

Research Assessment and Bibliometrics: Bringing Quality Back in

Michael Ochsner¹ and Sven E. Hug²

¹ ochsner@gess.ethz.ch

ETH Zurich, D-GESS, Mühlegasse 21, 8001 Zürich (Switzerland) and FORS, c/o University of Lausanne, Géopolis, 1015 Lausanne (Switzerland)

² sven.hug@gess.ethz.ch

ETH Zurich, D-GESS, Mühlegasse 21, 8001 Zürich (Switzerland) and University of Zurich, Evaluation Office, Mühlegasse 21, 8001 Zürich (Switzerland)

Introduction

Bibliometric indicators are used to compare research performances and also to assess and evaluate research performance (see, e.g. Gimenez-Toledo et al., 2007; Lane, 2010). However, recently scholars voice protest against bibliometric assessments (see, e.g., Lawrence, 2002; Molinier & Bodenhausen, 2010; Drubin, 2014). The arguments put forward are manifold. For example, the application of the impact factor, which is often used, but not meant, to evaluate individual researchers, is criticized (DORA, 2013). Then, there are myriads of perverse or unintended effects, like focus on high impact journals and mainstream topics, focus on review articles and short communications, strategic behavior, or lack of replication because of the low reputation of replication studies (e.g., Butler, 2007; Lawrence, 2003; Mooneshinghe et al., 2007). Furthermore, scholars from the social sciences and humanities (SSH) criticize that that bibliometric indicators cannot capture quality (e.g., Plumpe, 2009).

The authors of this paper were involved in a project to develop quality criteria and indicators for humanities research (see <http://www.psh.ethz.ch/crus>). Here, we argue that while bibliometric indicators and methods are powerful tools to describe research practices and, to some extent, scientific impact, there are some problems when they are readily used as quality indicators in research assessments. We feel that also other disciplines can learn from the critique of humanities scholars on simplistic quantitative assessments and from the findings of the research on quality in the humanities.

Notions of quality

The aim of the project “Developing and Testing Research Quality Criteria in the Humanities” was to find quality criteria and indicators that were at the same time accepted by the humanities scholars and implementable in different linguistic, cultural, and disciplinary settings. Analyzing the humanities scholars’ critique, we found that the development of criteria must take into account the disciplinary research practices, that the measurement must be

transparent and consensual, and that the notions of quality must be made explicit (Hug et al., 2014). We used the Repertory Grid technique to make the notions of quality explicit and base the development of quality criteria on the actual research practices. We found that there are two different conceptions of quality, a more traditional one, which can be described with individual, ground-breaking research that opens up new paradigms, and a more modern conception that can be described as interdisciplinary, project-focused, and public-oriented. Both kind of research can be good as well as bad (Ochsner et al., 2013). Hence, interdisciplinarity, for example, differentiates between two different ways of doing research but is not an indicator of quality (interdisciplinarity can point to good research, when it merges different theories and methods, but it can equally point to bad research that uses interdisciplinarity only for getting funding or for the career). Therefore, notions of quality should be taken into account in research evaluations. They might shed light on gaming strategies as well as on problems with indicators that are not linked to research practices or research quality.

Catalogue of quality criteria

Using the notions of quality, we developed a catalogue of quality criteria that are linked to the research practices in the humanities. Humanities scholars then rated these criteria as well as indicators measuring those criteria. We found that a broad range of quality criteria and aspects must be taken into account to adequately assess research quality (Hug et al., 2013) and that only about 3% to 32% of the scholars’ notions of quality can be quantified adequately, depending on the discipline. Furthermore, we found that there is a mismatch between the quality criteria put forward by the scholars and the quality criteria used in evaluation procedures (Ochsner et al., 2012). Hence, current evaluation procedures do not measure research quality in the humanities adequately. This does not mean that the existing evaluation procedures and criteria are useless (e.g., societal impact is not necessarily linked to research quality but is a legitimate criterion in evaluations), but it shows that

a very important dimension of research assessment is not reflected adequately: quality of research.

The humanities, so what?!

Our research bases on the humanities. What is the relevance of this research to the rest of academia? First, we argue that humanities scholars, while not specialised in quantification, are experts in critical thinking. Hence, their critique of evaluation procedures often points to the consequences of the instruments on research practices. This is what increasingly also happens in the natural sciences (e.g., DORA, 2013; Drubin, 2014) because some perverse effects start to become apparent. Hence, a focus on research practices in assessments could help minimise negative impact of indicators. Second, when we presented the criteria at conferences and workshops, also natural scientists were present. They surprisingly often said that the criteria we presented made also sense to them with a few exceptions. Hence, what could be learned from the case of the humanities would be the following: base evaluation procedures on research practices; be aware that the indicators used will affect the research practices; formulate quality criteria in a way that makes sense to the scholars; involve as many stakeholders as possible in the definition of quality criteria.

Bringing quality back in

While the bibliometric community is well aware of the possible drawbacks of bibliometric indicators, the most common reaction by the research evaluation community is to look for other sources of the same kind of indicators and altmetrics. We think that the problem is not a technical one but a conceptual. At the beginning of any research evaluation and science policy should be a reflection on the goals. Do we want scholars to use most of their time to feed Twitter, comment on Research Gate, or 'pimp' their statistics in Google Scholar? We think that research evaluation should bring quality back in. Evaluation and assessments should not solely *judge* the merits of scholars but help them to *enhance* their impact by fostering research quality. Hence, bibliometrics and altmetrics are powerful instruments to describe certain impacts, visibility, networks etc. But research assessments should also make clear statements about other aspects of research quality. Therefore, the disciplinary community should have a say in what criteria are applied in their assessments. New ideas of research evaluation based on research practices should lead scientific discussion much more than technical issues vaguely related to research quality.

Acknowledgments

This paper is based on work that was supported by the Rectors' Conference of the Swiss Universities

(CRUS) within the framework of the SUK B-05 Innovation and Cooperation Project "Mesurer les performances de la recherche". Matching funds were provided by the University of Zurich.

References

- Butler, L. (2007). Assessing university research: a plea for a balanced approach. *Science and Public Policy*, 34(8), 565-574.
- DORA. (2013). San Francisco Declaration on Research Assessment. <http://am.ascb.org/dora/>
- Drubin, D. (2014). Time to discard the metric that decides how science is rated. *The Conversation*. <http://theconversation.com/time-to-discard-the-metric-that-decides-how-science-is-rated-27733>
- Gimenez-Toledo, E., Roman-Roman, A. and Alcain-Partearroyo, M. D. (2007). From Experimentation to Coordination in the Evaluation of Spanish Scientific Journals in the Humanities and Social Sciences. *Research Evaluation*, 16(2), 137-48.
- 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.
- Hug, S. E., Ochsner, M., & Daniel, H.-D. (2014). A framework to explore and develop criteria for assessing research quality in the humanities. *International Journal for Education Law and Policy*, 10(1), 55-64.
- Lane, J. (2010). Let's Make Science Metrics More Scientific. *Nature*, 464(25), 488-489.
- Lawrence, P.A. (2002). Rank injustice. The misallocation of credit is endemic in science. *Nature*, 415, 835-836.
- Lawrence, P.A. (2003). The politics of publication. Authors, reviewers and editors must act to protect the quality of research. *Nature*, 422, 259-261.
- Molinie, A. & Bodenhausen, G. (2010). Bibliometrics as Weapons of Mass Citation. *Chimia* 64(1-2), 78-89.
- Mooneshinghe, R., Khoury, M. J., & Janssens, A. C. J. W. (2007). Most published research findings are false - but a little replication goes a long way. *PLOS Medicine*, 4(2), e28.
- 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, 79-92.
- Plumpe, W. (2009). Stellungnahme zum Rating des Wissenschaftsrates aus Sicht des Historikerverbandes. In C. Prinz & R. Hohls (Eds.), *Qualitätsmessung, Evaluation, Forschungsrating. Risiken und Chancen für die Geschichtswissenschaft?* (pp. 121-126). Historisches Forum. Berlin: Clionline.