Research Assessment in the Humanities:
Introduction

Michael Ochsner, Sven E. Hug and Hans-Dieter Daniel

Abstract Research assessments in the humanities are highly controversial. While citation-based research performance indicators are widely used in the natural and life sciences, quantitative measures for research performance meet strong opposition in the humanities. Since there are many problems connected to the use of bibliometrics in the humanities, new approaches have to be considered for the assessment of humanities research. Recently, concepts and methods for measuring research quality in the humanities have been developed in several countries. The edited volume ‘Research Assessment in the Humanities: Towards Criteria and Procedures’ analyses and discusses these recent developments in depth. It combines the presentation of state-of-the-art projects on research assessments in the humanities by humanities scholars themselves with a description of the evaluation of humanities research in practice presented by research funders. Bibliometric issues concerning humanities research complete the exhaustive analysis of humanities research assessment.

1 Introduction

Over the last decades, public institutions have experienced considerable changes towards greater efficiency and more direct accountability in many Western countries. To this end, new governmental practices, that is, new public management, have
been established. These practices did not stop at the gates of the universities (see e.g. Alexander 2000, p. 411; Mora 2001; Readings 1996; Rolfe 2013). In the past, scientific freedom guided practices at universities, and quality assurance was achieved endogeneously through peer review and rigorous appointment procedures for professorships. This sufficed as accountability to the public. Over the last decades, the university was increasingly understood as an institution that renders services to the economy, students and the public in general (see e.g. Mora 2001, p. 95; Rolfe 2013, p. 11). Such services were seen as value for money services, opening the door for new governance practices derived from theories based on market-orientation and efficiency (e.g. new public management).

While at first the natural and life sciences were in the focus of such new governance practices—the costly character of research projects in many natural and life science disciplines made such practices inevitable—, the humanities, which ignored such practices at first (and have been ignored by e.g. bibliometricians until lately), also came into focus (Guillory 2005, p. 28). However, the bibliometric approaches to research assessment used in the natural and life sciences yielded unsatisfying results when applied to the humanities due to different reasons, such as, amongst others, different publication practices and diverse publication channels (Hicks 2004; Mutz et al. 2013) or different research habits and practices and regional or local orientation (for an overview, see e.g. Nederhof 2006).

In light of these changes, the Swiss University Conference started a project organized by the Rectors’ Conference of the Swiss Universities (since 1 January 2015 called swissuniversities) entitled ‘B-05 mesurer la performance de la recherche’, with the goal to find ways to make more visible humanities’ and social sciences’ research performance and compare it on the international level (see the contribution by Loprieno et al. in this volume). The project consisted of three initiatives (research projects) and four actions (workshops and add-ons to the initiatives). The editors of this volume were involved in such an initiative entitled ‘Developing and Testing Research Quality Criteria in the Humanities, with an Emphasis on Literature Studies and Art History’ (see the contribution by Ochsner, Hug and Daniel in this volume1), which included one action that consisted of a series of colloquia about research quality and research assessment in the humanities. This series included a two-day international conference, a workshop on bibliometrics in the humanities and nine individual presentations between March 2009 and December 2012. This volume summarizes this series of presentations. The start of the series fell at a time when humanities scholars were repeatedly criticizing the evaluation and assessment practices by, for example, speaking up against two prominent initiatives to assess humanities research: the boycott of the research rating of the German Council of Science and Humanities (Wissenschaftsrat) by the Association of German Historians (Verband der Historiker und Historikerinnen Deutschlands) (see e.g. Plumpe 2009) and the rejection of the European Reference Index for the Humanities (ERIH) (see e.g. Andersen et al. 2009). Hence, the idea behind the series and this volume is

1See also the project’s website http://www.performances-recherche.ch/projects/developing-and-testing-quality-criteria-for-research-in-the-humanities.
letting humanities scholars themselves raise their voice about tools and procedures to evaluate humanities research. However, this volume also includes the view from the outside. To round out the picture, some scholars from the social sciences whose work focuses on research evaluation in the humanities are also present (see the chapters by Michèle Lamont and Joshua Guetzkow, by Ochsner, Hug and Daniel, by Thomas Koenig and by Björn Hammarfelt). Besides the fact that all authors come from the humanities and social sciences, the authors also represent a wide range of functional background: The selection of authors is well-balanced between humanities scholars, research funders and researchers on higher education.

The writing of this book started right after the two day international conference in Zurich entitled ‘Research Quality in the Humanities: Towards Criteria and Procedures for Evaluating Research’ in October 2010. The first contributions were submitted in early 2011. Because the series of colloquia continued, we soon realized that we wanted to expand the content of the book to other talks given in this series. Hence, the publication process was significantly extended. Many projects that are presented in the contributions have continued, and some of them have been concluded in the meantime. Thus, most chapters from 2011 had to be updated in 2014. We thank the authors for their patience with us, their understanding for the delay of the publication and their willingness to update their texts as well as their rapid revisions during the two rounds of peer review. We also want to thank the anonymous reviewers involved in the two review cycles at the early stage (book of extended abstracts) and final stage (full manuscript).

2 Structure of the Book

The book is structured in five parts. The first part presents the outset of the topic. On one hand, it describes the circumstances in which this book has been written, that is, the environment in which this project has been funded, and a description of the situation in which the humanities are concerning their competition with other subjects for funding at universities and funding institutions. On the other hand, it also comprises empirical studies on how peer review functions in the humanities as well as on the humanities scholars’ notions of quality. The second part presents the current state of quality-based publication rankings and publication databases. It focuses on projects that have their roots in the humanities and are led by a humanities scholar or focus specifically on the peculiarities of humanities research. The third part raises a delicate issue: bibliometrics in the humanities. It focuses on the problems in the application of bibliometric methods on humanities research as well as on the potential bibliometric analyses might bring if applied the right way. The fourth part focuses on the ex-ante evaluation of humanities research in practice, presenting humanities-specific evaluation procedures. The fifth part focuses on one influential ex-post practice of research evaluation that has been completely redesigned to match the needs of humanities research: The research rating of the subjects Anglistik and Amerikanistik by the German Council of Science and Humanities.
The first part starts with a contribution by Loprieno, Werlen, Hasgall and Bregy from the Rectors’ Conference of the Swiss Universities (CRUS, since 1 January 2015 called swissuniversities). They present the environment in which this volume was put together. It is a speciality of the humanities to understand the historicity of all knowledge, hence it is wise to start a volume on research assessment in the humanities presenting and reflecting on the context in which this volume has been created. Loprieno et al. present how the Swiss universities cope with the difficulty of evaluating humanities research. Their approach is scientific in nature: Following a case study in which the use of bibliometric methods in research assessment procedures for the humanities and social sciences was evaluated and found to be at least difficult if possible at all (CRUS 2009), a project was established that would scientifically investigate alternative instruments and approaches that measure aspects that cannot be captured by conventional bibliometry. The follow-up programme, drawing on the results of the first project, takes a step further and drops the concept of ‘measurement’ in favor of ‘visibility’.

The second chapter, by Wiljan van den Akker, takes the perspective from an established humanities scholar with many experiences in leading positions in university and research administration, as director of a research institute, as dean and as Director of Research at the Royal Academy of Sciences (KNAW) in the Netherlands. He argues that the humanities have to organize themselves to be able to play a role in science policy alongside the well-organized natural sciences. Hence, the humanities should also develop a system by which its research can be assessed. However, the humanities scholars should take the steering wheel in developing such a system to prevent being assessed by systems that are not suited to the nature of humanities research.

The contribution of Lamont and Guetzkow delves into how humanities and social sciences scholars assess research in expert peer review panels. They show the differences and commonalities between some humanities and social sciences disciplines in how research is evaluated by investigating informal rules, the impact of evaluation systems on such rules and definitions of originality. They show that cognitive aspects of evaluation cannot be separated from non-cognitive aspects and describe the evaluation process (by peer review) as interactional, emotional and cognitive. Peers mobilize their self-concept as well as their expertise in evaluation. Since there are different interpretations of criteria not only by peers but also by discipline, more emphasis must be put on the effect of panel composition in evaluations.

Ochsner, Hug and Daniel investigate how humanities scholars understand research quality. They take a bottom-up perspective and present quality criteria for research based on a survey administered to all scholars holding a PhD degree in three disciplines at the Swiss and the LERU universities. A broad range of quality criteria, they conclude, must be taken into account if humanities research is to be assessed appropriately. They also show that a vast majority of humanities scholars reject a purely quantitative approach to evaluation.

The first part thus provides information on the framework in which this volume has been put together and points to the ‘Swiss way to quality’, i.e. a scientific approach towards research evaluation. It furthermore puts forward reasons why the humanities
disciplines should take their evaluation into their own hands. Finally, it provides empirical evidence on how evaluation by experts works and contrasts it with the view on research quality by humanities scholars from the grass-roots perspective.

The second part of the book focuses on publication rankings and publication databases. Publications lie at the heart of scientific work. Therefore, publications are often used in research evaluations, be it simply by counting the number of publications of a unit or by the use of complex rankings of publication channels.

The chapter by Gerhard Lauer opens this part of the book. He reports on the initiative of several national research funders to establish a publication database for the social sciences and humanities (SSH). He describes the problems and opposition experienced with the ERIH project, lists the requirements for a comprehensive (open) publication database that can be useful to the SSH and depicts the future of ERIH.

Gunnar Sivertsen presents such a publication database on the national level, the so-called Norwegian Model. It serves as the foundation of a publication-based performance indicator applied in Norway that distributes extra funding for research in a competitive way. Evaluations of the model show that a comprehensive publication database can be useful not only for research administrators but also for the humanities scholars themselves: It makes visible humanities research and shows that humanities scholars are conducting at least as much research as scholars from the natural and life sciences. Additionally, it can also serve information retrieval purposes for humanities scholars.

Often, publications are not just counted but also weighted according to their academic value. This is an intricate task. Elea Giménez Toledo presents how SSH journals and books are evaluated in Spain using quality criteria for publication channels. She also shows how journal and book publisher lists are used in evaluations.

The contribution by Ingrid Gogolin, finally, summarizes the European Educational Research Quality Indicators (EERQI) project. This project was initiated as a reaction against the rising relevance of numerous university rankings and citation-based indicators that are not adequately reflecting the publication practices of (European) SSH research. The aim of EERQI is to combine different evaluation methods and indicators to facilitate review practices as well as enhance the transparency of evaluation processes.

Summarizing the second part of the book, there is a lack of information about SSH publications. Establishing a database for SSH publications can lead to more visibility of SSH research, which can serve scholars in terms of information retrieval. At the same time, it may also serve administrators for evaluation purposes. Thus, creating publication databases should go hand in hand with the development of standards regarding how to use or not use publication databases in SSH research evaluation.

One of the most commonly used instruments based on publication databases to evaluate research in the natural and life sciences are bibliometric indicators. The third part of the book investigates the limitations and potential of bibliometric instruments when applied to the humanities. The third part starts with the contribution by Björn Hammarfelt. He describes the state-of-the-art of bibliometrics in the humanities and sketches a ‘bibliometrics for the humanities’ that is based upon humanities’ publication practices. He argues that while it is true that conventional bibliometrics
cannot readily be used for the assessment of humanities research, bibliometrics might nevertheless be used to complement peer review if the bibliometric methods are adapted to the social and intellectual organization of the humanities.

In the second chapter of this part, Remigius Bunia, a German literature scholar, critically investigates why bibliometrics cannot be applied in the humanities with the example of German literature studies. While Bunia acknowledges that a part of the problem is due to technical and coverage issues of the commercial citation databases, he argues that there might also be a problem involved that is intrinsic to the field of literature studies: the fact that literature scholars seem not to read the works of other literature scholars or at least not to use (or cite) them in their own work. To test this claim, Bunia advocates applying bibliometrics to study what and how literary scholars cite and to critically examine the citation behaviour of literary scholars. Until light is shed on this issue, a bibliometric assessment of research performance in the humanities is not possible.

Thus, the third part of this volume shows that bibliometrics cannot be readily used to evaluate humanities research. Yet, bibliometrics adapted to the humanities can serve as tools to study publication and citation habits and patterns as well as to complement peer review. Knowing more about publication and citation habits also makes it possible to broach delicate issues in research practices.

Even though bibliometric assessment is not (yet) possible in the humanities, humanities research is assessed on a regular basis. Part four of this volume presents procedures regarding how humanities research is evaluated in practice and approaches regarding how an assessment of humanities research might look. In the focus of part four are ex-ante evaluations, i.e. evaluations of research yet to be conducted. Thomas König shares insights into the evaluation practices at the European Research Council (ERC). There was not much funding of SSH research on the European level until 2007. According to König, this is not only due to the reluctance of politicians to fund SSH in general but also because of the fact that (a) humanities researchers do not ask for funding as frequently as natural scientists and (b) SSH scholars are much less formally organized and thus cannot lobby as effectively on the political scene as natural scientists. However, the SSH’s share of funding for ERC grants is considerably higher than for the whole FP7—and rising. The distribution of applications for grants shows that there are differences between SSH disciplines in asking for funding. The results also show that despite some fears of disadvantages in interdisciplinary panels, SSH disciplines reach similar acceptance rates as the natural sciences in ERC grants.

For the next chapter we change to a private funding institution. Wilhelm Krull and Antje Tepperwien report how humanities research is evaluated in the Volkswagen Foundation, one of the largest private research funding institutions in Europe. In order to prevent falling into pitfalls by quantitative indicators not adapted to the characteristics of humanities research, they suggest guiding the evaluation of humanities research according to four ‘I’s’: infrastructure, innovation, interdisciplinarity and internationality. They also reveal important insights about evaluation practice in the humanities: Humanities reviewers even criticize proposals that they rate as excellent, a fact which can lead to disadvantages in interdisciplinary panels, as reviewers...
from the natural sciences do not understand why something might be very good even though it can be criticized.

The third chapter in this part presents evaluation procedures in France. After explaining the evaluation practices of the key actors in France—AERES, ANR, CNU and CNRS—Geoffrey Williams and Ioana Galleron describe two ongoing projects that aim at understanding the characteristics of French humanities research. The first project, DisValHum, aims at understanding the dissemination practices of French humanities scholars. The second, IMPRESHS, strives to bring about a better understanding of the variety of impacts humanities research can have.

The fourth part thus shows that humanities scholars do not apply for external funding as much as could be possible. Furthermore, humanities scholars are not organized well enough to lobby for humanities research on the national as well as the international level. Additionally, humanities research can be disadvantaged in interdisciplinary panels in ex-ante evaluations because of the fact that humanities scholars also criticize work they consider excellent, whereas natural scientists feel that no work should be funded that can be criticized.

The last part of the book is dedicated to a specific ex-post evaluation procedure that has been adapted for the humanities recently: the research rating of the German Council of Science and Humanities. The contribution by Christian Mair briefly describes the history of, and ideas behind, the research rating. He argues that the failure of the first attempt to conduct a pilot study for the research rating in the humanities was mainly a communication problem. He then describes the process of fleshing out a rating procedure adapted to the humanities by an expert group of humanities scholars that resulted in a pilot study of the research rating in Anglistik/Amerikanistik.

The joint contribution by Klaus Stierstorfer and Peter Schneck gives insight into the arguments for and against participating in such a rating exercise by the presidents of the two associations involved, the Deutsche Anglistenverband (German Association for English Studies) and Deutsche Gesellschaft für Amerikastudien (German Association for American Studies). Stierstorfer, then-president of the Deutsche Anglistenverband, argues that while research ratings as such are not naturally in the interest of humanities scholars but are likely to be here to stay, there might nevertheless accrue some collateral benefits. Hence, the rating has to be optimized to maximize such benefits. Peter Schneck, president of the Deutsche Gesellschaft für Amerikastudien from 2008 to 2011, also takes a very critical stance on the usefulness of research ratings. He acknowledges, however, that rating is an integral part of academic life, also in the humanities (e.g. grading students as well as rating applicants for a professorship). Therefore, he argues, the humanities should actively get involved in the discussion about standards for research assessments rather than boycott them.

The research rating Anglistik/Amerikanistik was finished in 2012. The third contribution of this part presents experiences from this pilot study from the perspective of the involved staff at the Council and members of the review board: It starts with the conclusions drawn from this pilot by the German Council of Science and Humanities. It describes the exact procedure of the research rating of Anglistik/Amerikanistik and concludes that the research rating is suitable for taking into account the specifics of the humanities research practice in the context of research assessments. The contribu-
tion continues with the perspective of Alfred Hornung, who chaired the review board of the rating as an Amerikanistik-scholar. He describes the critiques and concerns that accompanied the rating as well as the benefits of the exercise. Barbara Korte concludes this contribution with her insights into the pilot study as a member of the review board and as an Anglistik-scholar. She illustrates the difficulties of defining subdisciplines within a broad field. She warns that while the research rating helped to show the diversity of English studies, it also might have aroused more thoughts about divisions than about common research interests.

Finally, the contribution by Ingo Plag presents an empirical analysis of the ratings done during the research rating Anglistik/Amerikanistik. His analysis shows that there is a quite low variability in the ratings across raters, pointing to a high reliability of the research rating. Most criteria correlate highly with each other. However, third-party funding proves not to be a good indicator of research quality.

3 Synopsis, Outlook and Acknowledgements

The contributions in this volume show that there is no easy way to assess humanities research. The first part shows that there is no one-size-fits-all solution to research assessment: There are many disciplinary differences that must be taken into account. If humanities research is to be assessed, a broad range of criteria must be considered. However, as the second part of the book shows, there is a lack of information about humanities publications and dissemination practices. The presented projects suggest that the creation of publication databases should go hand in hand with the development of standards regarding how to use or not use publication databases in humanities research evaluation in order to protect the humanities from the perverse effects of the misuse of the information provided by such databases. Bibliometric analysis of publications cannot be used as a sole assessment tool, as is shown in the third part of the book. It is an instrument that is too simplistic and one-dimensional to take into account the diversity of impacts, uses and goals of humanities research. Publication databases and citation analysis could, however, help in providing information on dissemination patterns and their evolution if the databases were to be expanded to cover most of the humanities research.

Humanities scholars are not yet applying for external funding as much as they could. Funders that are willing to fund humanities research do exist, and there are funding instruments specifically created for humanities research. Yet, it seems that humanities scholars are not yet used to applying for grants. This might be due to the fact that they seem not to be organized formally enough to compete with the natural sciences on the political level so that many calls for proposals seem to exclude humanities research, and, consequently, humanities scholars think that their chances are too small to invest in the crafting of the proposal. Hence, it is obvious that humanities scholars not only have to organize themselves better but also that the evaluation procedures and criteria must be compatible with humanities research, as the fourth part of the book makes clear. This is not only true for ex-ante evaluation
but especially for ex-post assessments. Thus, humanities scholars should have a say in the design of assessment procedures in order to prevent negative effects of such assessments on research quality in the humanities. Assessments should be optimized in such a way that the benefits are maximized. This is the conclusion of the fifth part of the book.

This volume presents many different views on research assessment in the humanities. Scholars from very different fields of research as well as representing different functions within the assessment environment present contributions of different kinds: descriptions of projects, essays of opinions about assessments and empirical analyses of research assessments. Thus, we hope, this volume presents an interesting and diverse picture of the problems and advantages of assessments as well as of the opportunities and limitations that come with them. Despite different perspectives and opinions on research evaluation, all authors share the belief that, given that assessments are a reality, the humanities should take an active role in shaping the evaluation procedures that are used to assess humanities research in order to prevent negative consequences and to take as much benefit out of the exercise as possible.

The contributions in this volume also clearly show that in order to shape assessment procedures so that humanities research can benefit to a maximum, further research is needed: First, there needs to be more fine-grained knowledge about what exactly good research looks like in the humanities and what research quality actually means. Second, more knowledge on the social and intellectual organization of humanities research would also facilitate the organization of research assessments: What are the publication and dissemination habits in the humanities? Third, more research on peer review is needed, for example, to what extent can peers be informed by quantitative indicators in order to reduce subjectivity and prevent reinforcing old hierarchies? Fourth, investigations into the effect research assessments have on humanities research are also dearly needed. They provide important insights on what to avoid as well as what to focus on in future assessments.

These are only some of the possible routes for research on research assessments in the humanities. We think that if research is to be assessed, the assessments should also live up to scientific standards. Therefore, we need to base assessment procedures for the humanities on scientific knowledge about the organization of humanities research. While there is a hundred years of research on natural and life sciences, research on the humanities is still scarce. This volume presents some paths to take.

The creation of this volume lasted from 2010 until 2015. During this long time period, many people were involved in the production of this volume. We are very grateful for the commitment of these individuals. It all started in the fall of 2010 with the organization of an international conference on research quality in the humanities. We would like to thank Vanessa McSorley for her help contacting the scholars we had in mind for the conference. Special thanks are due to Heidi Ritz for her tireless commitment and the perfect organization of the event as well as for the communication with potential publishers and with the authors in the early phase of the creation of the book until 2011. Of course, we also thank Sandra Rusch and Fabian Gander, who were involved in the organization and realization of the conference. Many thanks are also due to Julia Wolf, who organized the workshop on bibliometrics in
the humanities. We are heavily indebted to Esther Germann, who supported us in many aspects of the final phase of the process from 2012 to 2015. She formatted many contributions, optimized the figures, controlled the process with the English editing and assisted us in all issues concerning the English language. She shared all the ups and downs that come with editing a book. We also want to thank the anonymous reviewers involved in the two cycles of peer review. Last but not least, we thank all the authors for their contributions and for their patience over the long publishing procedure.

Open Access  This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (http://creativecommons.org/licenses/by-nc/2.5/) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work’s Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work’s Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

Alexander, F. K. (2000). The changing face of accountability: Monitoring and assessing institutional performance in higher education. *Journal of Higher Education, 71*(4), 411–431. doi:10.2307/2649146.

Andersen, H., Ariew, R., Feingold, M., Bag, A. K., Barrow-Green, J., van Dalen, B., et al. (2009). Editorial: journals under threat: A joint response from history of science, technology and medicine editors. *Social Studies of Science, 39*(1), 6–9.

CRUS. (2009). Projet ‘Mesurer les performances de la recherche’—1er Rapport. Bern: CRUS.

Guillory, J. (2005). Valuing the humanities, evaluating scholarship. *Profession, 11*, 28–38. doi:10.1632/074069505X79071.

Hicks, D. (2004). The four literatures of social science. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems* (pp. 476–496). Dordrecht: Kluwer Academic Publishers.

Mora, J.-G. (2001). Governance and management in the new university. *Tertiary Education and Management, 7*(2), 95–110. doi:10.1023/A:1011338016085.

Mutz, R., Bornmann, L., & Daniel, H.-D. (2013). Types of research output profiles: A multilevel latent class analysis of the Austrian Science Fund’s final project report data. *Research Evaluation, 22*(2), 118–133. doi:10.1093/reseval/rvs038.

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.

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 Geschichtswissenschaften? (pp. 121-126). Historisches Forum. Berlin: Clio-online. Retrieved from http://edoc.hu-berlin.de/e_histfor/12/.

Readings, B. (1996). *The university in ruins*. Cambridge, MA: Harvard University Press.

Rolfe, G. (2013). *The university in dissent. Scholarship in the corporate university*. Abingdon: Routledge.