ENRESSH Work Group 1

Overview of Peer Review Practices

in the SSH

ENRESSH Interim Report

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
Michael Ochsner, FORS, Lausanne and ETH Zurich, Switzerland

With chapters by Gemma Derrick, Tim C. E. Engels, Raf Guns, Nina Kancewicz-Hoffman, Karolina Lendák-Kabók, Michael Ochsner, Janne Pölönen, Tony Ross-Hellauer, Marc Vanholsbeeck

And chapters previewed for the next version by: Mirjam Aeschbach, Elea Giménez-Toledo, Marek Hołowiecki, Sven E. Hug, Lai Ma, Jorge Mañana-Rodríguez, Ginevra Peruginelli, Elías Sanz-Casado, Jadranka Stojanovski

Chapters reviewed by Elea Giménez-Toledo, Jon Holm, Linda Sīle, Mimi Urbanc

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

Grant period to which the report/ deliverable is related: Grant Period 3

Version 1.1, May 2019

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 of 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 Open Science turn. To this, the general challenges of peer review are added, such as the risk of gender bias, conservative bias, work load for all parties involved.

While this interim report cannot fully sketch out the challenges and opportunities as well as recommendations for peer review in the SSH, it serves to identify and describe some of the major challenges to peer review in the SSH. It concludes that the SSH disciplines should develop their own ways to adequately evaluate their research. 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 by large.

The report proposes an approach for achieving such a goal: 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. Furthermore, reviewing should be part of academic education.

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 Galleron, Aldis Gedutis, Haris Gekić, Elea Giménez-Toledo, Agnė Girkontaitė, Raf Guns, Jon Holm, Marek Holowiecki, Todor Hristov Dechev, Dragan Ivanović, Nina Kancewicz-Hoffman, Andreja Isenič Starčič, Arnis Kokorevics, Litauras Kraniauskas, 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

Table of Content

Executive summary 2
Acknowledgements 3
PART I: General Framework: State of the art of peer review in the SSH in Europe 5
Introduction: Aim and Scope of the Report 5
Peer Review in SSH: In Need of Development? 8
PART II: Issues and Discussions Specific to and Most Relevant for SSH Peer Review 13
Evaluation Criteria and Methodology 13
PART III: Guidelines, Procedures and Formal Criteria Versus their Practical Application 20
Ambiguity in identification of scholarly peer-reviewed publications 20
PART IV: Current Challenges for Peer Reviewing: Towards More Open and Gender-Sensitive Peer Reviewing Practices in the SSH 25
Peer review in the context of the new modes of knowledge production, dissemination and evaluation 25
The Perception of Senior Sociologists Towards Peer Reviewing in the Context of the Current Changes in the SSH Assessment Systems 27
PART V: Conclusion 32
Bibliography: 34

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 and Michael Ochsner

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, Ochsner & Daniel, 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, Lamont & Guetzkow, 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. Three main drivers are behind this return of peer review as the preferred means of assessment in different evaluation situations after years of fascination with bibliometrics and quantitative methods: Bibliometrics and other quantitative methods have proved inadequate for the evaluation of SSH research (Hicks, 2004; Nederhof, 2006; Ochsner, Hug & Daniel, 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. It will sketch the topics addressed in the subgroup of Work Group 1 that investigates peer review processes and thus gives an overview of what the final report will address.

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 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 will focus on those aspects of evaluation which are related to peer review.

References

Burrows, R. (2012). Living with the h-index? Metric assemblages in the contemporary academy. *The Sociological Review, 60*(2), 355–372. doi: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.

Hicks, D. (2004). The four literatures of social science. In H. F. Moed, W. Glanzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 473–496). Dordrecht: Kluwer Academic.

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 of Education Law and Policy, 10*(1), 55–68.

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). Cham: Springer International Publishing. doi:10.1007/978-3-319-29016-4_12

Lack, E. (2008). Einleitung - Das Zauberwort "Standards". In E. Lack & C. Markschies (Eds.), *What the hell is quality? Qualitätsstandards in den Geisteswissenschaften* (pp. 9-34). Frankfurt: Campus.

Lawrence, P. A. (2002). Rank injustice. *Nature, 415*(6874), 835–836. doi:10.1038/415835a

Mallard, G., Lamont, M., & Guetzkow, J. (2009). Fairness as Appropriateness: Negotiating Epistemological Differences in Peer Review. *Science, Technology & Human Values, 34*(5), 573–606. doi:10.1177/0162243908329381

MacRoberts, M. H., & MacRoberts, B. R. (2017). The mismeasure of science: Citation analysis. *Journal of the Association for Information Science and Technology, 10*, 646–9. doi:10.1002/asi.23970

Molinié, A., & Bodenhausen, G. (2010). Bibliometrics as Weapons of Mass Citation. *CHIMIA International Journal for Chemistry, 64*(1), 78–89. doi: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. doi: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.

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). Cham: Springer International Publishing. doi:10.1007/978-3-319-29016-4_3

Peer Review in SSH: In Need of Development?

By Gemma Derrick and 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 is 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 (Ross-Hellauer & Derrick, 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. As stated in our introduction, 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 (Cornelius, 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, Engels, & Sivertsen, 2012; Ossenblok, Verleysen, & Engels, 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 consider the extent to which extended rejection, resubmission 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. 2019). 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, Deppe, & Schmidt, 2017; Rowley & Sbaffi, 2018; Ware, 2008). Rowley & 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 el al. (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 (Ross-Hellauer & Derrick, 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. doi:10.1108/03068290410518247

Björk, B.-C., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. Journal of Informetrics, 7(4), 914–923. doi:10.1016/j.joi.2013.09.001

British Academy. (2007). *Peer Review: the challenges for the humanities and social sciences. A British Academy Report*. Retrieved from https://www.thebritishacademy.ac.uk/publications/peer-review-challenges-humanities-and-social-sciences

Ceci, S. J., & Peters, D. P. (1982). Peer Review: A Study of Reliability. *Change: The Magazine of Higher Learning, 14*(6), 44–48. doi:10.1080/00091383.1982.10569910

Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences, 14*(1), 119–135. doi:10.1017/S0140525X00065675

Cornelius, J. L. (2012). Reviewing the review process: Identifying sources of delay. *The Australasian Medical Journal, 5*(1), 26–29. doi:10.4066/AMJ.2012.1165

Derricourt, R. (2012). Peer Review: Fetishes, Fallacies, and Perceptions. *Journal of Scholarly Publishing, 43*(2), 137–147. doi:10.1353/scp.2012.0000

Flyvbjerg, B. (2001) *Making social science matter: Why social inquiry fails and how it can succeed again*. Cambridge: Cambridge University Press.

Haffar, S., Bazerbachi, F., & Murad, M. H. (2019). Peer Review Bias: A Critical Review. *Mayo Clinic Proceedings, 94*(4), 670–676. doi:10.1016/j.mayocp.2018.09.004

Huisman, J., & Smits, J. (2017). Duration and quality of the peer review process: the author's perspective. *Scientometrics, 113*(1), 633–650. doi:10.1007/s11192-017-2310-5

Kratz, J. E., & Strasser, C. (2015). Researcher Perspectives on Publication and Peer Review of Data. *PLOS One, 10*(2), e0117619. doi:10.1371/journal.pone.0117619

Kulczycki, E., Engels, T. C. E., Pölönen, J., Bruun, K., Dušková, M., Guns, R., … Zuccala, A. (2018). Publication patterns in the social sciences and humanities: evidence from

eight European countries. *Scientometrics, 116*(1), 463–486. doi:10.1007/s11192-018-2711-0

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. doi:10.1002/asi.22784

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research, 1*(2), 161–175. doi:10.1007/BF01173636

Mallard, G., Lamont, M., & Guetzkow, J. (2009). Fairness as Appropriateness: Negotiating Epistemological Differences in Peer Review. *Science, Technology, & Human Values, 34*(5), 573–606. doi: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. doi: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. doi:10.1093/reseval/rvs019

Ossenblok, T. 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). *Journal of the Association for Information Science and Technology, 65*(5), 882–897. doi:10.1002/asi.23015

Pautasso, M., & Schäfer, H. (2010). Peer review delay and selectivity in ecology journals. *Scientometrics, 84*(2), 307–315. doi:10.1007/s11192-009-0105-z

Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research, 6*, 588. doi: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. doi:10.1371/journal.pone.0189311

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. doi:10.1177/0165551517740821

Shimp, C. P. (2004). Scientific Peer Review: A Case Study from Local and Global Analyses. *Journal of the Experimental Analysis of Behavior, 82*(1), 103–116. doi:10.1901/jeab.2004.82-103

Van Noorden, R. (2015). Interdisciplinary research by the numbers. *Nature, 525*(7569), 306-307. doi:10.1038/525306a

Van Rooyen, S., Black, N., & Godlee, F. (1999). Development of the Review Quality Instrument (RQI) for Assessing Peer Reviews of Manuscripts. *Journal of Clinical Epidemiology, 52*(7), 625–629. doi:10.1016/S0895-4356(99)00047-5

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. doi: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.

Yan, E. (2016). Disciplinary knowledge production and diffusion in science. *Journal of the Association for Information Science and Technology, 67*(9), 2223-2245.

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, Mutz & Daniel, 2008; Bornmann, Mutz, Marx, Schier & Daniel, 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, Galleron & Ionescu, 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, 1962) 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, multidimensional construct (see, e.g., Bazeley, 2010; Hemlin, 1993; Hug, Ochsner & Daniel, 2013; Ochsner, Hug & Daniel, 2013). Reducing research quality to some aspects might lead to adverse effects, such as goal displacement (see, e.g., Butler, 2007; De Rijke, Wouters, Rushforth, Franssen & Hammarfelt, 2016; Sousa & Brennan, 2014) or task reduction (De Rijke, Wouters, Rushforth, Franssen & Hammarfelt, 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, 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., 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, Laberge & McGuire, 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, Tejada-Artigas & Manana-Rodriguez, 2013; Gogolin & Stumm, 2014; Ochsner, Hug & Daniel, 2016; for an overview Ochsner, Hug & Galleron, 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, Hug & Daniel, 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, Mutz & Daniel, 2008; 2010; Mutz, Bornmann & Daniel, 2014; Tamblyn, Girard, Qian & Hanley, 2018). However, there are several methodological issues involved with these criticisms of the peer review process (see also Langfeldt, Bloch & Sivertsen, 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 *intra*rater 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, Dawes and Foddy (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, Hug & Daniel, 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. doi:10.1007/s10734-012-9519-2

Bazeley, P. (2010). Conceptualising research performance. *Studies in Higher Education, 35*(8), 889–903. doi: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. http://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.

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. doi: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. http://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, 174*(4), 857–879. doi: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.

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. doi:10.1073/pnas.1014871108

Daniel, H.-D. (1993). *Guardians of Science: Fairness and Reliability of Peer Review*. Weinheim: VCH Verlagsgesellschaft.

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. doi: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. doi:10.1057/palcomms.2017.40

Enserink, M. (2015). Dutch sexism study comes under fire. *Science, 364*(6446). doi:10.1126/science.aad4618

Giménez-Toledo, E., Tejada-Artigas, C., & Manana-Rodriguez, J. (2013). Evaluation of scientific books' publishers in social sciences and humanities: Results of a survey. *Research Evaluation, 22*(1), 64–77. http://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. doi:10.3917/socio.073.0261

Hemlin, S. (1993). Scientific quality in the eyes of the scientist. A questionnaire study. *Scientometrics, 27*(1), 3–18.

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. doi: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. doi:10.1111/j.1468-2273.2008.00378.x

KNAW. (2011). *Quality Indicators for Research in the Humanities*. Amsterdam: KNAW. Royal Netherlands Academy of Arts and Sciences.

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). Cham: Springer International Publishing. doi:10.1007/978-3-319-29016-4_13

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

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. doi: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. (2014). 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. J*ournal of the Association for Information Science and Technology, 66*(11), 2321–2339. http://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). Retrieved from: https://www.uantwerpen.be/images/uantwerpen/container41447/files/A5-BOOK_RESSH2017_170717-interactive.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.

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. doi: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. doi: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*. Cham: Springer International Publishing. http://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: Wuhan University.

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. Retrieved from: 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. http://doi.org/10.1057/palcomms.2017.20

Polanyi, M. (1967). *The tacit dimension*. London: 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). Dordrecht: Springer.

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. http://doi.org/10.1503/cmaj.170901

Tversky, A., & Kahnemann, D. (1974). Judgement under uncertainty: heuristics and biases. *Science, 185*, 1124–1131.

VSNU, NWO, KNAW. (2016). *Standard Evaluation Protocol 2015 - 2021. Protocol for Research Assessments in the Netherlands*. Voorburg: VSNU, NWO and KNAW.

Wissenschaftsrat. (2004). *Empfehlungen zu Rankings im Wissenschaftssystem*. Hamburg: Wissenschaftsrat.

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

Ambiguity in identification of scholarly peer-reviewed publications

By Janne Pölönen, Tim C. E. Engels and 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, 1991; Burnhill & Tubby-Hille, 2003; Verleysen & Engels, 2015) or book publisher (Verleysen, Ghesquière & Engels, 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., 2018; Pölönen, Engels & Guns, 2018).

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 3596 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. 2018; Pölönen, Engels & Guns, 2018).

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 (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.

ALLEA (2017). *The European Code of Conduct for Research Integrity*. Berlin: 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.

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.

British Academy. (2007). *Peer Review: the challenges for the humanities and social sciences. A British Academy Report*. Retrieved from 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.

Dahler-Larsen, P. (2019). *Quality: From Plato to Performance*. Palgrave Macmillan.

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.

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.

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.

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.

Pölönen, J. (2018). Applications of, and Experiences with, the Norwegian Model in Finland. *Journal of Data and Information Science, 3*(4), 31–44.

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.

Pölönen, J., Engels, T.C.E., & Guns, R. (forthcoming). Ambiguity in identification of peer-reviewed publications in the Finish and Flemish performance-based research funding systems. *Science and Public Policy*.

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.

Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications*, *3*, 17078.

Sivertsen, G. (2018). The Norwegian Model in Norway. *Journal of Data and Information Science, 3*(4), 3–19.

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. doi:10.1002/asi.22836

Verleysen, F.T., & Engels, T.C.E. (2015). *ERIH Plus in 2014: Stand van zaken en tij-dschriftselectie vergeleken met het VABB-SHW*. Antwerpen: 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). London: 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.

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

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

By Marc Vanholsbeeck and Karolina Lendák-Kabók

Introduction

There has been a tendency since the 1980s to "exoterise" knowledge production, dissemination and evaluation, i.e. to open them outside of the "esoteric" circles of the disciplinary peers who traditionally produce fundamental research (Vanholsbeeck, 2017). The concept of exoterisation of research goes beyond the linearity of knowledge transfer, but also beyond marketization, since it includes social innovation – which can take non-commercial forms –, and relates to the notion of knowledge society rather than to the sole knowledge economy.

In some cases, such as in recently institutionalized disciplines in Social Sciences and Humanities (SSH) like Communication Studies, interdisciplinarity 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 – 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 exoterisation of research. Worth mention 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 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" (2014), as well as the notion of "societal challenges" which constitutes one of the three pillars of the current Horizon 2020 Framework Programme (2014-2020). 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 agenda[1] – including open access to publications, open research data and citizen science – and the impact related policies – the so-called "impact agenda" of the EU Commission – also align with and foster this tendency to exoterization.

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 bring its contribution to the resolution of societal challenges. According to a recent monitoring report from the Commission though (EU Commission, 2018), it 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 to engage SSH researchers from diverse disciplinary backgrounds into this new type of peer reviewing of impact driven research projects. This supposes that they should be provided with the right skills, learning to debate and argue with peers from other disciplinary horizons and, in some cases, with no knowledge of SSH specific epistemologies and methods, and being able to assess impact.

SSH scholars, while they peer review together with colleagues from other disciplines or with stakeholders external to academia, may also lack a "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. As such, SSH peer reviewers may soon be faced, rather than with the egalitarian perspective that policy makers' discourses on interdisciplinary partnerships most often take for granted, with the diverse levels of prestige that are attached to the various disciplines inside academia, as well as – more generally – the level of legitimacy that STEM colleagues are ready to allow to their colleagues from the SSH fields. This question of legitimate interpretation concerns also SSH scholars' relations with non-academic actors potentially engaged in the review. Indeed, as J. 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, SSH researchers engaged in interdisciplinary evaluation may have to peer review topics for which disciplines adjacent to theirs are concerned. Ideally, SSH disciplinary

[1] https://ec.europa.eu/research/openscience/pdf/draft_european_open_science_agenda.pdf#view=fit&pagemode=none, last accessed on 20 October 2018.

expertise in peer reviewing should reflect the diversity of the individual SSH disciplines engaged in the interdisciplinary project to evaluate. In any case, policy makers and funders should avoid referring to the notions of "SSH expert" and "SSH expertise". 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.

References

Collins, H.M., & Evans, R. (2007). *Rethinking Expertise*. Chicago, IL: University of Chicago Press.

EU Commission (DG RTD) (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*. Luxembourg: Publications Office of the European Union.

Kattel, R., & Mazzucato, M. (2018). Mission-oriented innovation policy and dynamic capabilities in the public sector. *Industrial and Corporate Change, 27*(5), 787–801. doi:10.1093/icc/dty032

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.wixstatic.com/ugd/48bd50_66b801501d204ed3a123b84b947ac823.pdf

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.

The Perception of Senior Sociologists Towards Peer Reviewing in the Context of the Current Changes in the SSH Assessment Systems

By Marc Vanholsbeeck

Introduction

In her investigation of the changes affecting research evaluation in mathematics, physics, astronomy, biology, chemistry, physics, history, letters, law, management and economics, Lefèbvre underlined that researchers still consider traditional peer review as constructive and a source of improvement of the original article. On the contrary, more participatory online approaches and Open Peer Reviewing were considered with some caution - as potentially interesting additions to the traditional peer review but not as replacement - since it could be subject to interpersonal biases and non-scientific arguments (Lefèbvre, 2006).

In 2008, a major cross-disciplinary international study surveyed the perception of 3,040 academics, at an international level, as regards the peer review in scholarly journals (Ware, 2008). The main conclusions of this research were that peer review is considered as very important (93% of respondents disagree with the statement that peer review would not be necessary) and that peer review improves the quality of articles published (according to 90% of respondents). However, there is a desire to improve the procedure (slow and over-worked referees). Respondents prefer the double-blind review (56% of respondents), which is considered as the most effective (by 71% of respondents). However double-blind peer review is not deemed as being free from shortcomings such as, in particular, the possibility that even anonymous authors are identified. Peer review after publication is considered effective by 37% of the respondents but mostly as a complement rather than an alternative to the traditional peer review. There is no strong support from respondents for replacing the peer review with metrics or for Open Peer Reviewing (26% of respondents in favour of OPR), while respondents show ambivalent attitudes to peer review of data, which is considered by a majority to be desirable but difficult to incorporate into the peer review of articles. Finally, respondents are more likely to be against than in favour of a financial reward for the referees.

Later studies on peer review in the SSH have confirmed that researchers in these disciplines consider peer review as an important guarantee of quality in scholarly publications (Stieg Dalton, 2009; Harley et al, 2010; Albert et al., 2012). The 2015 STM report (Ware and Mabe, 2015) reaches the same conclusion, taking a broad diversity of SSH and STEM disciplines into account. Some evolutionary innovations in peer review such as the disclosure of reviewers' names or the publication of review reports become more common, but post-publication review is not strongly supported, at least as a replacement for traditional peer review. The 2018 edition of the STM report confirms that the "soundness not significance" peer review criterion adopted by open access "megajournals" like PLOS ONE is now well-established, contrarily to other innovations such as journal independent ("portable") peer review or post-publication review which still receive only limited support (Johnson et al., 2018).

Looking at the different threats menacing peer review, a survey conducted by Elsevier (Mulligan, 2005) with six focus groups of 59 respondents had revealed some more negative views: possibilities of biases and interpersonal conflicts, risk of plagiarism, increasing difficulty to recruit quality referees, lack of rewards for referees or lack of confidence in the scientific relevance of open comments online. Similarly, a UK Research Information Network (RIN) survey showed that peer review is criticized for several reasons such as: slow process, lack of ability to detect fraud, risk of bias and subjectivity in judgments, tendency to conservatism, discouragement of innovative and interdisciplinary research, cost of the system, lack of training of referees, risk of overwork linked to refereeing (RIN, 2010, p. 8).

The first cross-disciplinary survey dedicated to academics' attitudes towards Open Peer Review has been conducted for the OpenAIRE2020 project during September and October 2016 (Ross-Hellauer et al., 2017) and confirms a progressively more positive attitude of researchers towards open peer reviewing. The results show indeed that the majority (60.3%) of respondents consider that the general concept of OPR should be mainstreamed as a scholarly practice. Respondents also show high levels of experience with Open Peer

Review, as author, reviewer or editor. Respondents support open interaction, open reports and final-version commenting but open identities peer review was not generally favoured. Significantly, the study emphasises important variations in peer review systems across disciplines, warning against any "one size fits all" approach in (open) peer reviewing.

Methodology and Sample

In the context of a subgroup of the WG1[2], attitudes of senior SSH researchers towards research evaluation have been studied in 2018. The subgroup members conducted semi-structured interviews with sociologists 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 in native and/or state languages of the interviewees 16 scholars, male and female, about their perceived roles in the definition, the dissemination and the implementation of the quality criteria and rationales that are to be used in evaluation situations. In this report, respondents' perception of peer review in a changing evaluation context is presented.

Preliminary findings

Respondents' attitude towards peer review is to be put in the more general context of their perception of research evaluation as moving towards more consideration for internationalization and the use of quantitative indicators of performance. In some cases, peer review is considered as becoming ancillary to quantitative evaluation, but in others, performance indicators are, in some respondents' experience, to be included in the peer reviewing. 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: if they perceive clear assets linked to the prepublication traditional peer reviewing system, some researchers also consider the downsides of peer review, or even threats that may be linked to it.

On the one hand, peer review is considered by many as a top-quality criterion. It makes a clear-cut distinction between quality publications and – in the words of one male Icelandic sociologist – the other "small things" of lesser value that a sociologist produces, as well as between academics who are skilled in international publication – which relates to skills related to language, but also to how to manage peer review, as a reviewer and as a reviewed author – and the less educated others.

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 better their paper, and the reviewers to be updated on the last developments in certain fields. In small countries, involving international peer reviewers helps to overcome the problems inherent to small academic communities in which "everybody knows everybody" (Slovenian female sociologist). When international peer reviewers are involved, it can also encourage the internationalization of the SSH research national production as well: "If you

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

want to be funded you should publish internationally and this makes sense, actually a lot of sense to start publish 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 many as a potentially biased process. Anonymous 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 particularly 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-reviewing processes may have as an unwanted consequence that local contexts and societally relevant impact are not taken enough into account.

One respondent, Belgian male sociologist, goes as far as liking peer reviewing 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. only if you are not 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 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." (female English language professor from Cyprus)

Conclusions

Senior researchers recognize peer review as a fundamental building brick in the production, dissemination and evaluation system of research. But they certainly do not consider it as being free from any default. In particular, interpersonal relations between reviewed and reviewing researchers may introduce biases – particularly in small research communities –, including gender related ones, in the process. Peer review may also work as an instrument of domination and gate-keeping, on criteria that are sometimes not related to the epistemic quality of the contents to be assessed. It may also be an obstacle to the valorisation of more impact driven research.

In this perspective, the Open Science turn definitely opens new potentialities for more transparency and responsibility in the conduct of peer reviewing. But it also leads to new possible challenges still to be solved, in particular if more empirical evidence would in the future correlate non-anonymous peer review with gender biases, or if predatory publishers continue to threaten the reputation of Open Access publishing. This makes it necessary to pursue evidence and research-based monitoring of peer reviewing practices in the EU.

At the same time, new skills are also needed to assess and peer review interdisciplinary projects, making it urgent to educate researchers, and particularly early career investigators, in the conduct of more open, gender sensitive and interdisciplinary peer reviewing practices. It does not seem that SSH researchers are willing to get rid of peer review any time soon. Nevertheless, its effectiveness should not anymore be taken for granted, and any possible improvement has to be investigated, and whenever relevant be better integrated in SSH researchers' training and education.

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.

Harley, D., Acord, S.K, Earl-Novell, S., Lawrence, S., & King, C.J. (2010). *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines.* Berkeley: Center for Studies in Higher Education. https://cloudfront.escholarship.org/dist/prd/content/qt15x7385g/qt15x7385g.pdf

Johnson, R., Watkinson, A., & Wabe, M. (2018). *The STM report. An overview of scientific and scholarly publishing*. 5th edition. https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf

Lefèbvre, M. (2006). *Vers une remise en question des modalités traditionnelles d'évaluation des savoirs scientifiques ? Communication scientifique et valorisation de la recherche à l'heure d'Internet*, April 2006, Toulouse, Urfist Toulouse - SCD Toulouse1. https://archivesic.ccsd.cnrs.fr/sic_00112197

Mulligan, A. (2005). Is peer review in crisis? *Oral Oncology, 41*(2), 135–141.

RIN (Research Information Network) (2010). *Quality assurance and assessment of scholarly research. A guide for researchers, academic administrators and librarians*. London: RIN. http://www.rin.ac.uk/system/files/attachments/Quality_Assurance_screen_0.pdf

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.

Stieg Dalton, M. (2009). The publishing experiences of historians. In A.N. Greco (Ed.), *The state of scholarly publishing: challenges and opportunities* (pp. 107-146). New Brunswick: Transaction.

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.

Ware, M., & Mabe, M. (2015). *The STM report: An overview of scientific and scholarly journal publishing.* 4th edition. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1008&context=scholcom

PART V: Conclusion

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

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 of 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 documented in this report. 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 languages; challenges related to recent developments in research and its evaluation related to growing interdisciplinarity and Open Science turn. To this, the general challenges of peer review are added, such as the risk of gender bias, conservative bias, work load for all parties involved.

This report suggests ways for the SSH disciplines to respond to these challenges in a way that builds 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 by large.

The past has shown that automatically copying evaluation procedures from STEM disciplines did not always work out well. Therefore, Derrick and Ross-Hellauer 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. In the

following chapter Ochsner proposes an approach for achieving such a goal: 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. 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. Marc Vanholsbeeck and Karolina Lendák-Kabók discuss challenges of SSH integration to European impact-driven interdisciplinary research referring to the concept of exoterisation of research. Finally Marc Vanholsbeeck shows in his study on senior SSH scholars' attitudes towards research evaluation that although in their opinion peer review keeps its importance as an evaluation tool, at the same time they are aware of complexities of peer review practices in the context of current policies related to interdisciplinarity, internationalisation and societal impact. Vanholsbeeck recommends educating researchers in new skills needed in the conduct of more open, gender sensitive and interdisciplinary peer reviewing practices.

In this interim report, we were not yet able to fully sketch out the challenges and opportunities as well as recommendations for peer review in the SSH. Rather, this interim report serves to depict some issues of peer review in the SSH and some routes to take, which will be developed further in the final report. The final report will be structured in a similar way but include more chapters per part covering a number of relevant issues like definitions of impact in SSH and its evaluation, the use of English and national (local) languages in research and publications in the context of internationalisation of research; practices of peer review in Europe in different countries, for different research outputs and for different disciplines within SSH. Also challenges of gender equality and open science requirements will be treated in more depth.

Bibliography:

ENRESSH publications

Kulczycki, E., Engels, T. C. E., Pölönen, J., Bruun, K., Dušková, M., Guns, R., … Zuccala, A. (2018). Publication patterns in the social sciences and humanities: evidence from eight European countries. *Scientometrics, 116*(1), 463–486. doi:10.1007/s11192-018-2711-0

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.

Ochsner, M., & Dokmanović, M. (2017). Quality criteria and research obstacles in the SSH in Macedonia. Presentation at the conference "Research Evaluation for the Social Sciences and the Humanities (RESSH)" in Antwerp, 07.07.2017.

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. Retrieved from: 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. http://doi.org/10.1057/palcomms.2017.20

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*.

Pölönen, J., Engels, T.C.E., & Guns, R. (forthcoming). Ambiguity in identification of peer-reviewed publications in the Finish and Flemish performance-based research funding systems. *Science and Public Policy*.

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.

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.

ALLEA (2017). *The European Code of Conduct for Research Integrity*. Berlin: All European Academies.

Azar, O. H. (2004). Rejections and the importance of first response times. *International Journal of Social Economics, 31*(3), 259–274. doi:10.1108/03068290410518247

Baldwin, M. (2018). Scientific Autonomy, Public Accountability, and the Rise of "Peer Review" in the Cold War United States. *Isis, 109*(3), 538–558.

Bazeley, P. (2010). Conceptualising research performance. *Studies in Higher Education, 35*(8), 889–903. doi:10.1080/03075070903348404

Björk, B.-C., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. Journal of Informetrics, 7(4), 914–923. doi:10.1016/j.joi.2013.09.001

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. http://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.

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. doi: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. http://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, 174*(4), 857–879. doi:10.1111/j.1467-985X.2011.00689.x

British Academy. (2007). *Peer Review: the challenges for the humanities and social sciences. A British Academy Report*. Retrieved from https://www.thebritishacademy.ac.uk/publications/peer-review-challenges-humanities-and-social-sciences

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.

Burrows, R. (2012). Living with the h-index? Metric assemblages in the contemporary academy. *The Sociological Review, 60*(2), 355–372. doi: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.

Ceci, S. J., & Peters, D. P. (1982). Peer Review: A Study of Reliability. *Change: The Magazine of Higher Learning, 14*(6), 44–48. doi:10.1080/00091383.1982.10569910

Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences, 14*(1), 119–135. doi:10.1017/S0140525X00065675

Collins, H.M., & Evans, R. (2007). *Rethinking Expertise*. Chicago, IL: University of Chicago Press.

Cornelius, J. L. (2012). Reviewing the review process: Identifying sources of delay. *The Australasian Medical Journal, 5*(1), 26–29. doi:10.4066/AMJ.2012.1165

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.

Dahler-Larsen, P. (2019). *Quality: From Plato to Performance*. Palgrave Macmillan.

Daniel, H.-D. (1993). *Guardians of Science: Fairness and Reliability of Peer Review*. Weinheim: VCH Verlagsgesellschaft.

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. doi: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. doi:10.1057/palcomms.2017.40

Derricourt, R. (2012). Peer Review: Fetishes, Fallacies, and Perceptions. *Journal of Scholarly Publishing, 43*(2), 137–147. doi:10.1353/scp.2012.0000

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.

EU Commission (DG RTD) (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*. Luxembourg: Publications Office of the European Union.

Flyvbjerg, B. (2001) *Making social science matter: Why social inquiry fails and how it can succeed again*. Cambridge: Cambridge University Press.

Giménez-Toledo, E., Tejada-Artigas, C., & Manana-Rodriguez, J. (2013). Evaluation of scientific books' publishers in social sciences and humanities: Results of a survey. *Research Evaluation, 22*(1), 64–77. http://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. doi:10.3917/socio.073.0261

Haffar, S., Bazerbachi, F., & Murad, M. H. (2019). Peer Review Bias: A Critical Review. *Mayo Clinic Proceedings, 94*(4), 670–676. doi:10.1016/j.mayocp.2018.09.004

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.

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.

Harley, D., Acord, S.K, Earl-Novell, S., Lawrence, S., & King, C.J. (2010). *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines.* Berkeley: Center for Studies in Higher Education. https://cloudfront.escholarship.org/dist/prd/content/qt15x7385g/qt15x7385g.pdf

Hemlin, S. (1993). Scientific quality in the eyes of the scientist. A questionnaire study. *Scientometrics, 27*(1), 3–18.

Hicks, D. (2004). The four literatures of social science. In H. F. Moed, W. Glanzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 473–496). Dordrecht: Kluwer Academic.

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. doi: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 of Education Law and Policy, 10*(1), 55–68.

Huisman, J., & Smits, J. (2017). Duration and quality of the peer review process: the author's perspective. *Scientometrics, 113*(1), 633–650. doi:10.1007/s11192-017-2310-5

Johnson, R., Watkinson, A., & Wabe, M. (2018). *The STM report. An overview of scientific and scholarly publishing*. 5th edition. 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. doi:10.1111/j.1468-2273.2008.00378.x

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.

Kattel, R., & Mazzucato, M. (2018). Mission-oriented innovation policy and dynamic capabilities in the public sector. *Industrial and Corporate Change, 27*(5), 787–801. doi:10.1093/icc/dty032

KNAW. (2011). *Quality Indicators for Research in the Humanities*. Amsterdam: KNAW. Royal Netherlands Academy of Arts and Sciences.

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). Cham: Springer International Publishing. doi:10.1007/978-3-319-29016-4_12

Kratz, J. E., & Strasser, C. (2015). Researcher Perspectives on Publication and Peer Review of Data. *PLOS One, 10*(2), e0117619. doi:10.1371/journal.pone.0117619

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). Cham: Springer International Publishing. doi:10.1007/978-3-319-29016-4_13

Kulczycki, E., Engels, T. C. E., Pölönen, J., Bruun, K., Dušková, M., Guns, R., … Zuccala, A. (2018). Publication patterns in the social sciences and humanities: evidence from eight European countries. *Scientometrics, 116*(1), 463–486. doi:10.1007/s11192-018-2711-0

Lack, E. (2008). Einleitung - Das Zauberwort "Standards". In E. Lack & C. Markschies (Eds.), *What the hell is quality? Qualitätsstandards in den Geisteswissenschaften* (pp. 9-34). Frankfurt: Campus.

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

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. doi: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. doi:10.1038/415835a

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. doi:10.1002/asi.22784

Lefèbvre, M. (2006). *Vers une remise en question des modalités traditionnelles d'évaluation des savoirs scientifiques ? Communication scientifique et valorisation de la recherche à l'heure d'Internet*, April 2006, Toulouse, Urfist Toulouse - SCD Toulouse1. https://archivesic.ccsd.cnrs.fr/sic_00112197

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.wixstatic.com/ugd/48bd50_66b801501d204ed3a123b84b947ac823.pdf

MacRoberts, M. H., & MacRoberts, B. R. (2017). The mismeasure of science: Citation analysis. *Journal of the Association for Information Science and Technology, 10*, 646–9. doi:10.1002/asi.23970

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research, 1*(2), 161–175. doi:10.1007/BF01173636

Mallard, G., Lamont, M., & Guetzkow, J. (2009). Fairness as Appropriateness: Negotiating Epistemological Differences in Peer Review. *Science, Technology & Human Values, 34*(5), 573–606. doi: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.

Molinié, A., & Bodenhausen, G. (2010). Bibliometrics as Weapons of Mass Citation. *CHIMIA International Journal for Chemistry, 64*(1), 78–89. doi:10.2533/chimia.2010.78

Mulligan, A. (2005). Is peer review in crisis? *Oral Oncology, 41*(2), 135–141.

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. doi:10.1002/asi.22798

Mutz, R. D., Bornmann, L., & Daniel, H.-D. (2014). 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. J*ournal of the Association for Information Science and Technology, 66*(11), 2321–2339. http://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. doi:10.1007/s11192-006-0007-2

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). Retrieved from: https://www.uantwerpen.be/images/uantwerpen/container41447/files/A5-BOOK_RESSH2017_170717-interactive.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.

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. doi: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. doi: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*. Cham: Springer International Publishing. http://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: Wuhan University.

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. Retrieved from: 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. http://doi.org/10.1057/palcomms.2017.20

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. doi:10.1093/reseval/rvs019

Ossenblok, T. 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). *Journal of the Association for Information Science and Technology, 65*(5), 882–897. doi:10.1002/asi.23015

Pautasso, M., & Schäfer, H. (2010). Peer review delay and selectivity in ecology journals. *Scientometrics, 84*(2), 307–315. doi:10.1007/s11192-009-0105-z

Polanyi, M. (1967). *The tacit dimension*. London: Routledge & Kegan Paul.

Pölönen, J. (2018). Applications of, and Experiences with, the Norwegian Model in Finland. *Journal of Data and Information Science, 3*(4), 31–44.

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*.

Pölönen, J., Engels, T.C.E., & Guns, R. (forthcoming). Ambiguity in identification of peer-reviewed publications in the Finish and Flemish performance-based research funding systems. *Science and Public Policy*.

RIN (Research Information Network) (2010). *Quality assurance and assessment of scholarly research. A guide for researchers, academic administrators and librarians*.

London: RIN. http://www.rin.ac.uk/system/files/attachments/Quality_Assurance_screen_0.pdf

Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research, 6*, 588. doi: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.

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. doi:10.1177/0165551517740821

Shimp, C. P. (2004). Scientific Peer Review: A Case Study from Local and Global Analyses. *Journal of the Experimental Analysis of Behavior, 82*(1), 103–116. doi:10.1901/jeab.2004.82-103

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.

Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications*, *3*, 17078.

Sivertsen, G. (2018). The Norwegian Model in Norway. *Journal of Data and Information Science, 3*(4), 3–19.

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). Dordrecht: Springer.

Stieg Dalton, M. (2009). The publishing experiences of historians. In A.N. Greco (Ed.), *The state of scholarly publishing: challenges and opportunities* (pp. 107-146). New Brunswick: Transaction.

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. http://doi.org/10.1503/cmaj.170901

Tversky, A., & Kahnemann, D. (1974). Judgement under uncertainty: heuristics and biases. *Science, 185*, 1124–1131.

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

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). Cham: Springer International Publishing. doi:10.1007/978-3-319-29016-4_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.

Van Noorden, R. (2015). Interdisciplinary research by the numbers. *Nature, 525*(7569), 306-307. doi:10.1038/525306a

Van Rooyen, S., Black, N., & Godlee, F. (1999). Development of the Review Quality Instrument (RQI) for Assessing Peer Reviews of Manuscripts. *Journal of Clinical Epidemiology, 52*(7), 625–629. doi:10.1016/S0895-4356(99)00047-5

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. doi: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*. Antwerpen: 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). London: Portland Press.

VSNU, NWO, KNAW. (2016). *Standard Evaluation Protocol 2015 - 2021. Protocol for Research Assessments in the Netherlands*. Voorburg: VSNU, NWO and KNAW.

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.

Ware, M., & Mabe, M. (2015). *The STM report: An overview of scientific and scholarly journal publishing.* 4th edition. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1008&context=scholcom

Wissenschaftsrat. (2004). *Empfehlungen zu Rankings im Wissenschaftssystem*. Hamburg: Wissenschaftsrat.

Yan, E. (2016). Disciplinary knowledge production and diffusion in science. *Journal of the Association for Information Science and Technology, 67*(9), 2223-2245.

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.