

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

Evaluation Criteria and Methodology

By Michael Ochsner

Introduction

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

What is Quality?

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

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

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

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

Assessment of Research by Peer Review

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

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

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

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

Conclusions

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

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

References

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

- de Rijcke, S., Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use – a literature review. *Research Evaluation*, 25(2), 161–169. <https://doi.org/10.1093/reseval/rvv038>
- Derrick, G., & Samuel, G. (2017). The future of societal impact assessment using peer review: pre-evaluation training, consensus building and inter-reviewer reliability. *Palgrave Communications*, 3, palcomms201740. <https://doi.org/10.1057/pal-comms.2017.40>
- Enserink, M. (2015). Dutch sexism study comes under fire. *Science*, 364(6446). <https://doi.org/10.1126/science.aad4618>
- Giménez-Toledo, E., Tejada-Artigas, C., & Mañana-Rodríguez, J. (2013). Evaluation of scientific books' publishers in social sciences and humanities: Results of a survey. *Research Evaluation*, 22(1), 64–77. <https://doi.org/10.1093/reseval/rvs036>
- Gozlan, C. (2016). Les sciences humaines et sociales face aux standards d'évaluation de la qualité académique. *Sociologie*, 7(3), 261–21. <https://doi.org/10.3917/socio.073.0261>
- Hemlin, S. (1993). Scientific quality in the eyes of the scientist. A questionnaire study. *Scientometrics*, 27(1), 3–18. <https://doi.org/10.1007/BF02017752>
- Hug, S. E., Ochsner, M., & Daniel, H.-D. (2013). Criteria for assessing research quality in the humanities: a Delphi study among scholars of English literature, German literature and art history. *Research Evaluation*, 22(5), 369–383. <https://doi.org/10.1093/reseval/rvt008>
- Johnston, R. (2008). On structuring subjective judgements: Originality, significance and rigour in RAE2008. *Higher Education Quarterly*, 62(1-2), 120–147. <https://doi.org/10.1111/j.1468-2273.2008.00378.x>
- Koninklijke Nederlandse Akademie van Wetenschappen. (2011). *Quality Indicators for Research in the Humanities*. KNAW, Royal Netherlands Academy of Arts and Sciences. https://www.knaw.nl/en/news/publications/quality-indicators-for-research-in-the-humanities/@@download/pdf_file/20111024.pdf
- Krull, W., & Tepperwien, A. (2016). The four “I”s: Quality indicators for the humanities. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research Assessment in the Humanities. Towards criteria and procedures* (pp. 165–179). Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_13
- Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Harvard University Press.
- Langfeldt, L. (2006). The policy challenges of peer review: Managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation* 15(1), 31–41. <https://doi.org/10.3152/147154406781776039>
- Langfeldt, L., Bloch, C. W., & Sivertsen, G. (2015). Options and limitations in measuring the impact of research grants – evidence from Denmark and Norway. *Research Evaluation*, 24(3), 256–270. <https://doi.org/10.1093/reseval/rvv012>

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

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

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

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

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

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

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

ENRESSH Work Group 1

Overview of Peer Review Practices in the SSH

ENRESSH Report

Michael Ochsner, ETH Zurich and FORS, Lausanne, Switzerland
Nina Kancewicz-Hoffman, Institute of Literary Research, Polish Academy of
Sciences, Poland

Marek Hołowiecki, Adam Mickiewicz University Poznań, Poland
Jon Holm, National Research Council, Norway

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

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

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

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

Version 2.1, April 2020

Executive summary

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

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

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

Keywords

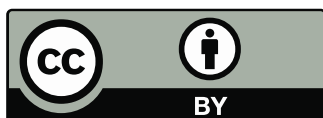
Peer Review, Evaluation, Criteria, Societal Impact, Books, Funding, Open Access

Acknowledgements

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

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

Georgios Afxentiou, Judit Bar-Ilan, Ondřej Daniel, Katya De Giovanni, Ioana Galeron, Aldis Gedutis, Haris Gekić, Elea Giménez-Toledo, Agnė Girkontaitė, Raf Guns, Jon Holm, Marek Hołowiecki, Todor Hristov Dechev, Dragan Ivanović, Nina Kancewicz-Hoffman, Andreja Isenič Starčič, Arnis Kokorevics, Litaugas 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



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

Recommended citation of the whole report:

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

Recommended citation of a chapter in the report:

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

Table of Content

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

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