Entre tendances à l’«exotérisation» et à la «gestionnarisation» de la

- recherche scientifique. *Revue française des sciences de l'information et de la communication*, 11. <https://doi.org/10.4000/rfsic.3241>
- van Leeuwen, T. N., van Wijk, E., & Wouters, P. F. (2016) Bibliometric analysis of output and impact based on CRIS data: A case study on the registered output of a Dutch university. *Scientometrics*, 106(1), 1–16. <https://doi.org/10.1007/s11192-015-1788-y>
- van Lent, M., IntHout, J., & Out, H. J. (2015). Peer review comments on drug trials submitted to medical journals differ depending on sponsorship, results and acceptance: A retrospective cohort study. *BMJ Open*, 5(9), e007961. <https://doi.org/10.1136/bmjopen-2015-007961>
- van Noorden, R. (2015). Interdisciplinary research by the numbers. *Nature*, 525(7569), 306–307. <https://doi.org/10.1038/525306a>
- Verleysen, F. T., & Engels, T. C. E. (2013). A label for peer-reviewed books. *Journal of the American Society for Information Science and Technology*, 64(2), 428–430. <https://doi.org/10.1002/asi.22836>
- Verleysen, F. T., & Engels, T. C. E. (2015). *ERIH Plus in 2014: Stand van zaken en tijdschriftselectie vergeleken met het VABB-SHW*. ECOOM.
- Verleysen, F. T., Ghesquière, P., & Engels, T. C. E. (2014). The objectives, design and selection process of the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW). In W. Blockmans, L. Engwall & D. Weaire (Eds.), *Bibliometrics: use and abuse in the review of research performance* (pp. 115–125). Portland Press.
- The Volkswagen Foundation. (2014). *What is intellectual quality in the humanities? Some guidelines*. <https://www.volkswagenstiftung.de/en/news-press/publications/what-is-intellectual-quality-in-the-humanities-some-guidelines>
- 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>
- VSNU, NWO, KNAW. (2016). *Standard Evaluation Protocol 2015–2021. Protocol for Research Assessments in the Netherlands*. VSNU, NWO and KNAW. <https://www.vsnunl.nl/files/documenten/Domeinen/Onderzoek/SEP2015-2021.pdf>
- Waaijer, C. J., Sonneveld, H., Buitendijk, S. E., van Bochove, C. A., & van der Weijden, I. C. (2016). The role of gender in the employment, career perception and research performance of recent PhD graduates from Dutch universities. *PloS ONE*, 11(10), e0164784. <https://doi.org/10.1371/journal.pone.0164784>
- Ware, M. (2008). Peer review in scholarly journals: Perspective of the scholarly community – Results from an international study. *Information Services & Use*, 28(2), 109–112. <https://doi.org/10.3233/ISU-2008-0568>

- Ware, M., & Monkman, M. (2008). *Peer review in scholarly journals: Perspective of the scholarly community—an international study*. Publishing Research Consortium. <http://www.publishingresearch.net/PeerReview.htm>
- Watermeyer, R. (2016). Impact in the REF: Issues and obstacles. *Studies in Higher Education*, 41(2), 199–214. <https://doi.org/10.1080/03075079.2014.915303>
- Watermeyer, R., & Chubb, J. (2019). Evaluating ‘impact’ in the UK’s Research Excellence Framework (REF): Liminality, looseness and new modalities of scholarly distinction. *Studies in Higher Education*, 44(9), 1554–1566. <https://doi.org/10.1080/03075079.2018.1455082>
- Watermeyer, R., & Hedgecoe, A. (2016). Selling ‘impact’: Peer reviewer projections of what is needed and what counts in REF impact case studies. A retrospective analysis. *Journal of Education Policy*, 31(5), 651–665. <https://doi.org/10.1080/02680939.2016.1170885>
- Wennerås, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, 387, 341–343. <https://doi.org/10.1038/387341a0>
- Whaley, A. L., Rodriguez, R., & Alexander, L. A. (2006). Development of a rating form to evaluate grant applications to the Hogg Foundation for Mental Health. *Evaluation Review*, 30(1), 3–26. <https://doi.org/10.1177/0193841X05275586>
- Wissenschaftsrat. (2004). *Recommendations for rankings in the system of higher education and research*. Wissenschaftsrat. https://www.wissenschaftsrat.de/download/archiv/6285-04-engl.pdf?__blob=publicationFile&v=2
- Wissenschaftsrat. (2010). *Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften* [Recommendations for comparative research assessment in the humanities]. Wissenschaftsrat. <https://www.wissenschaftsrat.de/download/archiv/10039-10.html> (for an English summary see Hornung et al., 2016).
- 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>
- Wroblewska, M. N. (2019). *Impact evaluation in Norway and in the UK: A comparative study, based on REF 2014 and Humeval 2015-2017*. ENRESSH. <https://doi.org/10.3990/4.2666-0776.2019.01>

- Yan, E. (2016). Disciplinary knowledge production and diffusion in science. *Journal of the Association for Information Science and Technology*, 67(9), 2223–2245. <https://doi.org/10.1002/asi.23541>
- Yarbrough, C., & Whitaker, J. A. (2009). Analysis of reviewer comments about quantitative manuscripts accepted by the Journal of Research in Music Education. *Journal of Research in Music Education*, 56(4), 287–292. <https://doi.org/10.1177/0022429408329448>
- Zacharewicz, T., Lepori, B., Reale, E., & Jonkers, K. (2018). Performance-based research funding in EU Member States – a comparative assessment. *Science and Public Policy*, 46(1), 105–115. <https://doi.org/10.1093/scipol/scy041>
- Zaruba, K. E., Toma, J. D., & Stark, J. S. (1996). Criteria used for qualitative research in the refereeing process. *Review of Higher Education*, 19(4), 435–460. <https://doi.org/10.1353/rhe.1996.0017>