Michael Ochsner · Sven E. Hug
Hans-Dieter Daniel *Editors*

Research Assessment in the Humanities

Towards Criteria and Procedures

Springer Open

Research Assessment in the Humanities

Michael Ochsner · Sven E. Hug
Hans-Dieter Daniel
Editors

Research Assessment in the Humanities

Towards Criteria and Procedures

Springer Open

Editors
Michael Ochsner
D-GESS
ETH Zürich
Zürich
Switzerland

and

FORS
University of Lausanne
Lausanne
Switzerland

Hans-Dieter Daniel
D-GESS
ETH Zürich
Zürich
Switzerland

and

Evaluation Office
University of Zürich
Zürich
Switzerland

Sven E. Hug
D-GESS
ETH Zürich
Zürich
Switzerland

and

Evaluation Office
University of Zürich
Zürich
Switzerland

ISBN 978-3-319-29014-0 ISBN 978-3-319-29016-4 (eBook)
DOI 10.1007/978-3-319-29016-4

Library of Congress Control Number: 2016932344

Printed on acid-free paper

This Springer imprint is published by SpringerNature
The registered company is Springer International Publishing AG Switzerland

Foreword

The volume that lays in front of you covers an important topic, namely the search for academic quality in research in the domain of the humanities and, particularly, how to come to terms on how to operationalize that in research assessment contexts. Over the last 20 years, we have witnessed, particularly in Europe, a growing influence of quantitative techniques on the measurement of research performance, mainly in the natural, life, biomedical and engineering sciences. And although it was clearly acknowledged that these quantitative, bibliometric techniques have lesser relevance in the social sciences, humanities and law (SSH), the pressure on these domains to adapt to the research assessment practices of a quantitative nature, as applied in the natural, life, biomedical and engineering sciences, grew steadily. And while some of these techniques did work in those few specialties of the social sciences, in which journal publishing has become the field's standard, it clearly was not applicable in most other specialties of the social sciences, nearly all of the humanities and in law.

This increasing pressure on SSH scholars to show quantitatively how they perform in research assessment procedures led to much protesting reactions from the social sciences and humanities communities. So we witnessed a fierce debate on the applicability of bibliometric techniques around a research assessment procedure in the field of psychology in the Netherlands, centred around the role of books in the assessment of psychology research. In Belgium, the application of the journal impact factor as part of the funding allocation model applied in Flanders, has led to the creation of an academic bibliographic system that could better serve the interests of scholars in the social sciences and humanities in that same funding model. And finally, in 2012/2013, German SSH scholars made clear statements, when first economists, followed by sociologists, historians and educationalists protested against academic rankings. And as these protests have created a higher degree of awareness on the importance of having a better insight in the publication output types and scholarly communication practices of scholars in the SSH domains, and initiated a variety of research on that topic, a more important development has been

that an academic interest grew with respect to the variety of research and communication practices all across the humanities and social sciences landscape.

And that is exactly what this new volume is demonstrating: a focus on the different aspects of scholarly practice in the humanities, and the ways these are reflected in research assessment procedures. Important in that respect is that this development is taking place by and through scholars in the humanities themselves. By consulting and listening to the scholars that are subject to research assessment, one can learn how the assessment of that type of research should be organized, by streamlining assessment practices towards local research and communication practices. An important question addressed in the volume is on how academic quality is perceived by scholars in the humanities, and not only through qualitative procedures, but also by quantitative means. Where peer review has been the backbone of research assessment in the humanities in the past, and will remain to be in the future, the initiative on the development of various quantitative approaches has to be welcomed as additional methodologies, informing peer-review processes. And while I realize that these quantitative methodologies do stir up a lot of discussion, this discussion is productive in the sense that it is the scholarly community within the social sciences and humanities itself that is involved now, thereby taking things in their own hands, rather than being confronted with top-down installed bibliometric techniques that do not fit into the variety of the academic work in the social sciences and humanities.

The editors of this volume have done a great job by joining together a wide variety of internationally highly reputed scholars from various academic ranks and backgrounds in the social sciences and humanities, all very well qualified to describe the most recent developments in the research assessment practices they are involved in, either locally or internationally. Furthermore, the volume is a display of the variety of research practices in various domains of the humanities, reflecting the heterogeneity of the scholarly research and communication practices within the humanities.

To conclude this preface, I sincerely hope that this volume contributes to a further extension of the academic efforts from within the humanities to think and develop procedures and methodologies that suit research assessment practices in the humanities.

Leiden Thed van Leeuwen
December 2015

Contents

About the Contributors

Jaromir Bregy finished his master of arts in sociology at the University of Bern in 2012, with a research focus on migration and integration. Since 2013, he has been working as a scientific assistant at the Rectors' Conference of the Swiss Universities (CRUS, since 1 January 2015 called swissuniversities). Amongst others he is involved in the programme "Performance de la recherche en sciences humaines et sociales".

Remigius Bunia is a former junior professor of comparative literature. Today he is an entrepreneur and a freelance author. He earned his master-level degrees in German literature and mathematics at the University of Bonn in 2002 and obtained his Ph.D. (Dr. phil.) at the University of Siegen in 2006. He is the author of three books, *Metrik und Kulturpolitik* (2014), *Romantischer Rationalismus* (2013) and *Faltungen* (2007), and of various articles on fictionality, aesthetics, semantic history, law/literature, quotation culture and the book as a medium. Website: http://litwiss.bunia.de

Hans-Dieter Daniel holds a dual professorship at ETH Zurich and at the University of Zurich. Since 2001, he has been the director of the evaluation office of the University of Zurich and since 2002, professor for social psychology and research on higher education at ETH Zurich. Dr. Daniel is a psychologist by training. Since 2011, he has been a member of the evaluation committee of the German Council of Science and Humanities (Wissenschaftsrat). His scholarly interests include research on peer review and evaluative bibliometrics. He is a highly cited researcher and a co-author of several highly cited journal articles in Essential Science Indicators from Thomson Reuters as well as author of the book *Guardians of Science—Fairness and Reliability of Peer Review*.

Ioana Galleron is a specialist in the eighteenth century French literature, senior lecturer at the University of South Brittany, France, and the former pro-vice-chancellor of the university in charge of administrative affairs and quality assessment. She is the co-founder and treasurer of the EvalHum initiative (www.evalhum.eu), a European association for research evaluation in the social sciences

and humanities (SSH). She is currently working on various projects dedicated to the evaluation of the SSH research, financed by the French national network for humanities centres (MSH): quality representation in SSH research, the publishing strategies of SSH researchers and the impact of their productions.

Elea Giménez Toledo holds a Ph.D. in information science. She is research fellow at the Spanish National Research Council (CSIC), and she is head of the Research Group on Scholarly Books (ÍLIA) (http://ilia.cchs.csic.es), which is devoted to the analysis of scholarly publishing in the social sciences and humanities as well as the relationship with its environment (authors, publishers, referees, readers and evaluation agencies). She is member of the EvalHum initiative (www.evalhum.eu), a European association for research evaluation in the SSH.

Ingrid Gogolin is professor of international comparative and intercultural education research at the University of Hamburg. Her research is focused on problems of migration and linguistic diversity in education. She was coordinator of the EERQI-project. Her recent research projects deal with the following topics: linguistic diversity management in urban areas (Research Cluster of Excellence at the University of Hamburg; www.lima.uni-hamburg.de), support of migrant children in schools (www.foermig.uni-hamburg.de), multilingualism and education (www.kombi.uni-hamburg.de) and multilingual development in a longitudinal perspective (MEZ; www.mez.uni-hamburg.de).

Joshua Guetzkow is assistant professor of sociology and anthropology at the Hebrew University of Jerusalem. His research on the impact of litigation on U.S. prisons is forthcoming in the *Law and Society Review*, and a study of how policymaking ideas influenced U.S. welfare policymaking was published in *The Annals of the American Academy of Social and Political Sciences* (2010).

Björn Hammarfelt (Ph.D.) is a senior lecturer at the Swedish School of Library and Information Science (SSLIS), University of Borås, and a visiting scholar at the Centre for Science and Technology Studies (CWTS), Leiden University. His research is situated at the intersection between information science and sociology of science, with a focus on the organization and communication of knowledge. Hammarfelt has published extensively on the use of bibliometric methods for analysing the humanities, and his papers can be found in journals such as *JASIST*, *Journal of Documentation*, *Scientometrics* and *Research Evaluation*. Currently he is engaged in a project studying how scholars respond to the further use of bibliometric measures for evaluating research.

Alexander Hasgall is the scientific coordinator of the programme Performances de la recherche en sciences humaines et sociales of the Rectors' Conference of the Swiss Universities (CRUS, since 1 January 2015 called swissuniversities) and is based at the University of Geneva. He studied philosophy and history at the University of Zurich and wrote his doctoral thesis on the politics of recognition in relation to the last military dictatorship in Argentina. Before starting his current

position, he worked in the NGO sector on the issue of combatting anti-Semitism, racism and other forms of intolerance and as a freelance journalist.

Alfred Hornung is professor and chair of English and American studies at the University of Mainz. He held guest professorships at various European, American, Canadian and Chinese universities. His publications are in the field of modernism, postmodernism, life writing, intercultural and transnational studies. He is editor and on the editorial board of the *Journal of Transnational American Studies*, *Atlantic Studies* and *Contemporary Foreign Literature* (Nanjing). He served as president of Multi-Ethnic Studies: Europe and the Americas (MESEA) and of the German Association for American Studies. As an elected member of the review board for literature of the DFG he chaired the research rating of *Anglistik/Amerikanistik* for the Wissenschaftsrat. In 2013 he received the Carl Bode-Norman Holmes Pearson Prize of the American Studies Association for outstanding contributions to American Studies in Washington. He is a member of Academia Europaea.

Sven E. Hug studied German language and literature as well as psychology at the University of Zurich and worked in various companies as a market research analyst. He is currently working as a project manager at the evaluation office of the University of Zurich and in addition acts as a research associate in the swissuniversities-organized project Application of Bottom-up Criteria in the Assessment of Grant Proposals of Junior Researchers in the Social Sciences and Humanities at the Professorship for Social Psychology and Research on Higher Education (ETH Zurich).

Veronika Khlavna studied social sciences at the Ruhr-University Bochum. Since 2008 Dr. Veronika Khlavna has worked in the research department at the head office of the German Council for Science and Humanities. Between 2008 and 2013 she was part of the Council's research rating project and coordinated the assessment group for the pilot study *Anglistik/Amerikanistik* from 2011 to 2012.

Thomas König studied political science and history and holds a Ph.D. from the University of Vienna. After fellowships at the Universities of Minnesota and Harvard, he is now researcher at the Institute for Advanced Studies (IHS), Vienna. Between 2010 and 2013, he served as scientific advisor to the president of the European Research Council (ERC). Among many other things, he observed the ERC evaluation panels in the social sciences and humanities domain, a task that brought him close insights into different epistemic communities in those fields.

Barbara Korte was appointed to a chair in English literature at the University of Freiburg in 2002. She previously held professorial positions at the universities of Chemnitz and Tübingen. She was a member of the German Research Council's (DFG) *Fachkollegium* (review board) for literary studies and a reviewer in the *Wissenschaftsrat*'s pilot study. Her main research area is British literature and culture from the Victorian period to the present, and she has written internationally published monographs on travel writing, war literature and the representation of poverty.

Wilhelm Krull has been running the Volkswagen Foundation since 1996—following his studies in German, philosophy, education and politics, an appointment as a DAAD lecturer at the University of Oxford and leading positions at the Wissenschaftsrat (German Council of Science and Humanities) and at the headquarters of the Max-Planck-Gesellschaft (Max Planck Society). Besides his professional activities in science policy as well as in the promotion and funding of research, Dr. Wilhelm Krull was and still is a member of numerous national, foreign and international boards and committees.

Michèle Lamont is professor of sociology and African and African American studies and Robert I. Goldman Professor of European Studies at Harvard University. Her recent publications include *Social Resilience in the Neoliberal Era* (with Peter A. Hall, 2013), *Responses to Stigmatization in Comparative Perspective* (coedited with Nissim Mizrachi, 2012), *Social Knowledge in the Making* (coedited with Charles Camic and Neil Gross, 2011) and *How Professors Think: Inside the Curious World of Academic Judgment* (2009).

Gerhard Lauer is a tenured professor for German philology at the Georg-August-Universität Göttingen as well as the founder and the current head of the Göttingen Centre for Digital Humanities. His research interests focus on (German) literary history, cognitive poetics and digital humanities.

Antonio Loprieno has been full professor of egyptology at the University of Basel since the year 2000. His main research areas include Near Eastern languages and Egyptian cultural history and religion. He served as rector of the University of Basel from 2006 to 2015 and as President of the Rectors' Conference of the Swiss Universities (CRUS) from 2008 to the birth of swissuniversities in 2015. He is member of the University Council of the University of Zurich.

Christian Mair Christian Mair has taught English linguistics at the Universities of Innsbruck, Austria, and—subsequently—Freiburg, Germany, where he has been a professor since 1990. He has published widely on English syntax and change and variation in World Englishes. From 2006 to 2012, he was a member of the German *Wissenschaftsrat* and involved in several of this advisory body's initiatives in the fields of research assessment and academic quality assurance.

Michael Ochsner finished his doctoral studies at the Institute of Sociology of the University of Zurich in 2012 and received his Ph.D. in 2014. Since 2009, he has been a research associate at the ETH Zurich in the CRUS-organized projects: Developing and Testing Research Quality Criteria in the Humanities, with an emphasis on Literature Studies and Art History and Application of Bottom-up Criteria in the Assessment of Grant Proposals of Junior Researchers in the Social Sciences and Humanities. Since 2013, he has also worked at the Swiss Centre of Expertise in the Social Sciences (FORS) at the University of Lausanne as a senior researcher in the team international surveys. He is vice-president of the EvalHum initiative, a European association for research evaluation in the SSH.

Ingo Plag is professor of English language and linguistics at Heinrich-Heine-Universität Düsseldorf. His monographs include *Morphological Productivity* (Mouton de Gruyter 1999), *Word-formation in English* (CUP 2003), *Introduction to English Linguistics* (with co-authors, Mouton de Gruyter 2009) and *A Reference Guide to English Morphology* (with co-Authors, Oxford University Press, 2013). He has published numerous articles in peer-reviewed journals and is editor of the journal *Morphology*. He has acted as referee nationally and internationally for journals, publishing houses, academic institutions and funding organizations. He is a member of the Academia Europaea.

Peter Schneck is dean of the faculty of languages and literatures and professor and chair of American literature and culture at Osnabrück University, Germany. He is also the co-founder and the co-director of the research cluster Cognition and Poetics, and director of the Osnabrück Summer Institute on the Cultural Study of the Law (OSI). He has been a fellow and a visiting scholar at numerous universities in the U.S. and Italy. Since 2006, he has been a member of the board of the German Association for American Studies (Deutsche Gesellschaft für Amerikastudien), and he served as the association's president from 2008 to 2011.

Gunnar Sivertsen is research professor and chair of the bibliometric research group at the Nordic Institute for Studies in Innovation, Research and Education (NIFU), Norway. He specializes in studies supporting research policy and assessment as well as in the development and use of indicators for statistics, evaluation, funding and science policy. Dr. Sivertsen has advised the development of institutional information and funding systems in Belgium, the Czech Republic, Denmark, Finland, Lithuania, the Netherlands, Norway, Portugal and Sweden. In 2010, he chaired a working group appointed by the European Science Foundation to look for solutions for comprehensive bibliometric data coverage in the social sciences and humanities. Sivertsen has a doctoral degree in eighteenth century Scandinavian literature.

Klaus Stierstorfer is professor of British studies at Westfälische Wilhelms-Universität Münster. He studied in Regensburg and Oxford. He wrote his doctoral dissertation on the Victorian drama at Oxford, his postdoctoral thesis on the history of English literature in Würzburg. In 2002, he accepted the offer of a professorship in Düsseldorf from where he moved in 2004 to his current position. From 2007 to 2010, he was president of the German Association for English Studies (Deutscher Anglistenverband). His research interests are history of literature, literary and cultural theory, law and literature, as well as literature in a European and transnational context. He is coordinator of the Marie Curie Initial Training Programme Diasporic Constructions of Home and Belonging and director of the graduate training programme Literary Form: Historical and Cultural Formations of Aesthetic Models as well as co-coordinator of the Europa-Kolleg in Münster.

Antje Tepperwien studied history and modern German literature in Tübingen, Aix-en-Provence and Durham and received her Ph.D. in modern history from Philipps-Universität Marburg. Before joining the Volkswagen Foundation as executive assistant to the secretary general in 2008, she worked at the Welcome Centre for International Academics at the University of Marburg. Since 2012, she has been a member of the advisory board of the German University Association of Advanced Graduate Training (UniWiND/GUAT).

Wiljan van den Akker is distinguished professor of modern poetry at Utrecht University. From 1993 to 2003, he was the director of the Research Institute for History and Culture, after which he became the director of research at the Royal Academy of Sciences (KNAW) in Amsterdam. From 2006 to 2014, he was dean of humanities and is presently serving as vice-rector for research at Utrecht University.

Raymond Werlen studied biochemistry at the ETH Zurich and the University of Geneva, where he did research at the department of medical biochemistry with a focus on the preparation of well-defined bio-conjugates through the regio-selective modification of proteins. After an engagement at the Swiss University Conference in 1996, Dr. Raymond Werlen was deputy secretary general of the Rectors' Conference of the Swiss Universities (CRUS) from 2001 to 2012, where he was mainly in charge of questions related to strategy, planning, quality and research. From 2013 to 2015, he was secretary general of the CRUS.

Geoffrey Williams is professor of applied linguistics at the University of South Brittany and is the former vice president for international relations. He has a particular interest in rankings and their application to the social sciences and the humanities. He holds an M.Sc. from Aston University and a Ph.D. from the University of Nantes. He is a former president of the European Association for Lexicography—EURALEX and has published widely in his field. He is a member of numerous academic societies. He is currently the director of the Department for Document Management and of the LiCoRN research group and the president of the EvalHum initiative, a European association for research evaluation in the SSH.

List of Figures

Assessment of Journal & Book Publishers in the Humanities and Social Sciences in Spain

European Educational Research Quality Indicators (EERQI): An Experiment

Beyond Coverage: Toward a Bibliometrics for the Humanities

Peer Review in the Social Sciences and Humanities at the European Level: The Experiences of the European Research Council

Research Assessment in a Philological Discipline: Criteria and Rater Reliability

List of Tables

Bottom Up from the Bottom: A New Outlook on Research Evaluation for the SSH in France

Research Assessment in a Philological Discipline: Criteria and Rater Reliability

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

M. Ochsner (✉) · S.E. Hug · H.-D. Daniel
Professorship for Social Psychology and Research on Higher Education,
Department of Humanities, Social and Political Sciences, ETH Zürich,
Mühlegasse 21, 8001 Zürich, Switzerland
e-mail: ochsner@gess.ethz.ch

S.E. Hug
e-mail: sven.hug@gess.ethz.ch

H.-D. Daniel
e-mail: daniel@gess.ethz.ch

M. Ochsner
FORS, University of Lausanne, Géopolis, 1015 Lausanne, Switzerland

S.E. Hug · H.-D. Daniel
Evaluation Office, University of Zürich, Mühlegasse 21, 8001 Zürich, Switzerland

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_1

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 volume[1]), 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

[1]See 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 reenforcing 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.

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.

Part I
Setting Sail into Stormy Waters

The 'Mesurer les Performances de la Recherche' Project of the Rectors' Conference of the Swiss Universities (CRUS) and Its Further Development

Antonio Loprieno, Raymond Werlen, Alexander Hasgall and Jaromir Bregy

Abstract The 'Mesurer les performances de la recherche' project was funded through project-related subsidies of the Swiss Confederation allocated by the Swiss University Conference. Over the period 2008–2012, the project supported the exploration of new approaches to measure aspects of research that cannot be captured by conventional bibliometry. The project followed the Swiss Way to Quality in the Swiss universities (CRUS 2008), where the Rectors' Conference of the Swiss Universities (CRUS, since 1 January 2015 called swissuniversities) is committed to a number of quality principles to guide its quest for university system quality. These principles were formulated on the basis of the CRUS understanding that quality is driven by the following two dimensions: international competition among each university related to specific stakeholder needs and cooperation through complementary specialization and coalition building among Swiss universities. In the long run, these quality principles should contribute to Switzerlands ambition to become a leading place for research, education and knowledge transfer. The project supported accounting for research performance rather than controlling the involved researchers. It also aimed to develop useful tools for the internal quality assessment procedure of Swiss universities according to the guidelines of the Swiss University Conference, devise strategies for Swiss universities and critique academic rankings. The project was successfully finalized by the end of 2012. As of 2013, the 'Performances de la recherche en sciences humaines et sociales' programme is up and running and pursues mainly the same goals as the previous project, but with a more specific focus on

A. Loprieno
University of Basel, Petersgraben 51, 4051 Basel, Switzerland
e-mail: a.loprieno@unibas.ch

R. Werlen · J. Bregy (✉)
swissuniversities, Effingerstr. 15, Postfach, 3000 Bern, Switzerland
e-mail: jaromir.bregy@swissuniversities.ch

R. Werlen
e-mail: raymond.werlen@swissuniversities.ch

A. Hasgall
swissuniversities, c/o University of Geneva, Service Recherche,
24 Rue du Général Dufour, 1211 Geneva, Switzerland
e-mail: alexander.hasgall@unige.ch

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_2

the humanities and social sciences. This project aims to develop instruments that will foster the visibility of research performance by scholars in the humanities and social sciences in terms of highlighting strengths of different research units located at Swiss universities. It will also strengthen a multiplicity-oriented approach to research evaluation, which aims to support the diversity that characterizes research in the social sciences and humanities.

1 Introduction

Although all Swiss universities share a strong focus on research, the effective monitoring of quality academic research has yet to be satisfactorily developed. The 'Mesurer les performances de la recherche' project was an attempt of the Rectors' Conference of the Swiss Universities to identify the best ways for Swiss universities to implement a system of research evaluation according to their specific needs and institutional strategy. The project was funded over the period 2008–2011 through project-related subsidies of the Swiss Confederation allocated by the Swiss University Conference. The project was finalized in 2012 and has since been followed by the 'Performances de la recherche en sciences humaines et sociales' programme, which will be funded from 2013–2016 through project-related subsidies as well. The main focus of this programme is the visibility of research performance and impact in terms of highlighting the quality and strengths of research in different fields and disciplines. In what follows, we will delimit the scope and intended purposes of the project and the programme while addressing the following five questions:

- What should be evaluated in research?
- For what purpose should we evaluate research?
- How should we evaluate research?
- What are the ties between evaluation and quality?
- How can the quality and impact of research be made visible to different stakeholders both within and outside the universities?

We will briefly describe the main features of the project and its results, detail current developments in the on-going programme and then present certain perspectives of swissuniversities on the remaining period of the programme.

2 Making a Variety of Research Visible

2.1 What Should We Evaluate in Research?

Academic research includes a wide array of aspects, from the discovery of new knowledge and promoting young researchers to potential impacts on the scientific community and society. However, the relevance of these aspects to different

stakeholders (universities, faculties, researchers, authorities and the public) varies according to disciplinary and institutional differences. Thus, the ‘Mesurer les performances de la recherche’ project paid particular attention to these differences, not only considering the impact of research evaluation on the scientific community, but also disciplinary diversity, the significance of interdisciplinary research, the interaction between research and teaching, technological innovation, and linguistic and cultural specificities, such as language and the form of publication. Many of these differences—like language and the form of publication—are particularly important in the social sciences and humanities (Huang and Chang 2008; Czellar and Lanares 2013).

Therefore, the understanding that all these aspects should be taken into account in research evaluation is one of the main reasons why the ‘Performances de la recherche en sciences humaines et sociales’ programme focuses on specific research circumstances in the humanities and social sciences.

2.2 For What Purpose Should We Evaluate Research?

The evaluation of research requires different levels of focus depending on whether a given body of research addresses authorities, peers, or the public at large. One important purpose of evaluating research is to make research accountable both to political authorities and the public. In this sense, research evaluation plays a major role in developing and adapting the institutional strategies of Swiss universities. At both the individual and institutional levels, attaining knowledge of research strengths and weaknesses is another crucial purpose of research evaluation. Lastly, research evaluation also serves to make quality and, consequently, the importance of research visible for external stakeholders. While the ‘Mesurer les performances de la recherche’ project explored various possibilities for measuring research performance and compared their scope, the ‘Performances de la recherche en sciences humaines et sociales’ programme fosters the development of instruments to increase the visibility of research performance and impact for the benefit of universities and their faculties.

2.3 How Should We Evaluate Research?

Conventional methods of research evaluation, particularly advanced bibliometric analyses based on the Web of Science or Scopus, both of which are online scientific citation indexing services, are quite useful for describing the impact of research in natural sciences, such as chemistry or medicine, within the scientific community (van Leeuwen 2013; Engels et al. 2012).

However, these methods are less useful for describing the social impact of research in the humanities. The 'Mesurer les performances de la recherche' project encouraged the exploration and the development of broader approaches that may better suit the needs of different disciplines and reflect the impact of other aspects of research, such as its social relevance or its applied uses, including teaching. The 'Performance de la recherche en sciences humaines et sociales' programme builds on the resulting activities of the previous 'Mesurer les performances de la recherche' project in order to develop further methods of evaluating research that will pay greater attention to specific circumstances in the humanities and social sciences, such as linguistic characteristics, informal researcher networks and different forms of publication in the respective disciplines.

2.4 Evaluation, Quality and Mission

As the CRUS points out in 'The Swiss Way to Quality in the Swiss universities' (CRUS 2008), the quality of research is not an end in itself, but rather is at the service of further aims that are derived from each university's self-determined strategy regarding its role in Switzerland and the international community. The CRUS underlines this principle while stressing the following aspects:

1. The CRUS recognizes that member universities are bound by different missions as established by their respective responsible bodies. The CRUS is therefore convinced that each university is responsible for setting its own strategy according to its mission, thereby autonomously determining its role in the Swiss and international university landscape.
2. The CRUS is further convinced that it is best that its member universities themselves determine the body of objective quality criteria that most appropriately fit the deliverables emanating from these strategies. However, no university shall abstain from committing itself to a body of objective quality criteria for its self-chosen deliverables or from communicating them broadly.

As a consequence of these statements, the 'Mesurer les performances de la recherche' project and the 'Performance de la recherche en sciences humaines et sociales' programme have supported accounting for research evaluation rather than controlling the researchers involved. Both the project and the programme have aimed to develop useful tools for internal quality assessment procedures, stakeholder communications and different approaches to deal with rankings and to achieve greater visibility of research performances. For these purposes, a dedicated decentralized network of specialists has been assembled.

3 The 'Mesurer les Performances de la Recherche' Project

Given the considerations mentioned above, the Swiss University Conference decided to finance the 'Mesurer les performances de la recherche' project to achieve three purposes:

- To establish university-based specialists that possess the necessary knowledge in the field of research evaluation.
- To generalize the use of bibliometry in Swiss universities in order to better judge its potential and its limits.
- To develop initiatives and actions for those aspects of research quality and performance that are not covered by conventional bibliometry.

The specialists in research evaluation established at every Swiss university represented a central pillar of the prior project and will remain as actors in the current programme. These specialists are organized within a network that guarantees the exchange of experiences and the diffusion of acquired competences by meeting several times a year.

For a better understanding and a more general use of bibliometry, Swiss universities conducted bibliometric analyses in collaboration with the Centre for Science and Technology Studies (CWTS) of Leiden. The main results of this bibliometry project can be summarized as follows: publications of Swiss universities recorded by the Web of Science are far more frequently cited than the world average. In contrast, research that is not published in the Web of Science, especially in the humanities and (to a lesser extent) in the social sciences, is not yet on the radar and remains largely invisible to the conventional bibliometry (CRUS 2009).

In addition to this bibliometric approach mentioned above, the 'Mesurer les performances de la recherche' project supported the following three peer-reviewed initiatives:

- 'Entwicklung und Erprobung von Qualitätskriterien in den Geisteswissenschaften am Beispiel der Literaturwissenschaften und der Kunstgeschichte [Developing and testing quality criteria for research in the humanities]', Universities of Zurich and Basel.
- 'Measuring Research Output in Communication Sciences and Educational Sciences between international benchmarks, cultural differences and social relevance', Universities of Lugano, Fribourg, Bern and Zurich.
- 'Décrire et mesurer la fécondité de la recherche en sciences humaines et sociales à partir d'études de cas [Describe and measure the fecundity of research in the humanities and social sciences from case studies]', Universities of Neuchatel, Lausanne and Lugano.

These three projects focused on different issues. 'Developing and testing quality criteria for research in the humanities' focused on quality criteria and indicators that researchers in the humanities and social sciences consider important (Hug et al. 2013, 2014; Ochsner et al. 2013, 2014). 'Measuring Research Output in Communication Sciences and Educational Sciences between international benchmarks, cultural

differences and social relevance' studies the different profiles within and between different research institutions in communication sciences (Probst et al. 2011). The project 'Describe and measure the fecundity of research in the humanities and social sciences from case studies' concentrates on making visible the manifold relationships between researchers, institutions and other stakeholders.

Additionally, the project supported four actions to achieve the following:

- Integrate another language into the initiative 'Measuring Research Output in Communication Sciences and Educational Sciences between international benchmarks, cultural differences and social relevance'.
- Organize workshops in an effort to transfer knowledge and experiences developed within the initiatives between the representatives of the involved universities.
- Organize a workshop to measure research performance in the field of law.
- Organize workshops and establish an experimental module on the added value of research assessments.

As the final report of the project (CRUS 2013) points out, the participation of all Swiss universities in the project as well as the development of different and complementary contributions represent the main achievements of the project. Both the participation and contributions of the Swiss universities—as leaders of the initiatives and actions or through participating in the experts network—built the foundation for frequent and constructive exchanges, especially within the specialists network. On the other hand, a number of goals were not fully achieved by the time the project was finalized. The CRUS decided to pursue these remaining goals during the period spanning 2013–2016.

4 The 'Performances de la Recherche en Sciences Humaines et Sociales' Programme

The financial efforts and implemented measures during the financing period 2008–2012 to support the project were not sufficient. The CRUS thus suggested to continue pursuing the goals of the project from 2013 to 2016 in the 'Performances de la recherche en sciences humaines et sociales' programme. This will allow for the sustainable development of competences in research evaluation in universities by allocating project-related subsidies to specialist posts. The launch of the programme also allows for calls for further initiatives with institutional partners that can cover domains and aspects of research not yet covered by the three initiatives of the previous project. The measures of the programme should further promote the development of competences at the national level and enhance international collaboration in the field of research evaluation.

The programme supports seven initiatives that were submitted either by a single university or as the result of collaboration among several universities:

- 'Developing indicators for the usage of research in Communication Sciences. Testing the productive interactions approach', Universities of Fribourg and Lugano
- 'Der Wertbeitrag betriebswirtschaftlicher Forschung in Praxis und Gesellschaft [The impact of economics research]', University of St. Gallen
- 'Scientometrics 2.0: Wissenschaftliche Reputation und Vernetzung [Scientometrics 2.0: academic reputation and networks]', University of St. Gallen
- 'Forschungsevaluation in der Rechtswissenschaft [Research evaluation in law]', Universities of Geneva and Bern
- 'Ressourcen-basiertes Instrument zur Abbildung geisteswissenschaftlicher Forschung am Beispiel der Theologie [Resource-based instrument for documenting and assessing research in the humanities and the social sciences as exemplified by theology]', Universities of Fribourg and Lucerne
- 'Cartographier les réseaux de recherche. Interactions et partenariats en sciences humaines et sociales [Mapping research networks. Interactions and partnerships in social sciences and humanities]', University of Neuchatel
- 'National vergleichbare Daten für die Darstellung und Beurteilung von Forschungsleistungen [Comparable data on national level for the presentation and evaluation of research performance]', University of Basel

As with the previous project, this programme has a special focus on the question of how the diversity concerning the approaches to research and its outcomes can be presented and evaluated appropriately in the context of research evaluation. This includes making visible the manifold interactions and co-operations between researchers and research institutions and the interactions of research institutions in social sciences and humanities with different external stakeholders. The project also investigates how research cultures and the specificities of different disciplines can be taken into account in order to find better ways of evaluating research. Additionally, two projects in law and theology are dedicated to making notions of quality in their disciplines more visible. It will thus also be possible to develop procedures for finding a consensus concerning quality criteria in a particular discipline.

Both programmes together include a total of ten projects. An additional eight so-called 'Implementation Projects' are being funded for the years 2015–2016. The aim of these smaller projects is to transfer the results of the initiatives into different institutional and thematic contexts and to test the applicability of the instruments and sets of indicators, examples of which include the following: Based on the results of the project 'Developing and testing quality criteria for research in the humanities', a rating form is going to be developed at the University of Zurich that serves to assess the research proposals of junior researchers in the humanities. In addition to ensuring a more appropriate evaluation of emerging researchers proposals, this will also demonstrate the potential of broader sets of qualitative indicators for research evaluation. The University of Lausanne is going to use the mapping tool developed in the project 'Describe and measure the fecundity of research in the humanities and social sciences from case studies' for a detailed analysis of this institutions collaborations and partnerships. Based on its own project, 'Scientometrics 2.0' (Hoffmann et al.

2015; Jobmann et al. 2014), the University of St. Gallen is incorporating alternative metrics of research impacts into its own repository.

In addition to the 18 total projects, a network consisting of specialists in bibliometrics and research evaluation from all Swiss universities and the individuals in charge of the different initiatives accompanies the programme. This network will allow for an important transfer of knowledge in a decentralized and university-based landscape. The network meets regularly and also invites national and international experts and representatives of the different stakeholders.

The programme has also received a further boost by hiring a full-time scientific coordinator. Besides coordinating the diverse components of the programme, he is also assigned a variety of additional tasks. He is responsible for the internal and external communication on a national and international level as well as the networking with the different stakeholders. He also elaborates on the synthesis of the results. Part of this synthesis is going to be a manual, which introduces the 'Swiss Way to Quality' and will enable practitioners to profit from the outcomes of the different projects.

Since the project is still ongoing, most of the results have not been published. However, a website (http://www.performances-recherche.ch) provides information about the current state of the project and the contact information of those responsible for the projects. Overall, both the Swiss universities unique approaches to the challenges in the field of research evaluation and the transfer of knowledge through the 'Mesurer les performances de la recherche' project and the 'Performances de la recherche en sciences humaines et sociales' programme represent crucial contributions toward an adequate system of research evaluation in the Swiss landscape of higher education, which is currently going through major changes due to the implementation of the new Federal Act on Funding and Coordination of the Swiss Higher Education Sector planned for 2015. At the same time, the programme is a Swiss contribution to the current research debate about how quality in research can best be evaluated.

References

CRUS. (2008). The Swiss way to quality in the university system. Bern: CRUS.
CRUS. (2009). Projet 'Mesurer les performances de la recherche'—1er Rapport. Bern: CRUS.
CRUS. (2013). Projet de coopération et d'innovation 2008–2011/2012 'Mesurer les performances de la recherche' (PCI B-05). Rapport final. Retreived from http://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/FR/UH/SUK_P-3/Rapport_final_B-05_251013.pdf.
Czellar, J., & Lanarés, J. (2013). Quality of research: Which underlying values? *Scientometrics, 95*(3), 1003–1021. doi:10.1007/s11192-012-0928-x.
Engels, T. C., Ossenblok, T. L., & Spruyt, E. H. (2012). Changing publication patterns in the social sciences and humanities, 2000–2009. *Scientometrics, 93*(2), 373-390. doi:10.1007/s11192-012-0680-2.
Hoffmann, C. P., Lutz, C., & Meckel, M. (2015). A relational altmetric? Network centrality on Research Gate as an indicator of scientific impact. *Journal of the American Society for Information Science and Technology*, online first. doi:10.1002/asi.23423.
Huang, M., & Chang, Y. (2008). Characteristics of research output in social sciences and humanities: From a research evaluation perspective. *Journal of the American Society for Information Science and Technology, 59*(11), 1819–1828. doi:10.1002/asi.20885.
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 for Education Law and Policy, 10*(1), 55–64.
Jobmann, A., Hoffmann, C. P., Künne, S., Peters, I., Schmitz, J., & Wollnik-Korn, G. (2014). Altmetrics for large, multidisciplinary research groups: Comparison of current tools. *Bibliometrie – Praxis und Forschung, 3*. Retrieved from http://www.bibliometrie-pf.de/article/view/205.
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*(4), 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 Supplement), 111–132. doi:10.1007/s11618-014-0576-4.
Probst, C., Lepori, B., De Filippo, D., & Ingenhoff, D. (2011). Profiles and beyond: Constructing consensus on measuring research output in communication sciences. *Research Evaluation, 20*(1), 73–88. doi:10.3152/095820211X12941371876102.
van Leeuwen, T. (2013). Bibliometric research evaluations. Web of Science and the social sciences and humanities: A problematic relationship? *Bibliometrie–Praxis und Forschung, 2*. Retrieved from http://www.bibliometrie-pf.de/article/view/173.

Yes We Should; Research Assessment in the Humanities

Wiljan van den Akker

Abstract In this contribution I argue that the Humanities, just like any other mature field of knowledge, should have or develop a system by which its research can be assessed. In a world that increasingly asks for justification of public funds, where public money becomes scarcer, so that less amounts have to be distributed among more players, where research funds are being concentrated and distributed on a highly competitive basis, we as humanists cannot shy away from research assessment with the argument that 'we are different from the rest' or that 'we don't need it'. Of course the humanities are a distinct member of the body of academic knowledge, but that holds true for every discipline. If we agree that for instance that bibliometry does not suit most players in our field, the question becomes: what will suit us better? Case-studies? This contribution also contains a warning: let us stop arguing about the language issue. English is the modern Latin of academia and its use enables us to communicate with one another, wherever we are or who we are. Without providing definite solutions, my argument is that we, humanists, should take the steering wheel ourselves in developing adequate forms of research assessment. If we leave it to others, the humanities will look like arms attached to a foot.

Suppose that I have learned something during the more than 25 years I am working within the humanities now—as a teacher, a researcher, a director and a dean. The attitude of my field towards research-assessment in any form, can be summed up as follows. 'We don't want it, because we don't have to, because we don't need it, because we are not like the others, and therefore we don't like it, and they shouldn't force us, because they don't know us, because they don't understand us, because they don't love us.' The image of the humanist working in solitude in the attic, writing a book that will replace all existing books and render superfluous all books that have not yet been written, is still alive and kicking.

The humanities have developed several defense-mechanisms against research assessment in general. I will name only three of them.

W. van den Akker
Utrecht University, Kromme Nieuwegracht 46, 3512 HJ Utrecht, Netherlands
e-mail: W.J.vandenAkker@uu.nl

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_3

1. The (much heard) argument of intuition: the quality of our research is not measurable, not quantifiable. We know quality when we see it. We have a perfect understanding of who is excellent and who is not. It is easy to see that although this argument may be (sometimes) true, it is also highly irrelevant. In fact, one could turn it around and say that this should make research assessment a lot easier, also the production of the top ten or top hundred. Anyone who has ever dared to ask such a question, knows that it equals a declaration of war.
2. A second mechanism is: the humanities as a whole are principally and practically completely different from all the other forms of science or knowledge fields, especially *the hard sciences*. But this is not true. There is not one common denominator that separates the humanities from the other academic fields. In fact the humanities are made of different disciplines and fields who hold their own positions within academia. Some are very familiar to fields like theoretical physics, like for instance linguistics. Others are close to social sciences, like for instance large parts of the historical disciplines. Some philosophers claim the same domain as mathematics.
3. The third defense mechanism mirrors the second: since there is no such identifiable and unifiable one thing as *the* humanities, since we are a habitat of different species, it is impossible to compare us to other parts of the body of knowledge. Again it is not a strong argument, since the same holds true for what we generally call the *(hard) sciences*medicinetechnical sciences, and so on and so forth. Think of the social sciences where the anthropological and the empirical approaches are totally different.

All these defense mechanisms are not effective for today's world and especially not for the future of the humanities. We cannot and should not insist on being 'different' just to shy away from any form of research assessment. If we continue doing that, we will be the young sister or brother who is tolerated at the dining table, at the mercy of the food that the rest of the family thinks it can spare and always looked down upon. Maybe with a friendly smile, but nevertheless.

In the near future, in a world that increasingly asks for justification of public funds, in a world where at the same time public money becomes scarcer and less amounts have to be distributed among more players, in a world where research funds are being concentrated and distributed on a highly competitive basis, we as humanists have to take the stand and declare that we are grownups who want to play the game.

Maybe our defense mechanisms were never effective in the past anyway, but the older brothers and sisters just left us alone, which could be one of the reasons that the humanities are underfunded in general, not only in research but especially in teaching. In that case we already have shot ourselves in the foot and it becomes a matter of healing as quickly as possible in order to be able to kick again real hard.

If we are not *essentially* different from other fields of academia, we also should recognize that, just like the other members of the family, we are not simple. It is clear that in discussing research assessment within the humanities, we are dealing with a complicated matter, complicated in the sense of a *complex* of several parameters, angles, similarities, issues etc. Just to name seven aspects:

1. There are substantial differences in scientific *practice* between the several disciplines within the humanities. These differences can and will have consequences for the selection of quality indicators. There are areas where groups of scholars work together on a common project—say the testing of a theory—and therefore they publish together in journals and an analysis of citations can or will be useful. In other areas individuals work on diverse topics and therefore publish individually and therefore an analysis of citations can be less useful.
2. The rotation time of humanities articles and books. Contrary to many other fields of science, much of what we humanists produce can have an effect in the long(er) run. Consider the fact that much research in for instance medicine will be outdated within 2 or 3 years, or perhaps even sooner.
3. The goals and products of research are different in different areas of the humanities. Unlike scholars in, say theoretical physics, much research in the humanities has the intention and maybe even the assignment by society to guard, disclose, save and interpret international and/or national heritage. Even though not all scholars like it or accept it, society in general often looks at us in this way. If *we* don't do it, who else will? This means that the products of such research will and cannot be seen only in terms of articles in scientific journals, but for instance also in the construction of large databases and the opening up of large data collections, exhibitions with catalogues, excavations of archeological sites etc. Think of the endless amounts of historical or cultural material lying in archives, museums, libraries. Data collections, also including books, are for the humanities the laboratories that make the work of our relatives in the *sciences* so expensive.
4. As a consequence the target group of the humanities is diverse. On the one hand, like in any other scientific field, our accumulation of knowledge is targeted at our peers, on the other hand we have a large, non-academic audience to serve. One of the problems scholars in the humanities face, is to define this wider group and to justify our relations with it. What astronomers perhaps would see as *translation* of scientific knowledge, and therefore regard as journalistic of the profession, is for many humanists core business. But not always, and there we have an immense problem to solve. To be quite clear, I don't have the answer, but I do think a possible solution lies within the realm of peer review.
5. All this shows that the publication channels of the humanities will vary. In some fields *traditional* books are still the main or even the only accepted way to transfer our knowledge, like in many parts of history or literary studies. In some areas, however, articles in journals have replaced the more traditional book, like in linguistics. There, books are mainly written in order to popularize knowledge or to use in classrooms for teaching purposes.
6. A highly inflammable aspect related to all this, is the language of our scholarly work. Inflammable because often there is a nationalistic side in the discussion, even when it is hidden and not explicitly mentioned. The argument mostly goes like this: since my scholarly object is Dutch poetry, I cannot but write about it in Dutch. Because of the linguistic nature of the field of study, there have to be journals in a language other than English. Tied to this is the more *sentimental* reasoning: a country like The Netherlands has its own cultural heritage and acad-

emia should honor the uniqueness of it, by allowing high quality scholarly work in Dutch.

Of course anyone can substitute Hungary or Switzerland for The Netherlands. Following this line, someone writing about Polish novels in Dutch, would not contribute to science, someone writing on the same subject in Polish on the other hand would. I am not convinced that this line of reasoning is strong enough but I also realize that my counter arguments are disputable and will be disputed.

First of all it is a mistake to think that most scholarly work is written in English. It looks and sounds like English but it is not. It is at the best Scholarly English, like Latin was centuries ago. The Latin those colleagues back then wrote and spoke in no way resembled the Latin from the Romans, as any specialist can confirm. It was agreed upon as the *lingua franca* of science, a fantastic way to communicate all over the world, regardless of one's country of origin and mother tongue. Seen from this point of view, there is no valuable reason why a scholar whose object is Dutch poetry should prevent the rest of the world to read his or her results by writing in Dutch about it. Why has the language of the object of research anything to do with the language in which we scholarly communicate about it? The mere fact that only a small part of the wide world is interested in Dutch poetry and a large part does not even know it exists at all, is totally irrelevant. Moreover: writing only in Dutch about Dutch poetry, will be absolutely the best guarantee that the world stays ignorant about the subject.

In the meantime there is a counterargument. Anyone who wants to work on a field that is specifically Dutch has to master the Dutch language. If not, all necessary documentary sources—the primary object of research—will not be accessible and stay unknown. Some examples can be found by looking at some of the most excellent American colleagues. Margaret Jacob for instance, a distinguished professor of history at UCLA, learned how to read Dutch, because she is interested in the field of European Enlightenment. She cannot write Dutch nor have scholarly conversations in Dutch, but she knows how to read the sources. Her books and articles are written in English though. And as a consequence, the Dutch influence on what was generally regarded as an Anglo-French movement, could be acknowledged.

Nationalism is a killer in the world of science, also in the humanities. My example is Dutch and therefore humble. But if I were French or German, I would say the same. Again, I am saying this in full awareness of the new nationalism that spreads its bad seeds all over Europe.

7. The final aspect is the level of organization within the humanities or maybe better formulated: the lack of it. If one still thinks of the humanities as a collection of individuals writing individual books, then there is absolutely no need whatsoever to have an internal or external form of organization. But if one agrees that this image of the humanities is no longer true or only partially true, organization becomes a substantial factor. Again the problem is that we are talking about something highly complex. Because there are several fields where scholars could—and to my opinion should—be better organized. Within the discipline or sub-discipline, within the managerial organization (departments, schools, research institutes, fac-

ulties of humanities), the national endowment organizations of the humanities, the European Science Foundation and/or the European Research Council.

To make a shortcut: we, humanists, are not well organized. Look at the astronomers. The amounts of public money that flows in their direction are not matched with any economic or social outcome at all. Only a few days ago one of the headlines in the Dutch media was the discovery of a new solar system thirteen billion lightyears away from us. The last known solar system is only 12.9 billion lightyears away. Experts said the discovery is of the highest importance. Why? They didn't tell. They almost never do. We speak about 'An Astronomous Amount'. Imagine we would speak of a 'Humanist Amount of Money'. Apart from many other reasons, the astronomers are extremely well organized. That is to say: they fight most of their paradigmatic battles inside their home, with the door shut, the windows closed and the curtains down. When they come outside, they are all astronomers in clean suits. *Nature* and *Science* are full of their latest discoveries and they have armies of well-trained scholars who are able and paid to translate the most obscure particles of new knowledge to a broader audience. They have agreed upon an excellent division of labor: doing this in one country, and that in the other. I always wondered why astronomy was such a big thing in The Netherlands: a country that the sun hates profoundly. They work on their research individually and at the same time in small and large groups. Fifteen years ago the Dutch government announced that a limited amount of research proposals could be awarded a large sum of money. The astronomers won by a landslide. Their proposal was written by a journalist and was called *Unraveling the Universe*. Can you imagine? Newspapers all over the world: 'Dutch unravel Universe!'

With regard to the humanities, there are fields that are highly successfull and well organized at the same time. Like archeology, but even more so linguistics and parts of history, especially social-economic history. If one takes linguistics: the domain is torn apart by fighting paradigms. Syntax, semantics, phonetics, neurocognition, Chomsky or not Chomsky. But they are well organized, share the same publication platforms, have their recognized international conferences, are willing to work on interdisciplinary projects—just think of neurolinguistics and the impact on questions of speech impediment over the last decade. It cannot be a coincidence that this part of the humanities is already working with laboratories and large data collections. Linguistics was recently put on the ESFRI-list, the European Roadmap for large scientific infrastructure.

Should we all copy linguistics? Of course not. But we should look from a more abstract point of view at the process of organization. We should start working at several levels at a time. At the lowest level, begin to look at the field of a discipline or of a group of disciplines. Let's say Literary Studies, to stick to my own academic field. At the same time maybe we should organize the process of research assessment on a national level, like Norway, Denmark and Belgium are doing. Of course benchmarking is one of the necessary factors, but in this way we could avoid sinking to the bottom immediately. I really am convinced that Germany is doing the right thing in selecting a limited number of universities and labeling them as research universities

and subsequently giving them proportionate more amounts of money. Of course one can criticize the criteria, but still.

I think that we as humanists do not prepare ourselves well enough for the future if we continue to put our research on the website only at the level of individual faculty members. We should have more research projects, more research institutes within the universities and not outside university. We should definitely stop telling the world that we are *different*. Research assessment is a complicated thing, not in the sense of too difficult or impossible, but in the sense of complex. Let's take all the different parameters into account, let's take time but move on. But the most important thing is: let's take or keep the lead.

Two years ago in The Netherlands a nationwide project started called *Sustainable Humanities*. It is a plea for more money for the Humanities. But not a traditional plea bargain in the sense of: o, world, look at those poor exotic disciplines, see how they are withering like beautiful flowers blossoming for the last time all alone in the desert with no water. On the contrary. The statement is: look at the enormous quantities of students in media studies, in history, in communication, see how our staff-student-ratio does not even come close to that of high schools. Many university professors in the humanities have such a heavy teaching load that it becomes almost impossible to do serious research. Look at our *Nachwuchs*: the ridiculous small amount of Ph.D. and Postdoc positions.

The project also contains a call to the Humanities itself to start a nationwide process of research assessment. To quote the report:

> In addition to peer review, international assessment of research increasingly makes use of bibliometric instruments such as citation indexes and impact factors. These are parameters which can be used in science, technology and medicine. But it is now widely acknowledged—also internationally—that these instruments are not necessarily suitable for determining the quality of research in the humanities. For example, in 2000 the European Science Foundation (ESF) concluded that the Arts and Humanities Citation Index (AHCI) and the Science Citation Index of the ISI (Institute for Scientific Information, Philadelphia) should not be used by policy makers in Europe. For the humanities these indexes are notoriously unreliable because of the predominance of English-language literature—particularly literature published in the United States—and because of the fact that books are not included in them. The European Reference Index for the Humanities (ERIH) which has since been developed under the auspices of the ESF has certainly not yet been operationalized to the point that it fills this gap. The problem is not so much that proper quality determination is impossible in the humanities. What is missing is an effective instrument that can take the specific character of humanities research into account while measuring quality across an academic field. Because of the special character of these subjects, the benchmarks used to assess them must always be special as well. The fact that relatively few *prizes* are awarded in this domain aggravates this lack of indicators and makes it even more difficult for outsiders to judge the quality of research (and researchers) in the humanities. Much too often this causes serious problems for top-ranking scholars in the humanities. (Committee on the National Plan for the Future of the Humanities 2009, p. 34)

Therefore the Dutch Royal Academy of Arts and Sciences has taken up the challenge and published a national report on research assessment within the humanities (Royal Netherlands Academy of Arts and Sciences 2011).

The recognition of the humanities as a distinct member of the body of academic knowledge, leads to the conclusion that humanists should take the steering wheel in developing adequate forms of research assessment themselves. If we leave it to others, the humanities will look like arms attached to the feet.

References

Committee on the National Plan for the Future of the Humanities. (2009). *Sustainable humanities: Report from the committee on the national plan for the future of the humanities*. Amsterdam: Amsterdam University Press. http://www.oapen.org/search?identifier=339995.

Royal Netherlands Academy of Arts and Sciences. (2011). *Quality indicators for research in the humanities*. Amsterdam: Royal Netherlands Academy of Arts and Sciences. https://www.knaw.nl/shared/resources/actueel/publicaties/pdf/quality-indicators-for-research-in-the-humanities.

How Quality Is Recognized by Peer Review Panels: The Case of the Humanities

Michèle Lamont and Joshua Guetzkow

Abstract This paper summarizes key findings of our research on peer review, which challenge the separation between cognitive and non-cognitive aspects of evaluation. Here we highlight some of the key findings from this research and discuss its relevance for understanding academic evaluation in the humanities. We summarize the role of informal rules, the impact of evaluation settings on rules, definitions of originality, and comparisons between the humanities, the social sciences and history. Taken together, the findings summarized here suggest a research agenda for developing a better empirical understanding of the specific characteristics of peer review evaluation in the humanities as compared to other disciplinary clusters.

1 Introduction

In *How Professors Think* (2009), Michèle Lamont draws on in-depth analyses of five fellowship competitions in the United States to analyse the intersubjective understandings academic experts create and maintain in making collective judgments on research quality. She analyses the social conditions that lead panelists to an understanding of their choices as fair and legitimate, and to a belief that they are able to identify the best and less good proposals. The book contests the common notion that one can separate cognitive from non-cognitive aspects of evaluation and describes the evaluative process as deeply interactional, emotional and cognitive, and as mobilizing the self-concept of evaluators as much as their expertise. Studies of the internal functioning of peer review reveal various 'intrinsic biases' in peer review like

M. Lamont (✉)
Department of Sociology, Harvard University, 510 William James Hall,
33 Kirkland Street, Cambridge, MA 02138, USA
e-mail: mlamont@wjh.harvard.edu

J. Guetzkow
Department of Sociology and Anthropology, Hebrew University of Jerusalem,
Mount Scopus, 91905 Jerusalem, Israel
e-mail: joshg@mscc.huji.ac.il

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_4

'cognitive particularism' (Travis and Collins 1991), 'favouritism for the familiar' (Porter and Rossini 1985), or 'peer bias' (Chubin and Hackett 1990; Fuller 2002).

These effects show that peer review is not a socially disembedded, quality-assessing process in which a set of objective criteria is applied consistently by various reviewers. In fact, the particular cognitive and professional lenses through which evaluators understand proposals necessarily shape evaluation. It is in this context that the informal rules peer reviewers follow become important, as are the lenses through which they understand proposals and the emotions they invest in particular topics and research styles. Thus, instead of contrasting 'biased' and 'unbiased' evaluation, the book aims to capture how evaluation unfolds, as it is carried out and understood by emotional, cognitive and social beings who necessarily interact with the world through specific frames, narratives and conventions, but who nevertheless develop expert views concerning what defines legitimate and illegitimate assessments, as well as excellent and less stellar research.

How Professors Think concerns evaluation in multidisciplinary panels in the social sciences and the humanities. It examines evaluation in a number of disciplines and compares the distinctive 'evaluative cultures' of fields such as history, philosophy and literary studies with those of anthropology, political science and economics. This paper first describes some of the findings from this study. Second, summarizing Lamont and Huutoniemi (2011), it compares the findings of *How Professors Think* with a parallel study that considers peer review at the Finish Academy of Science. These panels are set up somewhat differently from those considered by Lamont—for instance focusing on the sciences instead of the social sciences and the humanities, or being unidisciplinary rather than multidisciplinary. Thus we discuss how the structure of panels affects their functioning across fields. Finally, drawing on Guetzkow et al. (2004), we revisit aspects of the specificity of evaluation in the humanities, and more specifically, the assessment of originality in these fields. Thus, this paper contributes to a better understanding of the distinctive challenges raised by peer review in the humanities.

2 The Role of Informal Rules

Lamont interviews academic professionals serving on peer review panels that evaluate fellowship or grant proposals. During the interviews, panelists are asked to describe the arguments they made about a range of proposals, to contrast their arguments with those of other panelists, and to explain what happened in each case. Throughout the interviews, she asks panelists to put themselves in the role of privileged informer and to explain to us how 'it' works. They are encouraged to take on the role of the native describing to the observer the rules of the universe in which they operate. She also has access to the preliminary evaluations produced before panel deliberations by individual panelists and to the list of awards given.

Since *How Professors Think* came out, it has been debated within various academic communities, as it takes on several aspects of the evaluation in multidisciplinary panels in the social sciences and humanities. It is based on an analysis of twelve funding panels organized by important national funding competitions in the U.S.: those of the Social Science Research Council, the American Council for Learned Societies, the Woodrow Wilson Fellowship Foundation, a Society of Fellows at an Ivy League university and an important social science foundation in the social sciences. It draws on 81 interviews with panelists and program officers, as well as on observation of three panels.

A first substantive chapter describes how panels are organized. A second one concerns the evaluative culture of various disciplines, ranging from philosophy to literary studies, history, political science and economics. A third chapter considers how multidisciplinary panels reach consensus despite variations in disciplinary evaluative cultures. This is followed by two chapters that focus on criteria of evaluation. One analyses the formal criteria of evaluation provided by the funding organization to panelists (originality, significance, feasibility, etc.) as well as informal criteria (elegance, display of cultural capital, fit between theory and data, etc.). The following chapter considers how cognitive criteria are meshed with extra-cognitive ones (having to do with diversity and interdisciplinarity), finding that institutional and disciplinary diversity loom much larger than gender and racial diversity in decision making. A concluding chapter considers the implications of the study of evaluation cultures across national contexts, including in Europe.

The book is concerned not only with disciplinary compromise, but also with the pragmatic rules that panelists say they abide by, which lead them to believe that the process is fair (this belief is shared by the vast majority of academics interviewed). *How Professors Think* details a range of rules, which include for instance the notion that one should defer to expertise, and that methodological pluralism should be respected.

3 The Impact of Evaluation Settings on Rules

In an article with Katri Huutoniemi, Lamont explores whether these customary rules apply across contexts, and how they vary with how panels are set up. Their paper, 'Comparing Customary Rules of Fairness', (Lamont and Huutoniemi 2011) is based on a dialogue between *How Professors Think* and a parallel study conducted by Huutoniemi of the four panels organized by the Academy of Finland. These panels concern: Social Sciences; Environment and Society; Environmental Sciences; and Environmental Ecology. This analysis is explicitly concerned with the effects of the mix of panelist expertise on how customary rules are enacted. The idea is to compare panels with varying degrees of specialization (unidisciplinary vs. multidisciplinary panels) and with different kinds of expertise (specialist experts vs. generalists). However, in the course of comparing results from the two studies, other points of comparison beyond expert composition emerge—whether panelists 'rate' or 'rank'

proposals, have an advisory or decisional role, come from the social sciences and humanities fields or from more scientific fields, etc. The exploratory analysis points to some important similarities and differences in the internal dynamics of evaluative practices that have gone unnoticed to date and that shed light on how evaluative settings enable and constrain various types of evaluative conventions.

Among the most salient customary rules of evaluation, deferring to expertise and respecting disciplinary sovereignty manifest themselves differently based on the degree of specialization of panels: there is less deference in unidisciplinary panels where the expertise of panelists more often overlap. Overlapping expertise makes it more difficult for any one panelist to convince others of the value of a proposal when opinions differ; unlike in multidisciplinary panels, insisting on sovereignty would conflict with scientific authority. There is also less respect for disciplinary sovereignty in panels composed of generalists rather than experts specialized in particular disciplines and in panels concerned with topics such as Environment and Society that are of interest to wider audiences. In such panels, there is more explicit reference to general arguments and to the role of intuition in grounding decision-making.

While there is a rule against the conspicuous display of alliances across all panels, strategic voting and so-called 'horse-trading' appear to be less frequent in panels that 'rate' as opposed to 'rank' proposals and in those that have an advisory as opposed to a decisional role. The evaluative technique imposed by the funding agency thus influences the behaviour of panelists. Moreover, the customary rules of methodological pluralism and cognitive contextualism (Mallard et al. 2009) are more salient in the humanities and social science panels than they are in the pure and applied science panels, where disciplinary identities may be unified around the notion of scientific consensus, including the definition of shared indicators of quality. Finally, a concern for the use of consistent criteria and the bracketing of idiosyncratic taste is more salient in the sciences than in the social sciences and humanities, due in part to the fact that in the latter disciplines evaluators may be more aware of the role played by (inter)subjectivity in the evaluation process. While the analogy of democratic deliberation appears to describe well the work of the social sciences and humanities panels, the science panels may be best described as functioning as a court of justice, where panel members present a case to a jury.

The customary rules of fairness are part of 'epistemic cultures' (Knorr-Cetina 1999) and essential to the process of collective attribution of significance. In this context, considering reasons offered for disagreement, how those are negotiated, as well as how panelists interpret agreement is crucial to capture fairness as a collective accomplishment. Together, these studies demonstrate the necessity for more comparative studies of evaluative processes and evaluative culture. This remains a largely unexplored but promising aspect of the field of higher education, especially in a context where European research organizations and universities aim to standardize evaluative practices.

4 Defining Originality

We now turn to a closer examination of forms of originality scholars from different disciplines tend to favour, with a focus on contrasting the social sciences and the humanities. As described in Guetzkow et al. (2004), we construct a semi-inductive typology of originality. We use this typology to classify panelists' statements about the originality of scholarship, whether it is in reference to a proposal, the panelists' own work, their students' work, or that of someone whose work they admire. The typology is anchored in five broad categories. These categories concern which aspect of the work respondents describe as being original. They include the research topic, the theory used, the method employed, the data on which it is based and the results of the research (i.e. what was 'discovered'). It also includes two categories that have not been noted in previous research: 'original approach' (explained below) and 'under-studied area' (proposals set in a neglected time period or geographical region). As shown in Table 1, there are seven mutually exclusive categories of originality regarding the approach, under-studied area, topic, theory, method, data, and results.

Each of these generic categories consists of more specific types of originality, which are included in Table 1. Whereas 'Generic Types' refer to which aspects of the proposal are original, 'Specific Types' describe the way in which that aspect is original. Where applicable, the first specific type we list under each generic category refers to the most literal meaning that panelists attribute to this generic category, followed by other specific types in order of frequency. For instance, the first specific type for the generic category 'original approach' is 'new approach' and the other specific types are more particular, such as asking a 'new question', offering a 'new perspective', taking 'a new approach to tired or trendy topics', using 'an approach that makes new connections', making a 'new argument', or using an 'innovative approach for the discipline'. Table 1 also describes the distribution of the 217 mentions of originality we identify across the seven generic categories and their specific types.

Table 1 shows that the panelists we interviewed most frequently describe originality in terms of 'original approach'. This generic category covers nearly one third of all the mentions of originality made by the panelists commenting on proposals or on academic excellence more generally. Other generic categories panelists often use are 'original topic' (15 %), 'original method' (12 %) and 'original data' (13 %). Originality that involves an 'under-studied area' is mentioned only 6 % of the time.

5 What Is an Original Approach?

Previous research on the topic of peer review has not uncovered the category we refer to as 'original approach', and yet it appears that panelists place the greatest importance on this form of originality. But what is it, and how does it differ from original theory or method? 'Original approach' is used to code the panelists' comments on the novelty of the 'approach' or the 'perspective' adopted by a proposal, or on the innovativeness of the questions or arguments it formulates. In contrast to

Table 1 Typology of originality

Generic types	Specific types							Total
Original approach	New approach	New question	New perspective	New approach to tired/trendy topic	New connections	New argument	Innovative for discipline	
	5	21	11	10	8	6	6	67
	7 %	31 %	16 %	15 %	12 %	9 %	9 %	100 %
Under-studied area	Under-studied region	Under-studied period						
	7	6						13
	54 %	46 %						100 %
Original topic	New topic	Non-canonical	Unconventional					
	9	20	3					32
	28 %	63 %	9 %					100 %
Original theory	New theory	Connecting/ Mapping ideas	Synthesis of literature	New application of existing theory	Recon-ceptualizing	Unconventional use of theory		
	5	12	12	5	4	2		40
	13 %	30 %	30 %	13 %	10 %	5 %		100 %
Original method	Innovative method or research design	Synthesis of methods	New use of old data	Resolve old question/debate	Innovative for discipline			
	5	10	7	3	2			27
	19 %	37 %	26 %	11 %	7 %			100 %
Original data	New data	Multiple sources	Non-canonical					
	15	10	4					29
	52 %	34 %	14 %					100 %
Original results	New insights	New findings						
	5	4						9
	56 %	44 %						100 %

Note Some rows may not sum to 100 % due to rounding

original theory or method, an 'original approach' refers to originality at a greater level of generality: the comments of panelists concern the project's meta-theoretical positioning, or else the broader direction of the analysis rather than the specifics of method or research design. Thus in speaking of a project that takes a new approach in her discipline, an art historian applauds the originality of a study that is going to 'deal with [ancient Arabic] writing as a tool of social historical cultural analysis'. She is concerned with the innovativeness of the overall project, rather than with specific theories or methodological details. Whereas discussions of theories and methods start from a problem or issue or concept that has already been constructed, discussions of new approaches pertain to the *construction of problems* rather than to the theories and methodological approach used to study them. When describing a new approach, panelists refer to the proposals' 'perspective', 'angle', 'framing', 'points of emphasis', 'questions', or to their 'take' or 'view' on things, as well as their 'approach'. Thus a scholar in Women's Studies talks of the 'importance of looking at [Poe] from a feminist perspective'; a political scientist remarks on a proposal that has 'an outsider's perspective and is therefore able to sort of have a unique take on the subject'; a philosopher describes his work as 'developing familiar positions in new ways and with new points of emphasis and detail'; and an historian expresses admiration for an applicant because 'she was asking really interesting and sort of new questions, and she was asking them precisely because she was framing [them] around this problem of the ethics of [empathy]'. That 'original approach' is used much more often than 'original theory' to discuss originality strongly suggests a need to expand our understanding of how originality is defined, especially when considering research in the humanities and history, because the original approach is much more central to evaluation of research in these disciplines than in the social sciences, as we will soon see.

6 Comparing the Humanities, History and the Social Sciences

Can we detect disciplinary variations in the categories of originality that reviewers use when assessing the quality of grant proposals? We address this question only at the level of generic categories of originality, because the specific types include too few cases to examine disciplinary variation. For the purpose of our analysis we compare the generic categories of originality referred to by humanists, social scientists and historians.

Table 2 shows aggregate differences in the use of generic types of originality across disciplines and disciplinary clusters. A chi-square test ($\chi^2 = 34.23$ on 12 $d.f.$) indicates significant differences between the disciplines in the way they define originality at a high level of confidence ($p < 0.001$). The main finding is that a much larger percentage of humanists and historians than social scientists define originality in terms of the use of an original approach (with respectively 33 %, 43 % and 18 % of the

Table 2 Generic definitions of originality by disciplinary cluster

Originality type	Humanities		History		Social sciences		All disciplines	
	N	%	N	%	N	%	N	%
Approach	29	33	26	43	12	18	67	31
Data	19	21	6	10	4	6	29	13
Theory	16	18	11	18	13	19	40	18
Topic	13	15	6	10	13	19	32	15
Method	4	4	5	8	18	27	27	12
Outcome	3	3	4	7	2	3	9	4
Under-studied area	5	6	3	5	5	7	13	6
All generic types	89	100	61	100	67	100	217	100

Note Some rows may not sum to 100 % due to rounding

panelists referring to this category). Humanities scholars are also more likely than social scientists and historians to define originality in reference to the use of original 'data' (which ranges from literary texts to photographs to musical scores). Twenty-one percent of them refer to this category, as opposed to 10 % of the historians and 6 % of the social scientists. Another important finding is that humanists and historians are less likely than social scientists to define originality in terms of method (with 4 %, 8 % and 27 % referring to this category, respectively). Moreover humanists, and to a greater extent, historians, clearly privilege one type of originality—originality in approach—which they use 33 % and 43 % of the time, respectively. In contrast, social scientists appear to have a slightly more diversified understanding of what originality consists of, in that they privilege to approximately the same degree originality in approach (used by 18 % of the panelists in this category), topic (19 %) and theory (19 %), with a slight emphasis on method (27 %).

This suggests clearly that the scholars from our three categories privilege different dimensions of originality when evaluating proposals: humanists value the use of an original approach and new data most frequently; historians privilege original approaches above all other forms of originality; while social scientists emphasize the use of a new method. But this comparison is couched at a level of abstraction that allows us to compare these disciplinary clusters according to categories like 'approach', 'data' and 'methods'. This risks masking a deeper level of difference between the meaning of these categories for the social sciences, humanities and history. For example, when social scientists we interviewed refer to original 'data', they generally mean quantitative datasets; historians usually refer to archival documents and use the word 'evidence'; humanities scholars typically refer to written texts, paintings, photos, film, or music and often use words like 'text' and 'materials' to refer to the proposal's 'data'.

Likewise, there are sometimes distinct ways in which humanists and social scientists talk about taking a new approach. For example, humanists will often refer to a canonical text or author that is being approached in a way that is not novel *per se*, but is novel because nobody has approached that author or text in that way (e.g.

a feminist approach to Albert Camus). In contrast, social scientists rarely refer to novelty with regard to something that is 'canonical'. Relatively few social scientists describe originality in terms of approach and those who do so talk overwhelmingly in terms of 'new questions' (accounting for 8 out of 12 social science mentions of original approach). References to original approaches by historians and humanists are spread more evenly across the specific subtypes of 'original approach'. One third of humanists (8 out of 27) define it in terms of taking a 'new approach to an old/canonical topic', but refer to all the other types with nearly equal frequency. And although historians mention 'new questions' more than any other specific type of approach (32 % or 9 out of 28), they often mention other specific types as well. And, although we define 'methods' broadly to categorize the way that humanists, social scientists and historians describe original uses of data, this should not be taken to mean that 'method' means the same thing to all of them. Social scientists sometimes describe innovative methods as those which would answer 'unresolved' questions and debates (e.g. the question of why the U.S. does not have corporatism), whereas humanists and historians never mention this as a facet of methodological originality. Reviewers in the social sciences tend to refer to more methodological detail than others concerning, say, a research design. For instance, a political scientist says that an applicant 'inserted a comparative dimension into [his proposal] in a way that was pretty ingenious, looking at regional variation across precincts'. In contrast, an historian describes vaguely someone as 'read[ing] against the grain of the archives' and an English scholar enthuses about how one applicant was going to 'synthesize legal research and ethnographic study and history of art', without saying anything more specific about the details of this methodological mélange.

Arguably, the differences we find are linked to the distinct rhetorics (Bazerman 1981; Fahnestock and Secor 1991; Kaufer and Geisler 1989; MacDonnald 1994) and epistemological cultures (Knorr-Cetina 1999) of the different disciplines. We do not wish to make sweeping generalizations about the individual disciplines that compose each cluster. However, research on the distinct modes of knowledge-making in some of their constituent disciplines can inform the patterns we find.

In her comparison of English, history and psychology, MacDonnald (1994) shows that generalizations in English tend to be more text-driven than in the social sciences, which tend to pursue concept-driven generalizations. History is pulled in both directions (also see Novick 1988). In text-driven disciplines, the author begins with a text, which 'drives the development of interpretive abstractions based on it'. In contrast, with conceptually driven generalization, researchers design research 'in order to make progress toward answering specific conceptual questions' (MacDonnald 1994, p. 37). These insights map well onto our findings: original data excites humanities scholars because it opens new opportunities for interpretation. Social scientists value most original methods and research designs, because they hold the promise of informing new theoretical questions. The humanists' and historians' emphasis on original approaches is an indication that, while they are not as focused on the production of new generalized explanations ('original theories') or on innovative ways of answering conceptual questions ('original methods'), they value an 'original approach' that enables the researcher to study a text or an archive in a way that

will yield novel interpretations, but which does not necessarily aim at answering specific conceptual questions.

7 Conclusion

Together, the publications summarized in this paper suggest a research agenda for developing a better empirical understanding of the specific characteristics of peer review evaluation in the humanities as compared to other disciplinary clusters. More needs to do be done in order to fully investigate how the composition of panels and the disciplines of their members influence the customary rules of evaluation as well as the meanings associated with the criteria of evaluation and the relative weight put on them.

The comparative empirical study of evaluative cultures is a topic that remains in its infancy. Our hope is that this short synthetic paper, along with other publications which adopt a similar approach, will serve as an invitation to other scholars to pursue further this line of inquiry. More information is needed before we can draw clear and definite conclusions about the specific challenges of evaluating scholarship in the humanities. However, we already know that the role of connoisseurship and the ability to make fine distinctions is crucial given the centrality of 'new approaches' as a criterion for evaluating originality. Whether and how bibliometric methods can capture the real payoff of this type of original contribution is only one of the many burning topics that urgently deserve more thorough exploration.

References

Bazerman, C. (1981). What written knowledge does: Three examples of academic discourse. *Philosophy of the Social Sciences*, *11*(3), 361–388. doi:10.1177/004839318101100305.

Chubin, D., & Hackett, E. (1990). *Peerless science: Peer review and U.S. science policy*. Albany, NY: State University of New York Press.

Fahnestock, J., & Secor, M. (1991). The rhetoric of literary criticism. In C. Bazerman & J. Paradis (Eds.), *Textual dynamics of the professions: Historical and contemporary studies of writing in professional communities* (pp. 76–96). Madison, WI: University of Winsconsin Press.

Fuller, S. (2002). *Knowledge management foundations*. Woburn, MA: Butterworth-Heinemann.

Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is originality in the social sciences and the humanities? *American Sociological Review*, *69*(2), 190–212. doi:10.1177/000312240406900203.

Kaufer, D. S., & Geisler, C. (1989). Novelty in academic writing. *Written Communication*, *6*(3), 286–311. doi:10.1177/0741088389006003003.

Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.

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

Lamont, M., & Huutoniemi, K. (2011). Comparing customary rules of fairness: Evidence of evaluative practices in peer review panels. In C. Camic, N. Gross, & M. Lamont (Eds.), *Social science in the making* (pp. 209–232). Chicago, IL: University of Chicago.

MacDonnald, S. P. (1994). *Professional academic writing in the humanities and social sciences*. Carbondale, IL: Southern Illinois University Press.

Mallard, G., Lamont, M., & Guetzkow, J. (2009). Fairness as appropriateness: Negotiating epistemological differences in peer review. *Science, Technology and Human Values*, *34*(5), 573–606. doi:10.1177/0162243908329381.

Novick, P. (1988). *That noble dream: The 'Objectivity Question' and the American historical profession*. Cambridge, MA: Cambridge University Press.

Porter, A. L., & Rossini, A. (1985). Peer review of interdisciplinary research proposals. *Science, Technology and Human Values*, *10*(3), 33–38. doi:10.1177/016224398501000304.

Travis, G. D. L., & Collins, H. M. (1991). New light on old boys cognitive and institutional particularism in the peer review system. *Science, Technology and Human Values*, *16*(3), 322–341. doi:10.1177/016224399101600303.

Humanities Scholars' Conceptions of Research Quality

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

Abstract The assessment of research performance in the humanities is linked to the question of what humanities scholars perceive as 'good research'. Even though scholars themselves evaluate research on a daily basis, e.g. while reading other scholars' research, not much is known about the quality concepts scholars rely on in their judgment of research. This chapter presents a project funded by the Rectors' Conference of the Swiss Universities, in which humanities scholars' conceptions of research quality were investigated and translated into an approach to research evaluation in the humanities. The approach involves the scholars of a given discipline and seeks to identify agreed-upon concepts of quality. By applying the approach to three humanities disciplines, the project reveals both the opportunities and limitations of research quality assessment in the humanities: A research assessment by means of quality criteria presents opportunities to make visible and evaluate humanities research, while a quantitative assessment by means of indicators is very limited and is not accepted by scholars. However, indicators that are linked to the humanities scholars' notions of quality can be used to support peers in the evaluation process (i.e. informed peer review).

M. Ochsner (✉) · S.E. Hug · H.-D. Daniel
Professorship for Social Psychology and Research on Higher Education,
Department of Humanities, Social and Political Sciences, ETH Zürich,
Mühlegasse 21, 8001 Zürich, Switzerland
e-mail: ochsner@gess.ethz.ch

S.E. Hug
e-mail: sven.hug@gess.ethz.ch

H.-D. Daniel
e-mail: daniel@gess.ethz.ch

M. Ochsner
FORS, University of Lausanne, Géopolis, 1015 Lausanne, Switzerland

S.E. Hug · H.-D. Daniel
Evaluation Office, University of Zürich, Mühlegasse 21, 8001 Zürich, Switzerland

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_5

1 Introduction

In order to evaluate research performance adequately, there should be an explicit understanding of what 'good' research is. Thus, knowledge about research quality is necessary. However, little is known about research quality, especially in the humanities. Existing tools and procedures of evaluation or assessment of (humanities') research do not include an explicit understanding of quality. Even more so, the literature on research evaluation actively avoids the topic, reverting to 'impact', which is easier to measure but not necessarily congruent with research quality.

Yet, the assessment of research performance in the humanities must be linked to the question of what humanities scholars perceive as 'good research'. In a report, the League of European Research Universities (LERU) formulated this in the following way: 'senior administrators and academics must take account of the views of those 'at the coal-face' of research when developing assessment criteria and indicators (as should governments, funders and other external agencies)' (League of European Research Universities 2012, p. 15). If we do not know what 'good research' is, it is impossible to assess it, let alone to improve it. Explicating what characterizes 'good research' is not only important for the assessment of research, but it is also of value to the scholars themselves.

This chapter presents a project[1] in which humanities scholars' conceptions of research quality were investigated, and an approach to research evaluation in the humanities was developed. This chapter is structured as follows: In section one, we outline a framework for developing criteria and indicators for research quality in the humanities. In the subsequent section, we present the results of two studies in which we implemented this framework: In particular, section two describes humanities scholars' notions of quality derived from repertory grid interviews, and section three presents the results from a three-round Delphi survey that resulted in a catalogue of quality criteria and indicators as well as a list of consensual quality criteria and indicators. In section four, we discuss the advantages of basing quality criteria and indicators on scholars' notions of quality before we conclude the chapter with a summary and an outlook.

2 Framework

The bibliometric indicators that are widely used for evaluation in the natural and life sciences should not be applied to evaluate humanities research (Archambault et al. 2006; Bourke and Butler 1996; Butler and Visser 2006; Finkenstaedt 1990; Glänzel

[1]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' (see also http://www.performances-recherche.ch/). The project consisted of three initiatives (i.e. (sub-)projects) and four actions (i.e. workshops and add-ons to the initiatives). This chapter presents such an initiative entitled 'Developing and Testing Research Quality Criteria in the Humanities, with an emphasis on Literature Studies and Art History'. Even though *initiative* would be the correct term, we use the term *project* throughout this chapter for reasons of readability.

and Schoepflin 1999; Gomez-Caridad 1999; Guillory 2005; Hicks 2004; Moed et al. 2002; Nederhof 2006; Nederhof et al. 1989). Since many evaluation procedures are based on quantitative approaches, evaluation faces strong opposition by humanities scholars. Even though there have been different projects initiated to develop assessment tools that might fit to the humanities as well (e.g. Australian Research Council 2012; Engels et al. 2012; European Science Foundation 2011; Giménez-Toledo and Román-Román 2009; Gogolin et al. 2014; Royal Netherlands Academy of Arts and Sciences 2011; Schneider 2009; Sivertsen 2010; White et al. 2009; Wissenschaftsrat 2011b), they are discussed very controversially in the humanities, and some of them have even been rejected or faced boycott by the humanities scholars (e.g. the ERIH project of the European Science Foundation, see Andersen et al. (2009), or the Forschungsrating of the German Wissenschaftsrat, see e.g. Plumpe (2009)). We analysed this critique and identified four main reservations. We then developed a framework that addresses these four points of critique and that can serve as a foundation to develop criteria for research assessment. This framework has been published in Hug et al. (2014), and this section draws on this article.

2.1 The Four Main Reservations About Tools and Procedures for Research Evaluation

While humanities scholars criticize many different aspects of research evaluation and its tools and instruments, four main reservations can be identified that summarize many of these aspects: (1) the methods originating from the natural sciences, (2) strong reservations about quantification, (3) fear of negative steering effects of indicators and (4) a lack of consensus on quality criteria.

2.1.1 Methods Originating from the Natural Sciences

The first reservation relates to the fact that the methods used to assess research quality have their origin in the natural sciences (see e.g. Vec 2009, p. 6). Hence, they do not reflect the research process and the publication habits of humanities scholars, such as the importance of national language or the publication of monographs (see e.g. Lack 2008, p. 14), and this is also supported by bibliometric research (see e.g. Hicks 2004; Nederhof 2006). Furthermore, Lack (2008) warns that the existing procedures reflect a linear understanding of knowledge creation due to the natural sciences' notion of linear progress. However, humanities' and also much of the social sciences' conception of knowledge creation relies on the 'coexistence of competing ideas' and the 'expansion of knowledge' (Lack 2008, p. 14, own translation).

2.1.2 Strong Reservations About Quantification

Second, the quantification of research performance is met with scepticism. Some humanities scholars question the mere idea of quantifying research quality, as becomes evident in a joint letter by 24 philosophers to the Australian government as a reaction to their discontent with the journal ranking in the Excellence in Research for Australia (ERA) exercise: 'The problem is not that judgments of quality in research cannot currently be made, but rather that in disciplines like Philosophy, those standards cannot be given simple, mechanical, or quantitative expression' (Academics Australia 2008, p. 1). Particularly the intrinsic benefits of the arts and humanities are feared to be neglected by the use of quantitative measures. While Fisher et al. (2000) do not deny the possibility of a quantitative measurement of research performance, they stress that these indicators do not measure the important information: 'Some efforts soar and others sink, but it is not the measurable success that matters, rather the effort. Performance measures are anathema to arts because they narrow whereas the arts expand' (Fisher et al. 2000, 'The Value of a Liberal Education', para. 18).

2.1.3 Fear of Negative Steering Effects of Indicators

Third, indicators can have dysfunctional effects. Humanities scholars fear, for example, mainstreaming or conservative effects of indicators: 'Overall, performance indicators reinforce traditional academic values and practices and in trying to promote accountability, they can be regressive' (informant B in (Fisher et al. 2000), 'IV. Critiques of Current Performance Indicators', para. 8). A further negative effect frequently mentioned is the loss of diversity of research topics or even disciplines due to constraints and selection effects introduced by the use of research indicators—thus the reaction of nearly 50 editors of social sciences and humanities journals to the European Science Foundations' European Reference Index for the Humanities (ERIH). They argued as follows: 'If such measures as ERIH are adopted as metrics by funding and other agencies, [. . .] We will sustain fewer journals, much less diversity and impoverish our discipline' (Andersen et al. 2009, p. 8). On a more fine-grained scale, Hose (2009) describes the effect of a focus on citation counts as having 'the tendency to favour spectacular (and given certain circumstances, erroneous) results, and penalize fundamental research and sustainable results as well as those doing research in marginal fields' (Hose 2009, p. 95, own translation), an argument that has gained weight given the current discussion on spurious research findings in many disciplines in the life sciences (see e.g. Unreliable research. Trouble at the lab 2013; Mooneshinghe et al. 2007). Due to the poor reputation of replication and due to strong competition and the need to publish original research in high impact journals, research findings are hardly ever replicated (Unreliable research. Trouble at the lab 2013).

2.1.4 Lack of Consensus on Quality Criteria

The fourth reservation concerns the heterogeneity of paradigms and methods. If there is a lack of consensus on the subjects of research and the meaningful use of methods, a consensus on criteria to differentiate between 'good' and 'bad' research is difficult to achieve (see e.g. Herbert and Kaube 2008, p. 45). If, however, criteria do exist, they are informal, refer to one (sub)discipline and cannot easily be transformed to other subdisciplines [Kriterien werden 'informell formuliert, beziehen sich [...] auf die gleiche Fachrichtung und sind [...] nicht ohne weiteres auf andere Subdisziplinen übertragbar'] (Herbert and Kaube 2008, p. 40).

2.2 The Four Pillars of Our Framework to Develop Sustainable Quality Criteria

In order to take these criticisms into account, we developed a framework to explore and develop quality criteria for humanities research (Hug et al. 2014). It consists of four main pillars that directly address the four main criticisms. The four pillars are (1) adopting an inside-out approach, (2) relying on a sound measurement approach, (3) making the notions of quality explicit and (4) striving for consensus.

2.2.1 Adopting an Inside-Out Approach

If the goal of assessment is enhancing research or improving or assuring research quality, it is clear that we must know what quality actually is. In other words, we need to know what we want to foster. While many different stakeholders are involved in research policy (Brewer 2011; Spaapen et al. 2007, p. 79), it is also clear that only scholars can tell what really characterizes 'good research'. In 2012, the League of European Universities concluded that '[evaluators] must take account of the views of those "at the coal-face" of research when developing assessment criteria and indicators' (League of European Research Universities 2012, p. 15). It is, however, important that the different disciplines' unique quality criteria can emerge. Therefore, quality criteria for the humanities must be based on the humanities scholars' conceptions of research. This is best achieved by adopting an inside-out approach. Ideally, the development process should be rooted in the disciplines or even subdisciplines, since there are inter- and intradisciplinary differences within the humanities (e.g. Royal Netherlands Academy of Arts and Sciences 2011; Scheidegger 2007; Wissenschaftsrat 2011b). Furthermore, a genuine inside-out approach has an open outcome. This means that whatever the scholars define as a quality criterion will be accepted as such, no matter how different it might be from the already known criteria from the natural and life sciences. Finally, the inside-out approach implies a bottom-up procedure. This means that, on one hand, quality criteria should not

be determined solely by political stakeholders, university administrators or a few experts in the field in a top-down manner but rather by the scholarly community in its entirety. On the other hand, this means also that not only professors should have a say in what the important quality criteria are, but also younger researchers' conceptions of quality must be taken into account, since research practices can change and new ways of doing research should be reflected in the quality criteria as well. Applying an inside-out approach and developing specific quality criteria for each discipline is the obvious answer to the reservation that the methods in research evaluation stem from the natural and life sciences and do not take into account the research and communication practices of the humanities.

2.2.2 Relying on a Sound Measurement Approach

While it might seem paradoxical to those who argue against quantification as such, we think that applying a sound measurement approach when developing quality criteria and indicators can account for the reservations about quantification. Such an approach is necessary, because in many evaluation practices, indicators are only very loosely linked to definitions of quality. If we want to measure a concept, however, we must first understand it. This belongs to the basic knowledge in empirical sciences: 'Before we can investigate the presence or absence of some attribute [...], or before we can rank objects or measure them in terms of some variable, we must form the concept of that variable' (Lazarsfeld and Barton 1951, p. 155). However, very often theoretical and empirical studies live separate lives. Goertz concludes from his study of the social sciences that 'in spite of the primordial importance of concepts, they have received relatively little attention over the years' (Goertz 2006, p. 1). This is also true for biblio- and scientometrics. Brooks, for example, concludes in her review of major quality assessments in the U.S. that '[the assessments] often still make only a weak connection between theoretical definitions of quality and its measures by asserting a single rank or rating system that obscures the methodological and theoretical assumptions built into it' (Brooks 2005, p. 1). Donovan also points to the fact that there is a weak or no link between indicators and quality criteria, since the measurement in evaluation is very often data-driven: 'This leads us to the observation that research 'quality' comes to be defined by its mode of evaluation; and it is the measures and processes employed [...] that become the arbiters of research excellence' (Donovan 2007, p. 586). Hence, research quality seems to be defined by its measures instead of the other way round. Looking at one of the most important indicators of research performance, namely citations, Moed finds that 'it is [...] extremely difficult if not impossible to express what citations measure in one single theoretical concept [...]. Citations measure many aspects of scholarly activity at the same time' (Moed 2005, p. 221).

If there is such a weak or even missing link between the concept(s) and indicators of quality while at the same time indicators are ambiguous, it is no surprise that humanities scholars have reservations about the quantification attempts. Hence, it is important to rely on a sound measurement approach, since the issue is not 'first

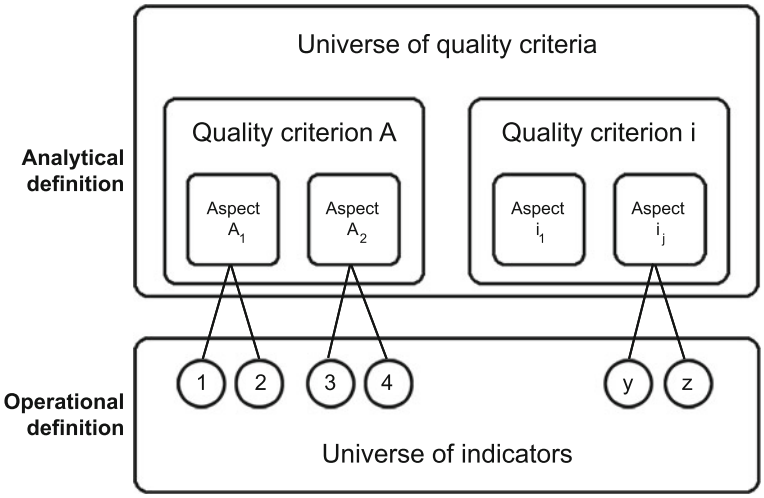


Fig. 1 Measurement model for developing quality criteria and indicators for the humanities. *Source* Hug et al. (2014)

to measure and then to find out what it is that is being measured but rather that the process must run the other way' (Borsboom et al. 2004, p. 1067). When it comes to measurement in research evaluation, it is therefore necessary to have an explicit understanding of quality (Schmidt 2005, p. 3).

We have therefore developed a measurement approach for the operationalization of research quality—the CAI-approach (Criteria, Aspect, Indicator). It is based on a measurement approach commonly used in the social sciences that includes an analytical and an operational definition of a concept (see Fig. 1) and consists of two parts. First, the concept, i.e. quality, has to be defined analytically. Every quality criterion is specified and defined explicitly by one or more aspects. These aspects can then be defined operationally: Each aspect is tied to one more indicators that specify how it can be observed, quantified or measured. Of course, it can be the case that, for a given aspect, no indicators can be found or thought of. Consequently, this aspect cannot be measured quantitatively. Therefore, this approach has the advantage that it is possible to identify quantifiable and non-quantifiable quality criteria. This might reduce scholars' reservations about quantification by disclosing what can be measured and what is exclusively accessible to the judgement of peers and by making clear that quality is not reduced to one simple quantitative indicator.

2.2.3 Making the Notions of Quality Explicit

The quotes by Brooks (2005), Donovan (2007) and Moed (2005) above show that it is not always clear what indicators are measuring. Hence, it is not evident along which criteria research is assessed and into which direction research is steered. The fact

that it is not exactly known what indicators measure and, none the less important, what they do not measure might cause unintended effects of research assessment and trigger fear of negative steering effects in scholars. However, even if it is clear what the indicators of an assessment procedure do measure, scholars still might fear negative steering effects, because the criteria used might not be congruent with their notions of quality. Therefore, it is very important to make the scholars' notions of quality explicit. Yet, to explicate the scholars' notions of quality, it is important not to simply ask them what quality is. They very likely will answer something along the lines of 'I can't define what quality is, but I know it when I see it'. Lamont's study on peer review processes in the social sciences and humanities documents such statements (Lamont 2009). It shows that scholars certainly have knowledge on research quality, as they evaluate research many times during a working day. However, they cannot articulate this knowledge clearly and in detail. Polanyi (1967, p. 22) calls this phenomenon *tacit knowing* and describes it as the 'fact that we can know more than we can tell' (p. 4). *Explicit knowledge*, on the other hand, is 'capable of being clearly stated'. Since knowledge about research quality is still mainly *tacit knowing*, it is important to transform it into *explicit knowledge* in order to develop quality criteria for research assessment in the humanities. To sum up, notions of quality must be as explicit as possible, and the notions of quality of humanities scholars must be taken into account in order to reduce scholars' fears of negative steering effects—and even to reduce the probability of negative steering effects in general.

2.2.4 Striving for Consensus

If we want to develop evaluation criteria that are accepted by the majority of scholars, we must adopt an approach that allows for consensus within a discipline or sub-discipline. By including all scholars in a particular research community or discipline—that is, scholars from all sub-fields as well as methodological backgrounds, young scholars as well as senior professors—it assures the diversity of research and helps foster the acceptance of the criteria while also corresponding to the bottom-up approach described above.

2.3 The Implementation of the Framework: The Design of the Project 'Developing and Testing Quality Criteria for Research in the Humanities'

The design of the project is divided into two main phases: (I) an exploration phase and (II) a phase to find consensus. Because there was not much known about what research quality exactly is in the humanities and because the scholars' knowledge about research quality is mainly tacit, there was a need to first explore what research

quality actually means to humanities scholars. Complying with the first and third pillars, i.e. to adopt an inside-out approach and to make notions of quality explicit, respectively, the exploration phase started the investigation into the notions of quality from scratch. For this aim, we conducted repertory grid interviews with 21 humanities scholars. This technique, developed by Kelly (1955), allows capturing subjective concepts that are used to interpret, structure, and evaluate entities that constitute the respondents' lives (see Fransella et al. 2004; Fromm 2004; Kelly 1955; Walker and Winter 2007). With this method, it is even possible to explicate tacit knowledge (Buessing et al. 2002; Jankowiecz 2001; Ryan and O'Connor 2009). Therefore, it is a very powerful instrument to explore researchers' notions of quality.

While it is possible to develop quality criteria from repertory grid interviews, we found it necessary to validate the criteria derived from the interviewed scholars' notions of quality, because we were able to conduct only a few repertory grid interviews due to the time-consuming nature of the technique. We also strove for consensus regarding the quality criteria according to the fourth pillar of the framework. Hence, we administered a Delphi survey to a large number of humanities scholars. The Delphi method makes use of experts' opinions in multiple rounds with anonymous feedback after each round in order to solve a problem (Häder and Häder 2000; Linstone and Turoff 1975). A Delphi survey starts with an initial round that delineates the problem. This can be done by the research team or, as in our case, by a first qualitative round surveying the experts. This was part of phase I (exploration). The result was a catalogue of quality criteria. In phase II (consensus), two more Delphi rounds, this time in the form of structured questionnaires, served to identify those quality criteria and indicators that reach consensus among the scholars. The Delphi method addresses three pillars from the above framework: By including all scholars of a discipline at the target universities, it (1) contributes to the inside-out approach; (2) it assures a sound measurement approach by structuring the communication process, that is, by linking indicators to the scholars' quality criteria; (3) it facilitates reaching a consensus.

Because both the repertory grid technique as well as the Delphi method are time-consuming methods, we could not investigate the quality notions of a broad range of disciplines. We decided to focus on three disciplines that are characterized by the fact that the commonly used approaches to research evaluation, that is, biblio- and scientometrics, are especially difficult to apply: German literature studies (GLS), English literature studies (ELS) and art history (AH).

3 Notions of Quality: The Repertory Grid Interviews

We conducted 21 repertory grid interviews with researchers from the universities of Basel and Zurich. The sample consisted of 11 women and 10 men, nine of whom were professors, five were senior researchers with a *Habilitation* qualification and seven were researchers holding a PhD.

The repertory grid interviews are built around entities and events meaningful to the respondents in the grid's thematic. These entities and events are called *elements*. We used 17 elements relevant to the scholars' research lives. They were defined by the research team and a repertory grid expert. For example, two of the elements were 'Outstanding piece of research' = Important, outstanding piece of research in the last twenty years in my discipline; 'Lowly regarded peer' = A person in my discipline whose research I do not regard highly. Using 'research' as topic for the elements, the interviewees generated words or syntagms, so-called *constructs*, they associated with pairs of elements they were presented. At the same time, they rated the constructs that they had just generated according to how much they corresponded with each of the 17 elements (for a comprehensive list of the elements as well as an in-depth description of the method and its implementation, see Ochsner et al. 2013).

Repertory grids generate qualitative, i.e. linguistic, and quantitative, i.e. numeric, data at the same time. A look at the linguistic material reveals that there is much communality between the three disciplines. The top categories in all disciplines include 'innovation' and 'approach' (see Table 1). Furthermore, 'diversity' is an important topic in all disciplines. Some differences exist between the disciplines as well. For example, 'cooperation' is mentioned quite a lot in GLS and especially in ELS but only receives a few mentions in AH. Art history is characterized further by the importance of 'scientific rigour' and 'internationality'. GLS, on the other hand, is characterized by the verbalization of 'careerist' mentality, which is not mentioned in ELS and only sparsely in AH. ELS scholars strongly emphasize 'cooperation' and do not mention 'inspiration' and 'careerist' mentality.

If we now combine the linguistic and the numeric data by using factor and cluster analysis to group the linguistic data according to the corresponding numeric data, we can reveal tacit, discipline-specific structures of the elements and constructs. In all three disciplines, the factor analysis yielded a three-dimensional representation of the elements and constructs defined by a quality dimension, a time dimension and a success dimension (in terms of success in the scientific system). In all three disciplines, the quality dimension explained the biggest portion of the variance, which means that quality is the most important factor in structuring the scholars' conception of their research lives. In GLS, the time dimension was the second factor, whereas it was the third factor in the other two disciplines (for details on the method and the statistical results, see Ochsner et al. 2013). Using these dimensions to interpret the linguistic data, we can see which constructs differentiate between, for example, 'good' and 'bad' research. This is obviously important information, since we are looking for notions of quality and quality criteria. We can show, for example, that constructs like interdisciplinarity, public orientation and cooperation have both positive and negative connotations. Interdisciplinary research and cooperation are both positively connoted if they serve diversity and complexity. However, if they are strategically used in order to obtain funding they are negatively connoted. Similarly, public-oriented research is positively connoted if it is innovative, and a connection with public issues is established. It is negatively connoted if the research is driven by public needs and, hence, is not free, or if it is economistic or career driven.

Table 1 Semantic categorization of the constructs from the repertory grid interviews

Category	Total	GLS	ELS	AH
Innovation	14.4	15.0	17.0	11.1
Approach	12.6	18.3	9.4	9.3
Cooperation	10.2	10.0	17.0	3.7
Diversity	6.6	6.7	5.7	7.4
Research autonomy	6.0	5.0	1.9	11.1
Interdisciplinarity	5.4	5.0	7.5	3.7
Skills	4.8	3.3	5.7	5.6
Public impact/applicability	4.8	3.3	5.7	5.6
Rigour	4.8	1.7	1.9	11.1
Resources	4.2	5.0	3.8	3.7
Career-oriented	3.6	8.3	0.0	1.9
Research agenda	3.6	1.7	5.7	3.7
Topicality	3.0	1.7	3.8	3.7
Inspiration	3.0	3.3	0.0	5.6
Internationality	3.0	0.0	1.9	7.4
Openness	3.0	1.7	5.7	1.9
Recognized by peers	2.4	3.3	3.8	0.0
Specialization	2.4	3.3	1.9	1.9
Varia	2.4	3.3	1.9	1.9
Column total	100.0	100.0	100.0	100.0

Note Measures in percent; Total of constructs mentioned: ($n = 167$); German literature studies: ($n = 60$); English literature studies: ($n = 53$); art history: ($n = 54$); Professors: ($n = 66$); Habilitated: ($n = 47$); PhDs: ($n = 54$); Male: ($n = 76$); Female: ($n = 91$); Basel: ($n = 94$); Zurich: ($n = 73$). Some columns might not sum to 100 % due to rounding

Furthermore, the combined analysis also reveals more details about how scholars structure their views regarding research. It showed that, in all disciplines, scholars differentiate between a 'modern' and a 'traditional' conception of research. 'Modern' research is characterized as being international, interdisciplinary, cooperative and public-oriented, whereas 'traditional' research is typically disciplinary, individual and autonomous. Hence, interdisciplinarity, cooperation and public orientation are not indicators of quality but of the 'modern' conception of research. It is notable that there is no clear preference for either conception of research (the 'traditional' conception received slightly more positive ratings). Hence, we can find four types of humanities research (see Fig. 2): (1) positively connoted 'traditional' research, which describes the individual scholar working within one discipline, who as a lateral thinker can trigger new ideas; (2) positively connoted 'modern' research characterized by internationality, interdisciplinarity and societal orientation; (3) negatively connoted 'traditional' research that, due to strong introversion, can be described as monotheistic, too narrow and uncritical; and finally (4) negatively connoted 'mod-

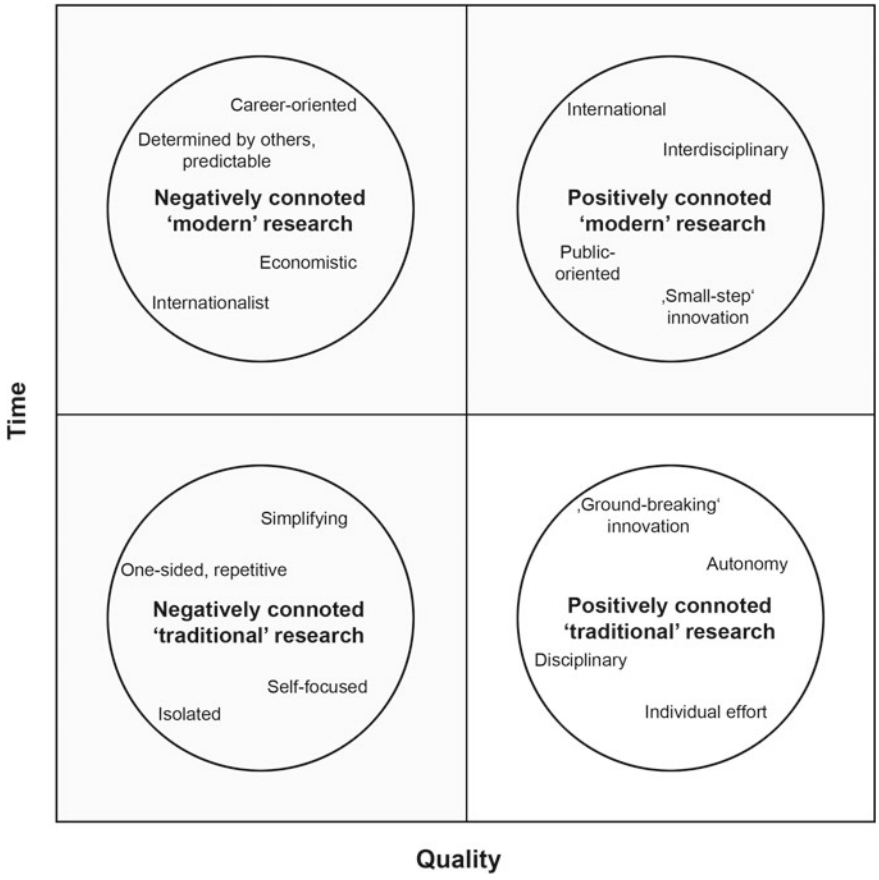


Fig. 2 Four types of research in the humanities. Commonalities across the disciplines. *Source* Ochsner et al. (2013), p. 86

ern' research that is characterized by pragmatism, career aspirations, economization and pre-structuring (see Fig. 2).

Using the time and success dimension, we can show that there are two forms of innovation. The first is connected to the 'modern' concept of research and is characterized as being an innovation of 'small steps'. It is based on new methods or current knowledge. The second is related to the 'traditional' concept of research. It is a 'ground breaking' innovation that is avant-gardist and brings about great changes (such as a paradigm shift). It is in all disciplines close to the element 'misunderstood luminary'. Hence, innovation, as a quality criterion, is double-edged along the success dimension. It can characterize successful research ('small-step' innovation) but also unsuccessful or not-yet-successful research ('ground breaking' innovation).

While the combined analysis of the quantitative and linguistic data is very useful to reveal insights into the implicit notions of quality and is therefore superior to the

traditional qualitative analysis of, for example, interview data (McGeorge and Rugg 1992, pp. 151–152; Winter 1992, pp. 348–351), the interpretation of the linguistic material presented as the first results of the repertory grid reveals valuable information about the salience of some constructs, for example, that innovation, approach and diversity are used often to describe research. Additionally, we can see that internationality is salient only in art history and comes only rarely to the mind of literature scholars when describing research. They talk more often of cooperation. In German literature studies, 'careerist' behaviour is often mentioned.

Getting into the details of the notions of quality, we can see, however, that despite these differences, the notions of quality are still similar. Figures 3, 4 and 5 show a visualization of the elements and clusters of constructs for the three disciplines. In these graphs, the distances between an element and another element, or between a cluster and another cluster, can be interpreted as similarity: The closer two elements are to each other, the more similar they are. However, because the elements and the clusters are scaled differently, the interpretation of the distances between elements and clusters is accessible exclusively via their relative positioning. For example, if a cluster lies closer to an element than a second cluster does, there is greater similarity between the first cluster and the element than between the second cluster and the element (e.g. in Fig. 3, cluster 11, 'productive', is more similar to the element 'research with reception' than cluster 4, 'self-focused'). We simplified the graphical representations for this publication to increase their readability. The clusters were

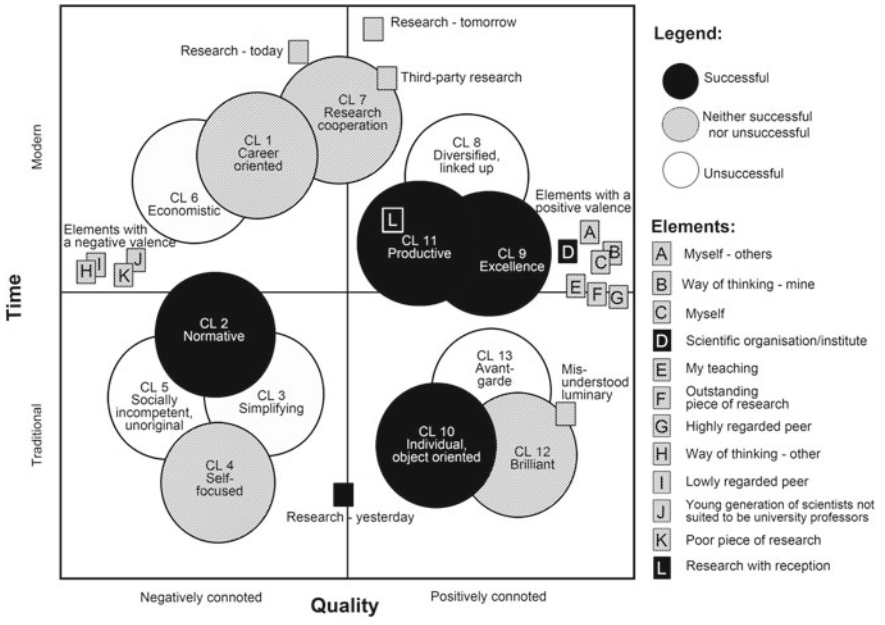


Fig. 3 Schematic representation of the clusters and elements in the discipline *German literature studies*. Slightly modified version of Ochsner et al. (2013), p. 84

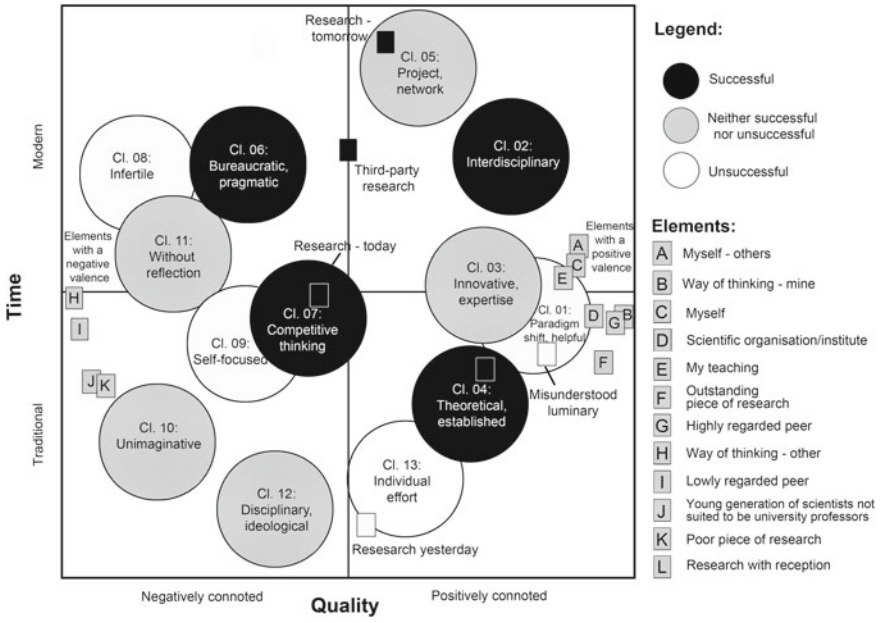


Fig. 4 Schematic representation of the clusters and elements in the discipline *English literature studies*

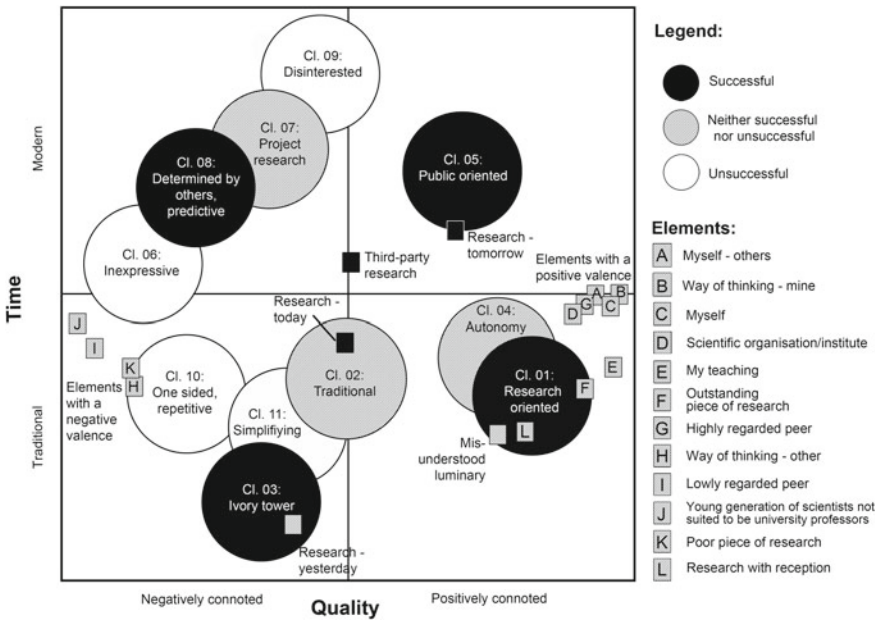


Fig. 5 Schematic representation of the clusters and elements in the discipline *art history*

placed schematically in the two-dimensional space with the axes *quality* and *time*, and the third dimension (*success*) was divided into three groups: successful, neither successful nor unsuccessful and unsuccessful.

The repertory grid for GLS is shown in Fig. 3. For example, cluster 1 represents 'career-oriented' research. Seen from the analysis of the linguistic material only, this is a concept solely salient in GLS. However, we can also find similar clusters in ELS and AH: In ELS, cluster 6, 'bureaucratic, pragmatic', describes applied research that is pragmatic and bureaucratic, associated with numbers-oriented evaluation. It is located in the negatively connoted 'modern' conception of research (see Fig. 4). In AH, cluster 8, 'determined by others', is located at a similar place in the grid and comprises research that is determined by others, elitist, overestimation of self and predictable, controllable and manageable (see Fig. 5). The three clusters encompass the same concept, career-focused strategies of research characterized by writing proposals and adapting to mainstream research. However, only the scholars of GLS clearly name it career-oriented, while in the other disciplines, it is more circumscribed and not clear-cut. However, there are also small differences. In GLS, this cluster's research is characterized by being neither successful nor unsuccessful, whereas in the other two disciplines this kind of research is characterized as successful. Furthermore, there is another cluster in ELS related to a careerist attitude: cluster 7, 'competitive thinking'. It shares the success-oriented approach to research. However, it is more focused on catching the attention of peers than on funding and social impact. This cluster is not restricted to the 'modern' conception of research but rather spreads across the time axis.

There are also clusters that are very similar in all three disciplines: Cluster 7 in GLS, cluster 5 in ELS and cluster 7 in AH are about project or network research. They are part of the 'modern' conception of research and are characterized by differentiation, cooperation, concerted activities and economization pressure. Also in the positively connoted 'traditional' conception of research, there is a cluster that is very similar in all disciplines: Cluster 13 in GLS ('avant-garde'), cluster 1 in ELS ('paradigm shift, helpful') and cluster 4 in AH ('autonomy'). They are all closely related to the element 'misunderstood luminary' and consist of research that is bringing about a paradigm shift by means of theoretical advancement and that is characterized by autonomy and unpredictability. This kind of research is not successful (yet): In GLS and ELS, it belongs to the unsuccessful clusters and in AH, to the neither successful nor unsuccessful clusters.

A peculiarity of AH is that there is only successful research in the positively connoted 'modern' conception of research. In Fig. 5, we can see that there is a positive correlation between the success and the quality dimensions in AH. There is no unsuccessful research both in the positively connoted 'modern' and in the positively connoted 'traditional' conception of research (the correlation of the two dimensions is $r = 0.43$) in AH). In the other two disciplines, the correlation is less striking (GLS: $r = 0.29$); ELS: $r = 0.26$).

4 Consensual Quality Criteria: The Delphi Survey

In order to validate our catalogue of quality criteria, we used the Delphi method. Complying with the bottom-up approach, our panel consisted of all research-active faculty at Swiss universities holding a PhD in GLS, ELS or AH. In order to ensure international standards and comparability, the panel also included all research-active faculty holding a PhD in the three disciplines at the member universities of the League of the European Research Universities (LERU). The first round of the Delphi served to complete the catalogue. The respondents could check or uncheck the existing quality criteria and aspects as well as name new criteria and/or aspects. We also asked for indicators that measure the quality aspects. Because of the heavy workload required to respond to this questionnaire, it was administered to only a part of the sample ($n = 180$) scholars). The first round achieved a response rate of 28 % and resulted in a more refined catalogue of quality criteria, comprising 19 criteria specified by a total of 70 aspects (for a description of the method and the results, see Hug et al. 2013). In the second Delphi round, which was administered to the whole sample $N = 664$), the scholars rated the aspects on a scale from 1 to 6 as to whether they agreed with a given statement. The statement consisted of a generic part that was the same for all aspects (i.e. 'My research is assessed appropriately if the assessment considers whether I . . .') and a second part consisting of the aspect (e.g. '. . . introduce new research topics') of a given criterion (e.g. *Innovation*, *Originality*); 1 was labelled 'I strongly disagree with the statement', 2: 'I disagree', 3: 'I slightly disagree', 4: 'I slightly agree', 5: 'I agree' and 6: 'I strongly agree with the statement'. The second round achieved a response rate of 30 %.

The second Delphi round showed that a broad palette of quality criteria and aspects are needed to appropriately assess research quality in the humanities. Table 2 lists the 19 criteria for research quality in the humanities (for a list of all the 70 aspects, see Hug et al. 2013). In GLS, only 10 out of the 70 aspects scored a mean of less than 4, of which only two received a median lower than 4. The same numbers apply for AH. In ELS, however, 13 aspects scored a mean of less than 4, and five aspects had a median lower than 4. The grand mean of the aspect was 4.71 (range = 3.34–5.74), 4.64 (range = 3.15–5.6) and 4.56 (range = 2.88–5.56) in GLS, AH and ELS, respectively. Of the aspects that have received a negative rating (i.e. mean lower than 4), seven were rejected in all three disciplines—namely, 'reputation in society' and 'insights are recognized by society' (*recognition*), 'continuation of research traditions' and 'long-term pursuit of research topics' (*continuity*, *continuation*), 'establishing a new school of thought' (*impact on research community*), 'responding to societal concerns' (*relation to and impact on society*) and 'research has its impact mainly in teaching' (*connection between research and teaching*, *scholarship of teaching*). Furthermore, in all three disciplines, no criterion was rejected altogether since each criterion had at least one aspect that had been rated with a 4 ('I slightly agree') by at least 50 % of the scholars ($mean > 4$). Hence, the catalogue that resulted from the repertory grid and the first Delphi round aptly reflects the notions of quality of the humanities scholars in the three disciplines.

Table 2 Quality criteria for humanities research: consensuality in the three disciplines

1.	Scholarly exchange^{GLS,ELS,AH}	8.	Continuity, continuation^{GLS}	15.	Scholarship, erudition^{GLS,ELS,AH}
2.	Innovation, originality^{GLS,ELS,AH}	9.	Impact on research community^{GLS,ELS,AH}	16.	Passion, enthusiasm^{GLS,ELS,AH}
3.	Productivity	10.	Relation to and impact on society	17.	Vision of future research^{GLS,ELS,AH}
4.	Rigour^{GLS,ELS,AH}	11.	Variety of research^{GLS,AH}	18.	Connection between research and teaching, scholarship of teaching^{GLS,ELS,AH}
5.	Fostering cultural memory^{GLS,ELS,AH}	12.	Connection to other research^{GLS,ELS,AH}	19.	Relevance^{GLS}
6.	Recognition^{ELS}	13.	Openness to ideas and persons^{GLS,ELS,AH}		
7.	Reflection, criticism^{GLS,AH}	14.	Selfmanagement, independence^{GLS,ELS}		

Note GLS = criterion reached consensus in German literature studies; ELS = criterion reached consensus in English literature studies; AH = criterion reached consensus in art history

However, regarding some aspects and criteria, the scholars were divided (i.e. while some scholars supported the aspect, a large number of others rated the same aspect very low). Therefore, and in order to comply with the fourth pillar of our framework (striving for consensus), we identified those aspects that were clearly approved by a majority and disapproved by very few scholars (i.e. consensual aspects). Consequently, we classified an aspect as consensual when at least 50 % of the discipline's respondents rated the aspect with at least a '5', and not more than 10 % of the discipline's respondents rated the aspect negatively, that is, with a '1', '2' or '3'. Accordingly, we classified a criterion as consensual when at least one of its aspects reached consensus. In GLS, 36 aspects pertaining to 16 criteria reached consensus, in AH, 31 aspects connected to 13 criteria did so and 29 aspects related to 13 criteria reached consensus in ELS. For simplicity reasons, we focus on the criteria in the further analysis. For information regarding the aspects, please refer to Hug et al. (2013).

The data revealed a set of shared criteria consisting of 11 criteria that reached consensus in all three disciplines. Note, however, that not all these criteria are specified with the same consensual aspects in the three disciplines. For example, the criterion *connection to other research* was specified differently in the three disciplines. In GLS, all three aspects of this criterion reached consensus: 'building on current state of research', 're-connecting to neglected research' and 'engaging in on-going research debates'; in ELS, the two aspects 'building on current state of research' and 're-connecting to neglected research' reached consensus; and in AH, only one aspect reached consensus: 'engaging in on-going research debates'. Moreover, six criteria

were consensual in one or two disciplines and can be considered discipline-specific criteria. Finally, two criteria did not reach consensus in any discipline, namely *productivity* and *relation to and impact on society*. Table 2 indicates the consensuality of the criteria in the respective disciplines.

The fact that all criteria reached acceptable mean scores shows that in order to assess research quality in the humanities appropriately, a broad spectrum of quality criteria must be taken into account. Ten of the presented criteria are well known and are already used in evaluation procedures, and nine are less known—namely, fostering cultural memory, reflection/criticism, variety of research, openness to ideas and persons, self-management/independence, scholarship/erudition, passion/enthusiasm, vision of future research, connection between research and teaching/scholarship of teaching. Two of these criteria are also mentioned in the empirical literature on quality criteria in the humanities—reflection/criticism corresponding to reflexivity, deliberation and criticism (Oancea and Furlong 2007) and passion/enthusiasm corresponding to engagement (Bazeley 2010). However, if we look at the criteria that reached consensus, we see that all the nine less known criteria reach consensus in at least two disciplines, whereas some criteria that are very often used, i.e. productivity, recognition, relation to and impact on society and relevance, reach consensus in only one discipline or in none at all. Hence, from the point of view of the humanities scholars' notions of quality, there is doubt as to whether current evaluation criteria can capture research quality in the humanities (VolkswagenStiftung 2014, p. 1).

In order to investigate this issue further, we gathered indicators that are used or are suggested for use in evaluation procedures. These were collected in two steps. The first step consisted of an extensive literature review looking for documents that included criteria or indicators for research in the humanities and related disciplines or documents that addressed criticisms or conceptual aspects of research assessments. This resulted in a bibliography of literature on quality criteria and indicators for humanities research that is accessible on the project's website[2] (Peric et al. 2013). In the second step, the collection of indicators was expanded with indicators that were named by the humanities scholars themselves in our repertory grid interviews and the first Delphi round. Because we identified an abundance of indicators, we had to group them into clusters. The grouping procedure resulted in 62 groups of indicators by following two principles: The indicators of a group must be of similar kind and—in order to comply with our measurement model—it should be possible to assign each group to a specific quality criterion or aspect (for a detailed description of the documents used and the assigning procedure, see Ochsner et al. 2012).

By assigning the indicator groups to the quality criteria and aspects, we are able to quantify the proportion of aspects that can be measured quantitatively. We were able to identify indicators for only about half of the aspects that reached consensus,

[2]See http://www.performances-recherche.ch/projects/developing-and-testing-quality-criteria-for-research-in-the-humanities.

53 % in GLS, 52 % in ELS and 48 % in AH, respectively. In other words, indicators can capture only about half of the humanities scholars' notions of quality.

The scholars rated these groups of indicators in the third Delphi round according to a clear statement on a scale ranging, again, from 1 to 6, where (1) meant 'I strongly disagree with the statement', (2) 'I disagree', (3) 'I slightly disagree', (4) 'I slightly agree', (5) 'I agree' and (6) 'I strongly agree with the statement'. The third Delphi round was designed similarly to the second round. Again, the statements consisted of two parts: a generic part (i.e. 'The following quantitative statements provide peers with good indications of whether I ...') and an aspect (e.g. '... realize my own chosen research goals') of a criterion (e.g. self-management/independence). This statement was followed by the groups of indicators assigned to the given aspect. Because every discipline had its own set of consensual aspects, the questionnaires differed between the disciplines.

In the third Delphi round, which achieved a response rate of 20 %, most items received ratings above 4 (i.e. agreement) by at least 50 % of the respondents. However, in order to be able to use the indicators in assessment procedures, they have to be accepted by most scholars. Hence, we identified the consensual indicators (consensus was defined the same way as in round two: that is, at least 50 % of the discipline's respondents rated the item with at least a '5', and not more than 10 % of the discipline's respondents rated the item with a '1', '2' or '3'). In GLS, 10 indicator groups reached consensus (12 %); in ELS, only one indicator group reached consensus (1 %) and in AH, 16 indicator groups reached consensus (22 %). This is considerably less than in round two, where 51 % of the aspects reached consensus in GLS, 41 % in ELS and 44 % in AH.

The participants also responded to a question asking whether they think that it is conceivable that experts (peers) could evaluate the participants' own research performance appropriately based solely on the quantitative data that the participants had just rated. This question was dismissed by a vast majority of the respondents (GLS: 88 %; ELS: 66 %; AH: 89 %).

5 Discussion: Notions of Quality at the Base of Assessment

Because other projects on research evaluation in the humanities have faced strong opposition (e.g. Andersen et al. 2009; Plumpe 2009, p. 209), we expected a very low willingness of the scholars to participate in our surveys. However, the first two Delphi rounds received quite high response rates of 28–30 %, respectively. Similar studies that surveyed professors report lower or similar response rates (e.g. Braun and Ganser 2011, p. 155; Frey et al. 2007, p. 360; Giménez-Toledo et al. 2013, p. 68). However, in the third Delphi round, where the topic moved from quality criteria to indicators for research performance, only 11 % of the scholars responded to the survey within the same timeframe as in the first two rounds. Even by significantly prolonging the field

period, the response rate did not exceed 20 %. This constitutes initial evidence of the fact that scholars are ready and willing to discuss research quality by defining quality criteria but are not willing to narrow down quality to purely quantitative measures, i.e. indicators. This is further confirmed by the comments we received in response to our surveys. Whereas in the first two rounds the comments were predominantly positive, in the third round a clear majority of the comments was negative (for an analysis of the comments, see Ochsner et al. 2014). Also, the data reveal a clear divide between evaluation by criteria as opposed to evaluation by indicators. In all disciplines, the ratings of the aspects were clearly higher than those of the indicators. This holds true for the grand mean, the share of aspects or indicators that received a positive rating (i.e. $mean \geq 4$) and was even more pronounced for the share of aspects or indicators that reached consensus (for a more detailed integration and comparison of the three Delphi rounds and the repertory grid interviews, see Ochsner et al. 2014).

Hence, we can conclude that humanities scholars prefer a qualitative approach to research evaluation. They are willing to talk about notions of quality and to cooperate in developing quality criteria based on those notions of quality if a bottom-up approach is applied. In order to adequately assess research performance in the humanities, a broad range of quality criteria has to be taken into account. While there is strong reluctance to accept a quantitative approach, it is not rejected altogether. However, the indicators have to be connected to the scholars' notions of quality, i.e. quality criteria.

When on one hand most indicators were accepted by most of the respondents (i.e. most indicators scored a mean of above 4) but failed to reach consensus, the question arises as to why some scholars are reluctant to accept indicators and others approve of them. There are many different reasons for this, but our studies point to two possible reasons that have not yet gained much attention. Firstly, there is a mismatch of quality criteria and indicators between evaluators and humanities scholars, and secondly some quality criteria are double-edged in nature. The mismatch can be described as follows: Some criteria that are frequently used in evaluations are not perceived as indicative of research quality by the humanities scholars (e.g. reputation, societal impact, productivity). On the other hand, there are quality criteria that humanities scholars perceive as important to assess research quality which are not known or are not commonly used in evaluation protocols (e.g. fostering cultural memory, reflection/criticism, scholarship/erudition, passion/enthusiasm). Additionally—and due to constraints of space not reported in this article—the indicators most often used in research evaluations (e.g. citations, prizes, third-party funding, transfers to economy and society) measure criteria that do not reach consensus in all disciplines (i.e. recognition, impact on research community, relevance, relation to and impact on society; see Ochsner et al. 2012, pp. 3–4). The double-edged nature of some quality criteria is revealed in the results of the repertory grid study. Interdisciplinarity, cooperation, public orientation and internationality are often used as quality criteria in evaluation schemes. However, the repertory grid interviews reveal that they are indicators of the

'modern' as opposed to the 'traditional' conception of research and are not necessarily related to quality. If these criteria are used as quality criteria, the 'traditional' conception of research would be forced to 'take a back seat'. However, it has to be kept in mind that the 'traditional' conception of research is highly regarded by the scholars and is connected to an important aspect of innovation: the 'ground-breaking' innovation that establishes new paradigms and theories. Evaluators must not confuse the dichotomy of the 'modern' and 'traditional' conceptions of research with 'new/innovative/promising' versus 'old-fashioned/conservative'. Both are valuable, innovative and important in the humanities.

If humanities research is to be assessed appropriately, it is important that indicators for the 'traditional' conception of research are also used. Using the repertory grid and the Delphi method, we were able to also identify indicators for the 'traditional' conception of research (e.g. the indicator group 'number of sources, materials and original works used in publications or presentations', which measures the aspect 'rich experience with sources' from the criterion 'scholarship/erudition'). However, it is an open question as to whether the 'traditional' conception of research can be measured prospectively at all. The repertory grid interviews point clearly towards the prerequisite of autonomy for such achievements. Quantitative assessments are even explicitly a characteristic of the 'modern' conception of research—more specifically, the negatively connoted 'modern' conception of research (see Ochsner et al. 2013, pp. 91–92). On one hand, the measurement of some characteristics of the 'traditional' conception of research could make visible important contributions of humanities research that might be overlooked otherwise. It also might help promote humanities-specific notions of quality. On the other hand, the measurement of research performance might never capture the true notion of the 'traditional' conception of research, described as an individual researcher who is bringing about a paradigm change by conducting disciplinary research locked up in his study. Hence, many humanities scholars will likely be critical if not disapproving of quantitative measurement and purely indicator-based assessments, having in mind the ideal of the erudite scholar.

6 Conclusion

The assessment of humanities research is a controversly discussed topic. Particularly, the humanities scholars' acceptance of the assessment criteria is an unresolved problem. While most initiatives investigating ways to assess research quality in the humanities focus on enlarging databases, building new rankings or ratings, expanding the quantitative measures to societal impact or studying the peculiarities of humanities' research production (see, e.g. Australian Research Council 2012; Engels et al. 2012; Guetzkow et al. 2004; Hammarfelt 2012; Hemlin 1996; Lamont 2009; Nederhof 2011; Royal Netherlands Academy of Arts and Sciences 2011; Schneider 2009; Sivertsen 2010; White et al. 2009; Wissenschaftsrat 2011a, b; Zuccala 2012), we offer a different approach by starting with the humanities scholars' notions of quality and

linking indicators to the quality criteria that are generated in a bottom-up procedure from within the humanities.

We suggest a framework for developing quality criteria for the humanities that comprises a bottom-up approach, a sound measurement approach, the explication of the humanities scholars' notions of quality and the principle of consensus (Hug et al. 2014). We implemented this framework using the repertory grid technique to explicate the scholars' implicit knowledge of quality, thereby making visible the scholars' notions of quality and generating a first catalogue of quality criteria. We then applied the Delphi method to survey all scholars of the three disciplines covered in this project—German literature studies, English literature studies and art history—at the Swiss and the LERU universities, thereby following a bottom-up procedure. The Delphi method made it possible to find a consensus on quality criteria.

From the results of the four studies we conducted during this project (repertory grid and three rounds of the Delphi survey), we can formulate opportunities for and limitations of research assessments in the humanities.

The limitations of research assessments in the humanities can be formulated as follows: We could identify quantitative indicators for only about 50 % of the notions of quality of the humanities scholars. As long as this holds true, humanities scholars will be very critical of purely indicator-based approaches to research assessment. Furthermore, those indicators that are most commonly used in procedures for research evaluation measure exactly those quality criteria and aspects that are not consensual among scholars (see Ochsner et al. 2012, p. 4). While the humanities scholars emphasize the importance of the 'traditional' conception of research, most indicators used in current research assessment procedures measure the 'modern' conception of research (see Ochsner et al. 2013, pp. 85–86).

However, while the humanities scholars' opposition to purely *indicator-based* research assessments will likely persist given the issues mentioned above, an approach towards research assessment relying on *quality criteria* based on the scholars' notions of quality presents opportunities (such as e.g. the guidelines of the VolkswagenStiftung: VolkswagenStiftung 2014). If a bottom-up approach is chosen and the humanities scholars are involved in formulating the quality criteria, and if a broad range of quality criteria are applied, humanities research can be assessed adequately. Using caution when linking indicators to relevant quality criteria, quantitative data can be used to inform judgements on these quality criteria. Hence, an *informed peer review* process based on the relevant quality criteria creates an opportunity to make humanities research more visible and to assess humanities research adequately. It furthermore facilitates the communication between different stakeholders in the evaluation process, and it helps young researchers to focus on quality criteria.

Of course, the research presented has some limitations. First, it is based on three humanities disciplines only. Future research should include a broader range of disciplines in the humanities and neighbouring disciplines. Second, while the response rates were quite high given the composition of the panel and the topic of the research as well as the workload of filling in the questionnaires, the results are based only on the responses of a third of the contacted scholars. Hence, future research should

involve more scholars. Third, scholars are only one of several stakeholders involved in research assessments. Our approach could be used to investigate the notions of quality of other stakeholders.

Acknowledgments This work 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' (Measuring Research Performance) as part of the cooperative initiative of the Universities of Zurich and Basel entitled 'Developing and Testing Research Quality Criteria in the Humanities, with an Emphasis on Literature Studies and Art History'. Matching funds for the initiative were provided by the University of Zurich.

References

Academics Australia. (2008). Letter to senator the honourable Kim Carr, minister for innovation, industry, science and research (tech. rep. No. 15. September 2009). Retrieved from https://web.archive.org/web/20091221195149/, http://www.academics-australia.org/AA/ERA/era.pdf.

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.

Archambault, E., Vignola-Gagné, E., Côté, G., Lavrivére, V., & Gringras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics, 68*(3), 329–342. 1007. doi:10.1007/s11192-006-0115-z.

Australian Research Council. (2012). The excellence in research for Australia (ERA) initiative. Retrieved from http://www.arc.gov.au/excellence-research-australia.

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

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. doi:10.1037/0033-295X.111.4.1061.

Bourke, P., & Butler, L. (1996). Publication types, citation rates and evaluation. *Scientometrics*, *37*(3), 473–494. doi:10.1007/BF02019259.

Braun, N., & Ganser, C. (2011). Fundamentale Erkenntnisse der Soziologie? Eine schriftliche Befragung von Professorinnen und Professoren der deutschen Soziologie und ihre Resultate. *Soziologie*, *40*(2), 151–174.

Brewer, J. D. (2011). The impact of impact. *Research Evaluation*, *20*(3), 255–256. doi:10.3152/095820211X12941371876869.

Brooks, R. L. (2005). Measuring university quality. *The Review of Higher Education*, *29*(1), 1–21. doi:10.1353/rhe.2005.0061.

Buessing, A., Herbig, B., & Ewert, T. (2002). Implizites Wissen und erfahrungsgeleitetes Arbeitshandeln. Entwicklung einer Methode zur Explikation in der Krankenpflege [Implicit knowledge and experience guided working: Development of a method for explication in nursing]. *Zeitschrift für Arbeits- und Organisationspsychologie, 46*(1), 2–21. doi:10.1026//0932-4089.46.1.2.

Butler, L., & Visser, M. S. (2006). Extending citation analysis to non-source items. *Scientometrics*, *66*(2), 327–343. doi:10.1007/s11192-006-0024-1.

Donovan, C. (2007). The qualitative future of research evaluation. *Science and Public Policy*, *34*(8), 585–597. doi:10.3152/030234207X256538.

Engels, T. C., Ossenblok, T. L., & Spruyt, E. H. (2012). Changing publication patterns in the social sciences and humanities, 2000–2009. *Scientometrics*, *93*(2), 373–390. doi:10.1007/s11192-012-0680-2.

European Science Foundation. (2011). European Reference Index for the Humanities (ERIH). Retrieved from http://www.esf.org/erih.

Finkenstaedt, T. (1990). Measuring research performance in the humanities. *Scientometrics*, *19*(5–6), 409–417. doi:10.1007/BF02020703.

Fisher, D., Rubenson, K., Rockwell, K., Grosjean, G., & Atkinson-Grosjean, J. (2000). *Performance indicators and the humanities and social sciences*. Vancouver, BC: Centre for Policy Studies in Higher Education and Training.

Fransella, F., Bell, R., & Bannister, D. (2004). *A manual for repertory grid technique* (2nd ed.). Chichester: Wiley.

Frey, B. S., Humbert, S., & Schneider, F. (2007). Was denken deutsche Ökonomen? Eine empirische Auswertung einer Internetbefragung unter den Mitgliedern des Vereins für Sozialpolitik im Sommer 2006. *Perspektiven der Wirtschaftspolitik*, *8*(4), 359–377. doi:10.1111/1468-2516.00256.

Fromm, M. (2004). *Introduction to the repertory grid interview*. Münster: Waxmann.

Giménez-Toledo, E., Tejada-Artigas, C., & Mañana-Rodriguez, J. (2013). Evaluation of scientific books' publishers in social sciences and humanities: Results of a survey. *Research Evaluation*, *22*(1), 64–77. doi:10.1093/reseval/rvs036.

Giménez-Toledo, E., & Román-Román, A. (2009). Assessment of humanities and social sciences monographs through their publishers: A review and a study towards a model of evaluation. *Research Evaluation*, *18*(3), 201–213. doi:10.3152/095820209X471986.

Glänzel, W., & Schoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences. *Information Processing & Management*, *35*(1), 31–44. doi:10.1016/S0306-4573(98)00028-4.

Goertz, G. (2006). *Social sciences concepts: A user's guide*. Princeton, NJ: Princeton University Press.

Gogolin, I., Åström, F., & Hansen, A. (Eds.). (2014). *Assessing quality in European educational research. Indicators and approaches*. Wiesbaden: Springer VS.

Gomez-Caridad, I. (1999). Bibliometric indicators for research evaluation: Interfield differences. *Science Evaluation and Its Management*, *28*, 256–265.

Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is originality in the social sciences and the humanities? *American Sociological Review*, *69*(2), 190–212. doi:10.1177/000312240406900203.

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

Häder, M., & Häder, S. (2000). *Die Delphi-Technik in den Sozialwissenschaften. Methodische Forschungen und innovative Anwendungen*. Wiesbaden: Westdeutscher Verlag.

Hammarfelt, B. (2012). Following the footnotes: A bibliometric analysis of citation patterns in literary studies. Doctoral dissertation. Skrifter utgivna vid institutionen för ABM vid Uppsala Universitet (Vol. 5). Uppsala: Uppsala Universitet. Retrieved from http://www.diva-portal.org/smash/get/diva2:511996/FULLTEXT01.pdf.

Hemlin, S. (1996). Social studies of the humanities. A case study of research conditions and performance in ancient history and classical archaeology and english. *Research Evaluation*, *6*(1), 53–61. doi:10.1093/rev/6.1.53.

Herbert, U., & Kaube, J. (2008). Die Mühen der Ebene: Über Standards, Leistung und Hochschulreform. In E. Lack & C. Markschies (Eds.), *What the hell is quality? Qualitätsstandards in den Geisteswissenschaften* (pp. 37–51). Frankfurt a. M.: Campus.

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.

Hose, M. (2009). Glanz und Elend der Zahl. In C. Prinz, & R. Hohls (Eds.), *Qualitätsmessung, Evaluation, Forschungsrating. Risiken und Chancen für die Geschichtswissenschaft?* (pp. 91–98). Historisches Forum. Berlin: Clioonline.

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 for Education Law and Policy, 10*(1), 55–64.

Jankowiecz, D. (2001). Why does subjectivity make us nervous? Making the tacit explicit. *Journal of Intellectual Capital*, *2*(1), 61–73. doi:10.1108/14691930110380509.

Kelly, G. A. (1955). *The psychology of personal constructs*. New York, NY: Norton.

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 a. M.: Campus.

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

Lazarsfeld, P. F., & Barton, A. H. (1951). Qualitative measurement in the social sciences. Classification, typologies, and indices. In D. Lerner, & H. D. Lasswel (Eds.), *The policy sciences*. Stanford: Stanford University Press.

League of European Research Universities. (2012). Research universities and research assessment. Retrieved from http://www.leru.org/files/publications/LERU_PP_2012_May_Research_Assesment.pdf.

Linstone, H. A., & Turoff, M. (1975). Introduction. In H. A. Linstone, & M. Turoff (Eds.), *The delphi method. Techniques and applications* (pp. 3–12). Don Mills: Addison-Wesley.

McGeorge, P., & Rugg, G. (1992). The uses of 'contrived' knowledge elicitation techniques. *Expert System*, *9*(3), 149–154. doi:10.1111/j.1468-0394.1992.tb00395.x.

Moed, H. F., Luwel, M., & Nederhof, A. J. (2002). Towards research performance in the humanities. *Library Trends*, *50*(3), 498–520.

Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht: Springer.

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. *PLoSMedicine, 4*(2), e28. doi:10.1371/journal.pmed.0040028.

Nederhof, A. J., Zwaan, R. A., De Bruin, R. E., & Dekker, P. (1989). Assessing the usefulness of bibliometric indicators for the humanities and the social sciences: A comparative study. *Scientometrics*, *15*(5–6), 423–435. doi:10.1007/BF02017063.

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.

Nederhof, A. J. (2011). A bibliometric study of productivity and impact of modern language and literature research. *Research Evaluation*, *20*(2), 117–129. doi:10.3152/095820211X12941371876508.

Oancea, A., & Furlong, J. (2007). Expressions of excellence and the assessment of applied and practice-based research. *Research Papers in Education*, *22*(2), 119–137. doi:10.1080/02671520701296056.

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). Retrieved from http://www.bibliometrie-pf.de/article/view/157/192.

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*(4), 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 Supplement), 111–132. doi:10.1007/s11618-014-0576-4.

Peric, B., Ochsner, M., Hug, S. E., & Daniel, H.-D. (2013). *Arts and Humanities Research Assessment Bibliography (AHRABi)*. Zürich: ETH Zurich. doi:10.3929/ethz-a-010610785.

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/.

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

Royal Netherlands Academy of Arts and Sciences. (2011). *Quality indicators for research in the humanities*. Amsterdam: Royal Netherlands Academy of Arts and Sciences. Retrieved from https://www.knaw.nl/shared/resources/actueel/publicaties/pdf/quality-indicators-for-research-in-the-humanities.

Ryan, S., & O'Connor, R. V. (2009). Development of a team measure for tacit knowledge in software development teams. *Journal of Systems and Software*, *82*(2), 229–240. doi:10.1016/j.jss.2008.05.037.

Scheidegger, F. (2007). *Darstellung, Vergleich und Bewertung von Forschungsleistungen in den Geistes- und Sozialwissenschaften. Bestandesaufnahme der Literatur und von Beispielen aus dem In- und Ausland*. Bern: Zentrum für Wissenschafts- und Technologiestudien.

Schmidt, U. (2005). Zwischen Messen und Verstehen. Anmerkungen zum Theoriedefizit in der deutschen Hochschulevaluation (evaNet-Positionen 06/2005). Retrieved from http://www.forschungsnetzwerk.at/downloadpub/messen%20und%20verstehen.pdf.

Schneider, J. W. (2009). An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway. *European Political Science*, *8*(3), 364–378. doi:10.1057/eps.2009.19.

Sivertsen, G. (2010). A performance indicator based on complete data for scientific publication output at research institutions. *ISSI Newsletter*, *6*(1), 22–28.

Spaapen, J., Dijstelbloem, H., & Wamelink, F. (2007). *Evaluating research in context: A method for comprehensive assessment*. The Hague: Consultative Committee of Sector Councils for Research and Development.

Unreliable research. Trouble at the lab. (2013, October 19). *The Economist*. Retrieved from http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble.

Vec, M. (2009). Die vergessene Freiheit. Steuerung und Kontrolle der Geisteswissenschaften unter der Prämisse der Prävention. In C. Prinz, & R. Hohls (Eds.), *Qualitätsmessung, Evaluation, Forschungsrating. Risiken und Chancen für die Geschichtswissenschaft?* (pp. 79–90). Historisches Forum. Berlin: Clioonline.

VolkswagenStiftung. (2014). *What is intellectual quality in the humanities? Some guidelines*. Hannover: VolkswagenStiftung. Retrieved from http://www.volkswagenstiftung.de/uploads/media/Humanities_Quality_Guidelines.pdf.

Walker, B. M., & Winter, D. (2007). The elaboration of personal construct psychology. *Annual Review of Psychology*, *58*, 453–477. doi:10.1146/annurev.psych.58.110405.085535.

White, H. D., Boell, S. K., Yu, H., Davis, M., Wilson, C. S., & Cole, F. T. H. (2009). Libcitations: A measure for comparative assessment of book publications in the humanities and social sciences. *Journal of the American Society for Information Science and Technology*, *60*(6), 1083–1096. doi:10.1002/asi.21045.

Winter, D. (1992). *Personal construct psychology in clinical practice: Theory, research and applications*. London: Routledge.

Wissenschaftsrat. (2010). *Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften*. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/download/archiv/10039-10.pdf.

Wissenschaftsrat. (2011a). *Forschungsrating Anglistik/Amerikanistik*. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/nc/arbeitsbereiche-arbeitsprogramm/forschungsrating/anglistikamerikanistik.html.

Wissenschaftsrat. (2011b). *Zum Forschungsrating allgemein*. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/arbeitsbereiche-arbeitsprogramm/forschungsrating.html.

Zuccala, A. (2012). Quality and influence in literary work: evaluating the 'educated imagination'. *Research Evaluation, 21*(3), 229–241. 1093, doi:10.1093/reseval/rvs017.

Part II
The Current State of Quality-Based Publication Rankings and Publication Databases

The ESF Scoping Project 'Towards a Bibliometric Database for the Social Sciences and Humanities'

Gerhard Lauer

Abstract This paper is a brief report on the European Science Foundation (ESF) Scoping Project, installed in 2009, results published in 2010, which examines the potential for developing some form of research output database that could be used for assessing research performance in Social Sciences and Humanities (SSH). Suggestions were made as to how such a database might look.

Bibliometrics is loved neither in the natural sciences, nor in the life sciences, nor in engineering. However, it is a more or less common practice in all of these areas of research. In the humanities and some social sciences, it is neither loved nor practiced—to put it simply. The situation hasn't changed since the European Research Index in the Humanities' (ERIH)[1] was established in 2002. ERIH was established both for humanities 'purposes and in order to present their ongoing research achievements systematically to the rest of the world'. The Index adds: 'It is also a unique project because, in the context of a world dominated by publications in English, it highlights the vast range of world-class research published by humanities researchers in the European languages'. It was, and is, its major goal to improve the unsatisfactory coverage of European Humanities' research through better bibliometric tools.

In 2009, Bonnie Wheeler, President of the Council of Editors of Learned Journals, raised serious objections against ERIH (Zey 2010). She argued: 'ERIH claims that its goal is to aid journals and their contributors, but it will inevitably inform institutional assessments and may result in rigid common protocols for scholarly journals' (Wheeler 2009; cf. Wheeler 2011). Wheeler's concerns are those of many editors regardless of whether their journals are ranked in the ERIH list or not. Maybe not the best, but certainly the most common argument is a different one: In principle, research output in the humanities is not countable and even social sciences are to be treated differently from the science, technology, engineering and medicine (STEM)

[1] http://www.esf.org/erih.

G. Lauer (✉)
Universität Göttingen, Seminar für Deutsche Philologie, Käte-Hamburger-Weg 3, 37073 Göttingen, Germany
e-mail: gerhard.lauer@phil.uni-goettingen.de

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_6

disciplines. Finally, there is an incongruity between the steadily growing numbers of publications and the need for a fair and effective practice of peer review for sufficient library budgets and preservation services. Because the entire system is heavily dependent on tax-payer money, research organizations are calling for an alternative. They advocate for university-based and open-access publishing models (Harley and Krzys Acord 2011). Not only bibliometrics, but the whole system of scholarly publication is challenged and will be under much more pressure in the next few years than it is today (Leydesdorff 2001).

The Agence National de la Recherche (ANR), the Arts and Humanities Research Council (AHRC), the Deutsche Forschungsgemeinschaft (DFG), the Economic and Social Research Council (ESRC) and the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NOW) are working together with the European Research Foundation to meet the challenges presented by the current pressure to establish a more robust bibliometric database for assessing the impact of all types of research output in the domains of social sciences and humanities (SSH). They ask how a bibliometric database for the humanities and social sciences can be developed that more accurately represents humanist work than current citation indices like ERIH or newer 'usage' indices. A European scoping project was established in 2009 to answer the question: 'What is the potential for developing some form of research output database that could be used for assessing research performance in SSH?' In the field of social sciences and humanities the main problems are well known, i.e. the wider scale and variety of research outputs from SSH, the need to consider national journals (in particular those published in languages other than English) and the highly variable quality of existing SSH bibliographical databases due to the lack of a standardized database structure for the input data. On the other hand, it's obvious how rapidly Web of Science (Thomson-Reuters), which is the former Science Citation Index/Social Sciences Citation Index/Arts and Humanities Citation Index, and Scopus (Elsevier) have expanded their coverage of social sciences and humanities journals in the last years. Web of Science has increased the covered number of SSH journals from 1,700 in 2002 to 2,400 in 2009. And Scopus, much stronger in the field, added 1,450 SSH journals in 2009 to its collection of more than 3,500 SSH journals. Moreover, Scopus has already started to add bibliographic meta-data on highly cited books in its database. So-called regional journals are an increasing part of these two main bibliometric database providers. In March 2014, Elsevier indexed 30,000 books, expecting to index around 75,000 by the end of 2015 (Scopus blog, see Dyas 2014). And, as Henk Moed puts it, Google is already the poor man's bibliometrics (Moed et al. 2010, p. 19; cf. Harzing and van der Wal 2009). The driving force, however, is the interest of many researchers and universities to make their results more visible.

Within this situation, the European Scoping Project (cf. SPRU 2009) understands bibliometrics in a broad sense, from bibliographic to statistics, and has taken political, strategic and operational issues into account. Two experts—Diana Hicks and Henk Moed—were asked to give a short report on the actual situation of SSH bibliometrics (Hicks and Wang 2009; Moed et al. 2010). After having discussed the evaluations by Hicks and Moed, the scoping project board members developed a variety of solutions and examined more closely six suggestions: First, to create more comprehensive

national bibliographic systems through the development of institutional repositories. Second, to enhance and build upon existing national documentation systems like METIS in the Netherlands or the DRIVER initiative through the creation and standardization of institutional research management systems. The third suggestion discussed the possibilities for a new database of SSH research outputs from publishers' archives and institutional repositories, and adding to this appropriate data on enlightenment literature and curated events. A further point considered was to take advantage of the competition between Web of Science and Scopus to strengthen the coverage of SSH research outputs, and of the potential of Google Scholar to become a more rigorous bibliometric database provider. The fifth suggestion was whether it would be suitable to integrate the specialized SSH bibliographic lists into one comprehensive bibliographic database. And last, there was a discussion on the chances to encourage the further development of the Open Access approach, since it offers a potential means to overcome barriers of accessibility and to enhance the visibility of SSH journals and books published by small European publishers.

Advantages and disadvantages of each approach were weighed and recommendations were given. These recommendation were based on a combination of top-down and bottom-up actions, with an emphasis on extensive bottom-up involvement in the development of an SSH bibliometric database. Main functions of the recommendations were to provide accountability with regard to the use of public funds, to assess research quality, to provide a comprehensive overview of SSH research outputs in Europe, to map the directions of SSH research and to identify new emerging areas of interdisciplinary SSH research. The four recommendations were:

1. Defining the criteria for inclusion of SSH research outputs and establishing a standardized database structure for national bibliometric databases;
2. exploring the option of involving a commercial supplier in the construction of a single international SSH bibliometric database;
3. conducting a pilot study of one or several specific SSH disciplines; and
4. longer-term expansion and enhancement of the SSH bibliometric database.

The required actions for each recommendation were laid out, to mark very concrete further steps. The roadmap was described as a two year path towards a bibliometric database for the humanities and social sciences. The full report was published with both research reports by Moed and Hicks (Martin et al. 2010; Moed et al. 2010; Hicks and Wang 2009).

The European Science Foundation has already reacted and recently signed a memorandum of understanding with the Norwegian Social Science Data Services (NSD). The decision was made to transfer the ERIH to the NSD website, where it will be possible to submit new journals. However, no decision has been reached whether ERIH should play a larger role, while the oligopoly of major publishing houses and their bibliometrics steadily enlarge their positions. New ways of open review ratings with self-publishing have stepped into the field. The rise of ResearchGate is but one example of an alternative scoring system based on a scholarly social network which, however, still faces the same problems of fair indexing (Murray 2014). How to change the conduct of social sciences and humanities and their reputation-based

system towards a more data-based is still an open question. Neither the established reputation-based system nor a more quantitative combination of many indices is better, more abstract or more valuable. Fairness cannot be born from the head of computers and of scholarly networks alone.

References

Dyas, E. (2014). *Scopus content update: The Arts & Humanities*. Retrieved from http://blog.scopus.com/posts/scopus-content-update-the-arts-humanities.

Harley, D., & Krzys Acord, S. (2011). *Peer review in academic promotion and publishing: Its meaning, locus and future*. Berkeley, CA: Center for Studies in Higher Education. Retrieved from http://escholarship.org/uc/item/1xv148c8.

Harzing, A.-W., & van der Wal, R. (2009). A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business. *Journal of the American Society for Information Science and Technology*, *60*(1), 41–46. doi:10.1002/asi.20953.

Hicks, D., & Wang, J. (2009). *Towards a bibliometric database for the social sciences and humanities-a European scoping project* (Annex 1 of the report 'Towards a bibliometric database for the social sciences and humanities-a European scoping project'). Sussex: Science and Technology Policy Research Unit. Retrieved from http://works.bepress.com/diana_hicks/18/.

Leydesdorff, L. (2001). The challenge of Scientometrics. The development, measurement, and self organization of scientific communications. Boca Raton, FL: Universal publishers.

Martin, B., Tang, P., Morgan, M., Glänzel, W., Hornbostel, S., Lauer, G., & Zic-Fuchs, M. (2010). Towards a bibliometric database for the social sciences and humanities-A European scoping project (A report produced for DFG, ESRC, AHRC, NWO, ANR and ESF). Sussex: Sience and Technology Policy Research Unit. Retrieved from https://globalhighered.files.wordpress.com/2010/07/esf_report_final_100309.pdf.

Moed, H. F., Linmans, J. A. M., Nederhof, A. J., Zuccala, A., López Illescas, C., & de Moya Anegón, F. (2010). *Options for a comprehensive database of research outputs in social sciences and humanities* (Annex 2 of the report 'Towards a bibliometric database for the social sciences and humanities-a European scoping project'). Sussex: Science and Technology Policy Research Unit.

Murray, M. (2014). Analysis of a scholarly social networking site. The case of the dormant user. In K. S. Floyd, P. Rutner, & A. Gardiner (Eds.), *Proceedings of the Seventeenth Annual Conference of the Southern Association for Information Systems (SAIS)* (Paper 24). Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1023&context=sais2014.

SPRU. (2009). *Towards a bibliometric database for the social sciences and humanities: A European scoping project*. Sussex: Sience and Technology Policy Research Unit. Retrieved from http://www.sussex.ac.uk/Units/spru/esf/.

Wheeler, B. (2009). The fragmentation and reification of the scholarly journal. Remarks from the president of the Council of Editors of Learned Journals (CELJ). *The Council of Editors of Learned Journals*. Retrieved from http://www.celj.org/fragmentation.

Wheeler, B. (2011). Journal 'ranking' issues and the state of the journal in the humanities. *Journal of Scholarly Publishing*, *42*(3), 323–381. doi:10.1353/scp.2011.0017.

Zey, C. (2010). Wissenschaftliche Zeitschriften. Aktuelle Situation und Perspektiven. Eine Einführung. In R. Groe (Ed.), *Revues scientifiques. État des lieux et perspectives (Table ronde, organisée par l'Institut historique allemand et le département d'histoire de l'université de Zurich, 15 octobre 2009) - Wissenschaftliche Zeitschriften. Aktuelle Situation und Perspektiven (2. Tag der Geisteswissenschaften, veranstaltet vom DHIP und dem Historischen Seminar der Universität Zürich, 15. Oktober 2009)*. Retrieved from http://www.perspectivia.net/content/publikationen/discussions/3-2010/zey_zeitschriften.

Publication-Based Funding: The Norwegian Model

Gunnar Sivertsen

Abstract The 'Norwegian Model' attempts to comprehensively cover all the peer-reviewed scholarly literatures in all areas of research—including the preferred formats and languages of scholarly publishing in the humanities—in one single weighted indicator which makes the research efforts comparable across departments and faculties within and between research institutions. This article describes the main components of the model and how it has been implemented, as well as the effects and experiences in three of the countries that are making use of the model, and where it has been evaluated: Belgium (Flanders), Denmark and Norway. The article concludes with a discussion of the model from the perspective of the humanities.

1 Introduction

The so-called 'Norwegian Model' (Ahlgren et al. 2012; Schneider 2009), which so far has been adopted at the national level by Belgium (Flanders), Denmark, Finland, Norway and Portugal, as well as at the local level by several Swedish universities, has three components:

(A) A complete representation in a national database of structured, verifiable and validated bibliographical records of the peer-reviewed scholarly literature in all areas of research;
(B) A publication indicator with a system of weights that makes field-specific publishing traditions comparable across fields in the measurement of 'Publication points' at the level of institutions;
(C) A performance-based funding model which reallocates a small proportion of the annual direct institutional funding according the institutions' shares in the total of Publication points.

G. Sivertsen (✉)
Nordic Institute for Studies in Innovation, Research and Education (NIFU),
P. O. Box 2815 Toyen, NO-0608 Oslo, Norway
e-mail: gunnar.sivertsen@nifu.no

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_7

In principle, component C is not necessary to establish components A and B. The experience is, however, that the funding models in C support the need for completeness and validation of the bibliographic data in component A. Since the largest commercial data sources, such as *Scopus* or *Web of Science*, so far lack the completeness needed for the model to function properly, the bibliographic data are delivered by the institutions themselves through Current Research Information Systems (CRIS).

The Norwegian model is designed to represent all areas of research equally and properly. The typical mode of implementation in each country has been for the governments to involve prominent researchers in each major area of research, e. g. deans appointed by the rector's conference to represent the respective faculties at all universities, or experts appointed by the learned societies on the national level. The representative researchers have then been involved directly in the national adaptation and design of the publication indicator (component B). The result of these design processes has been one single and simple pragmatic compromise—the first bibliometric indicator to cover all areas of research comprehensively and comparably—rather than several separate and ideal representations of scholarly publishing standards in each individual field.

The Norwegian model usually attracts more attention in the social sciences and humanities than in the other areas. Initially, the reaction is negative or sceptical because the model turns scholarly values into measurable points. There are also concerns about the fact that, although it covers book publishing and the national level of publishing better than other indicators, it still disregards other valuable publication practices by concentrating on the peer-reviewed literature and giving extra incentives to publishing on the international level.

The model has been evaluated three times. I will refer results from the evaluation in Belgium (Flanders) here in the introduction and return to the evaluations in Denmark and Norway later on.

Flanders introduced a performance-based funding model called the BOF-key for the five Flemish universities in 2003. The bibliometric part of the funding formula was initially based on data from the Web of Science only. As a response to criticisms from the social sciences and the humanities, the Government decided in 2008 to supplement the commercial data source by introducing modifications of component A and B in the Norwegian model. Since 2009, the Flemish Academic Bibliographic Database for the Social Sciences and the Humanities (Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen, VABB-SHW) has collected supplementing bibliographic data from the five universities (Engels et al. 2012). An evaluation of the VABB-SHW was performed in 2012 by the Technopolis Group for the Flemish Government. They found these effects of the initiative (Technopolis Group 2013, pp. 9–10):

- 'The VABB-SHW protects certain types of publications in the SSH from becoming marginal.
- The VABB-SHW boosts publications in peer-reviewed journals and those with publishers who are using peer review procedures. It thus provides some guidance to publication behaviour of researchers in the SSH domain.

- More generally, the VABB-SHW has led to a greater emphasis on using peer review procedures in journals and by publishers.
- The VABB-SHW has contributed to an increased visibility of both the SSH and the recognition of SSH publications within the academic community.
- The VABB-SHW has also contributed to an increased quality of the bibliographic databases in the SSH domain of the university associations. This provides, in turn, new opportunities for strategic intelligence'.

In the following, I will shortly present the three components of the Norwegian model in more detail. I will then present more results from evaluations of the model. I will conclude by discussing the model from the perspective of the humanities.

My contribution here is not a neutral and objective study of the Norwegian model as seen from the outside. I designed the model in 2003–2004 in collaboration with academic representatives from Norwegian universities and as a consultant to the Norwegian Association of Higher Education Institutions and the Norwegian Ministry of Education and Research (Sivertsen 2010). I still have a role in the further development of the model, both in Norway and in Denmark.

2 Component A: Delimitation and Collection of Data

The Norwegian model is designed to serve a partly indicator-based funding system for research institutions. Since institutions have different research profiles (e. g. a general university versus a technical university), the model needs to represent all research areas in a comprehensive and comparable way.

There is no single comprehensive international data source for all scholarly publications in all research areas. Figure 1 exhibits the patterns and degrees of coverage in the two largest commercial data sources, Scopus and Web of Science. We know from the complete data set that we use here for comparison, which is based on data from the Norwegian model in Norway since 2005, that the deficiencies in coverage of the social sciences and humanities are mainly due to incomplete coverage of the international journals, limited or no coverage of national scholarly journals and very limited coverage of peer-reviewed scholarly books (Sivertsen 2014).

The data for the Norwegian model are delimited by a *definition* which all areas of research contributed to develop and agree on before it was published in 2004 (Sivertsen and Larsen 2012, p. 569). According to this definition, a scholarly publication must:

1. present new insight
2. in a scholarly format that allows the research findings to be verified and/or used in new research activity
3. in a language and with a distribution that makes the publication accessible for a relevant audience of researchers
4. in a publication channel (journal, series, book publisher) which represents authors from several institutions and organizes independent peer review of manuscripts before publication.

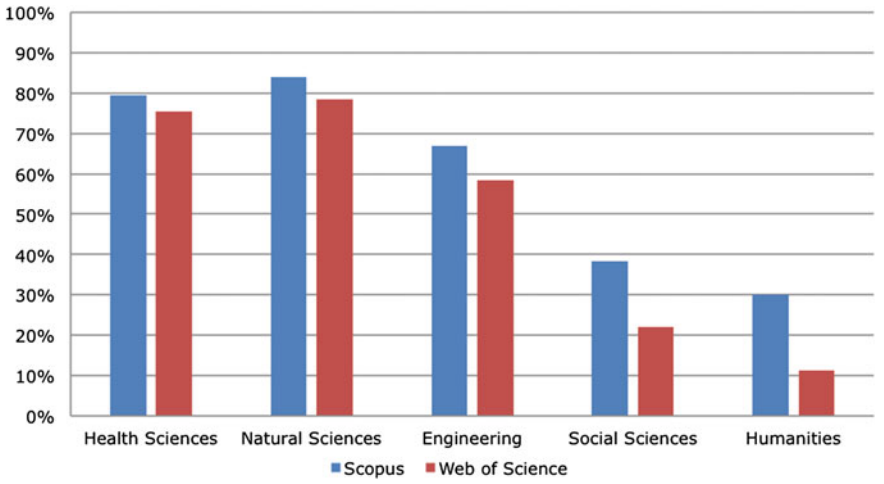


Fig. 1 Coverage in Scopus and Web of Science of 70,500 peer-reviewed scholarly publications in journals, series and books from the higher education sector in Norway 2005–2012

While the first two requirements of the definition demand originality and scholarly format in the publication itself, the third and fourth requirement are supported by a dynamic register of approved scholarly publication channels at http://dbh.nsd.uib.no/kanaler/. Suggestions for additions can be made at any time through the same web page.[1] Publications in local channels (serving only one institution's authors) are not included in the definition, partly because independent peer-review cannot be expected in local channels, and partly because the indicator connected to institutional funding of research is not meant to subsidize in-house publishing.

The definition is not meant to cover the researchers' publishing activities in general. It is meant to represent *research*, not publications. Accordingly, it is limited to original research publications.

In addition to a definition, there is need for a comprehensive data source with bibliographic data that can be connected to persons and their institutional affiliations. These data need to be well-structured (thereby comparable and measurable), verifiable (in external data sources, e. g. in the library) and validated (inter-subjective agreement on what is included according to the definition). These needs are now possible to serve due to the development during the last two decades of Current Research Information Systems (CRIS). They can be designed to produce quality assured metadata at the level of institutions or countries.

CRIS systems on the institutional level have become widespread recently, both in locally and commercially developed solutions. Norway is one of a few countries that has a fully integrated non-commercial CRIS system at the national level. *Cristin* (The

[1] A parallel service at the Norwegian Social Science Data Services was recently established for ERIH PLUS, formerly ERIH (European Reference Index for the Humanities) in collaboration with the European Science Foundation: https://dbh.nsd.uib.no/publiseringskanaler/erihplus/.

Current Research Information System in Norway; cristin.no) is a shared system for all research organizations in the public sector: universities, university colleges, university hospitals and independent research institutes. The Norwegian model, which is now used for institutional funding in all sectors, was a driver in the development of a *shared* system. One reason is that many publications are affiliated with more than one institution and need to be treated as such in the validation process and in the indicator. Another reason is that transparency across institutions stimulates data quality. Every institution can see and check all other institutions' data. The publication database in the CRIS system is also online and open to society at large.

The costs of running Cristin would not be legitimate without multiple use of the same data. References to publications are registered only once, after which they can be used in CV's, applications to research councils, evaluations, annual reports, internal administration, bibliographies for Open Archives, links to full text, etc.

3 Component B: Comparable Measurement

In the measurement for the funding formula by the end of each year, the publications are *weighted* as they are counted. The intention is to balance between field specific publishing patterns, thereby making the publication output comparable across research areas and institutions that may have different research profiles. In one dimension, three main publication types are given different weights: articles in journals and series (ISSN), articles in books (ISBN) and books (ISBN). In another dimension, publication channels are divided into two levels in order to stimulate publishing in the most prestigious and demanding publication channels within each field of research. The highest level is named 'Level 2'. It includes only the leading and most selective international journals, series and book publishers. There is also a quantitative restriction, since the publication channels selected for Level 2 can only in total represent up to 20 % of the world's publications in each field. The weighting of publications by type and channel is shown in Table 1.

Publication points are measured at the level of institutions, not at the level of individual researchers. The points for publications with multiple authors representing several institutions are *fractionalized* among the participating institutions according to their number of participating authors.

Table 1 Publication points in Norway

	Channels at (the normal) level 1	Channels at (the high) level 2
Articles in ISSN-titles	1	3
Articles in ISBN-titles	0.7	1
Books (ISBN-titles)	5	8

The list of journals, series and book publishers on 'Level 2' is revised annually in collaboration with national councils in each discipline or field of research (Sivertsen 2010). These councils propose changes to an interdisciplinary National Publishing Board, which governs the process on behalf of all institutions and has the final decision. Bibliometric statistics (world production versus national production in channels on both levels, and citation statistics for publication channels) are used as an aid in this process, but not as criteria by themselves.

4 Component C: Incentives and Funding

There are two main variants of performance-based funding of research institutions in Europe: the evaluation-based variants (United Kingdom and Italy, also being developed in the Czech Republic and in Sweden), and the indicator-based variants (many smaller European countries). The Norwegian model was developed for indicator-based funding. It is, however, not an alternative to research evaluation. In all of the countries using the Norwegian model presently, research evaluations with expert panels are also practiced, but not with direct consequences for institutional funding.

Countries with indicator-based funding of research institutions do not rely solely on bibliometric indicators. Other indicators may be for example be external funding or the number of doctoral degrees. In addition, the indicators usually reallocate only a minor part of the total funding. Consequently, the economic consequences of an institution's score on the publication indicator in the Norwegian model are therefore relatively small in all countries. In Norway, the publication indicator reallocates less than 2 % of the total expenses in the Higher Education Sector. One publication point represents less than 5,000 Euro.

Still, the publication indicator receives a lot of attention from the researchers, much more attention than is given other and more consequential parts of the funding system. A reason might be that this indicator can be influenced directly by the researchers themselves. Consequently, the Norwegian model seems to be able to change the behaviour of researchers—and that might be a problem.

5 Evaluations of Effects and Experiences

There have been several studies already of the effects of the Norwegian model in different contexts in Denmark, Flanders, Norway and Sweden (Ahlgren et al. 2012; Hammarfelt and de Rijcke 2014; Ossenblok et al. 2012). In addition, there have been three evaluations commissioned by the Governments in Denmark, Flanders and Norway. Above, we referred to the Flemish evaluation in 2012.

The evaluation of the model in Denmark (Sivertsen and Schneider 2012) covered all of the universities and their research areas. As it was performed only three years after the implementation, not much could be said about the effects and possible

unintended consequences. Instead, based on a dialogue with each university, the evaluation identified a number of ideas for improvement of the model which have been taken forward into development work.

The Norwegian model, introduced in 2004, has influenced the funding of Norwegian research institutions since 2005. An evaluation of the effects and experiences was undertaken in 2013. The evaluation was commissioned by the Norwegian Association of Higher Education Institutions and performed by the Danish Centre for Studies in Research and Research Policy at Aarhus University. The report from the evaluation (Dansk Center for Forskningsanalyse 2014), which is in Danish with a ten page summary in English, is being supplemented by a journal article that discusses the results (Aaagaard et al. 2015).

Interviews with researchers and surveys to a large number of them was part of the evaluation in Norway. Since no broad general discontent with the model was found except for the identified problems (see below), and since unintended changes in the researchers' behaviour could not be detected, at least at the macro level, the Ministry of Education and Research has decided to continue using the model as part of the performance-based funding.

The evaluation identified one major effect of the indicator, increased productivity, along with three major problems, all of which I will discuss shortly here.

A main finding was an increased publication rate above what could be expected from the increase of funding. Figure 2 below shows the increase in publication points in the higher education sector since 2004. Figure 3 below has a more independent measurement based on Web of Science. It shows the development in world shares

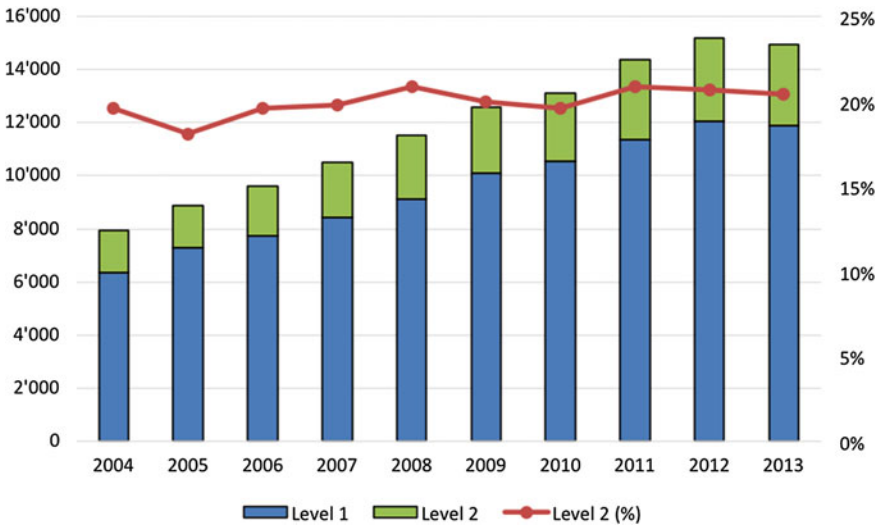


Fig. 2 Publication points in the Norwegian Higher Education Sector 2004–2013. Level 2 represents internationally leading publication channels expected to publish around 20 % of the total. The *red line* and the axis on the *right side* represent the observed percentages on Level 2

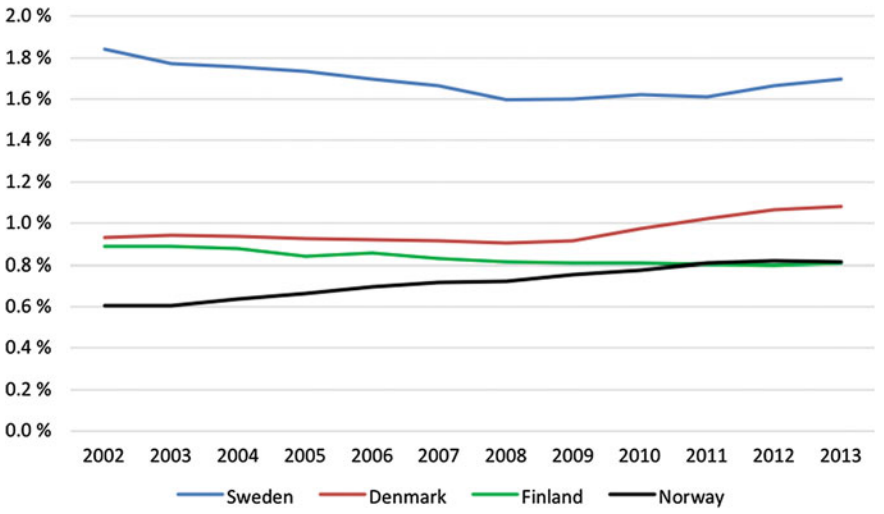


Fig. 3 Shares in the world's scientific output in Web of Science 2000–2013. *Source* National Science Indicators (NSI), Thomson Reuters

of articles for four Scandinavian countries. Note that the incentive to publish was introduced in Norway in 2004 and in Denmark and Sweden in 2009. It will be introduced in Finland in 2015.

The evaluation in Norway found no other changes in the publication patterns than the increase. The balances between publication types (books, articles in books, articles in journals and series) and publication languages (the native language versus international languages) remain the same. Collaboration in authorship is increasing at the same rate as in other countries of the same size. The length of publications remains the same. The citation impact on country level is also stable. And, as seen in Fig. 2, the percentage publications in the most internationally influential publication channels has been stable around 20 %, while the absolute number of those publications has almost doubled.

The evaluation in Norway identified three major problems with the model; one problem in the design of the indicator, and two problems with how the model is practiced.

As mentioned above, the publication points for publications with multiple authors representing several institutions are *fractionalized* among the participating institutions according to their number of participating authors. The evaluation found that this method of fractionalization favours the social sciences and humanities. The average annual publication points per researcher are higher in these areas. Without fractionalization, however, it would be the other way round. Researchers in science, technology and medicine on average contribute to a significantly higher number of publications per year—with the help of their co-authors. The intermediate solution seems to be to use the square root of the institution's fraction of the publication.

The *transparency* and thereby the *legitimacy* of the annual nomination process for Level 2 (described above in component B) is the second problem identified in the evaluation. Here, the Norwegian Association of Higher Education Institutions has started a project to make the whole process of decisions (and their explicit grounds) available in an internet portal open to all researchers, both for influence and for information.

The third problem is the *local use of the indicator*. Although the Norwegian model was developed for institutional funding on the national level, the indicator has become widely used also for internal purposes at the level of institutions, faculties, departments, etc. Some of these practices may be reasonable; other practices can be highly problematic, especially if the indicator replaces responsible leadership and human judgment. Norwegian research institutions are relatively autonomous and cannot be instructed from the outside with regard to leadership practices. However, a large national conference was arranged early in 2015 where leaders of research organizations at all levels shared their views and experiences related to the use of the publication indicator at the local level.

6 Discussion: The Norwegian Model from the Perspective of the Humanities

The humanities are known to have more heterogeneous publication patterns than other areas of research. On the one hand, original peer-reviewed research is published in a wider range of formats. Book publishing (monographs or articles in edited volumes) may even be more important than journal publishing in some of the disciplines (Sivertsen and Larsen 2012). On the other hand, scholars in the humanities, more often than their colleagues in the sciences, publish directly for a wider audience in the societies and cultures that they relate to in their research (Bentley and Kyvik 2011). Even the peer-reviewed scholarly publications may appear in the national language if this is more relevant with regard to contents and outreach (Hicks 2004). In addition, nationally adapted textbooks for students are often preferred over international standard editions. Consequently, scholars in the humanities more often appear as authors of textbooks and other educational material.

Publications for wider audiences and for students can be regarded as the most important expression of societal relevance for the humanities. Furthermore, it can often be difficult to draw a line between publications resulting from new research and publications for students and wider audiences. From this perspective, the Norwegian model seems to be restrictive and disincentivising. However, publishing for wider audiences has in fact increased in Norway after the implementation of the model (Kyvik and Sivertsen 2013). From another perspective, the limitation of the indicator to peer-reviewed publications representing original research can be questioned in relation to its purpose: Does it give a balanced representation of the humanities compared to other research areas? The experience is that it does; the research efforts in the humanities can in fact be matched to the efforts in other areas.

The disciplines *within* the humanities are heterogeneous in their publication patterns. As an example, the degree of international publishing differs a lot across disciplines, and even within them (e. g. in classical versus local archaeology). However, generally, one will find that humanistic scholars will be publishing in a minimum of two languages, one of which is the native language and the other the dominant international language of the field (which in certain humanistic disciplines needs not be English). This is not a new phenomenon; it has been a humanistic practice for two thousand years. Certainly, in our time, we see a gradual and stable increase in English language publishing in the humanities, but there are also large differences between the disciplines (van Leeuwen 2006; Ossenblok et al. 2012), indicating that the bilingual situation will prevail in the humanities due to the societal obligations and wider audiences, as explained above. Furthermore, there is no evidence that book publishing is being replaced by journal publishing in the humanities. The monograph, the edited volume and the journal article, all exist in the humanities because they represent supplementing methodologies in the research itself. Accordingly, all publication types and all languages need to be represented comprehensively in a publication indicator from the perspective of the humanities. From this point of view, the Norwegian model represents a defence of the humanities in a situation where other bibliometric indicators are misrepresenting the disciplines or even creating tensions between them (because there are large variations within the humanities in the representation of the disciplines in commercial data sources).

Access to other publications is perhaps the most important *research infrastructure* in the humanities. It is a paradox, therefore, that this infrastructure is not in place in the humanities as comprehensively as in other research areas. Web of Science, Scopus, PubMed, Chemical Abstracts, etc., were not created for the purpose of research evaluation, but for bibliographic information retrieval. Figure 1 above is, from this perspective, a demonstration of the deficiency of the library system in serving the humanities with an international infrastructure. Figure 1 also illustrates how the Norwegian model can detect this deficiency. A move forward in the direction of making the scholarly output of the humanities searchable and accessible across countries and languages is more needed now, but also more feasible, with the internationalization of research communication. Visibility and availability can be gained for the humanities by the same move forward. However, this goal is less attainable if we regard the humanistic literatures as endless and want everything that we write to be included. As a first step, the Norwegian model provides definitions, thresholds and empirical statistics that can help delimit the scholarly literatures from other literatures and thereby make them internationally searchable and available.

References

Aagaard, K., Bloch, C. W., & Schneider, J. W. (2015). Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. *Research Evaluation, 24*(2), 106–117. doi:10.1093/reseval/rvv003.

Ahlgren, P., Colliander, C., & Persson, O. (2012). Field normalized rates, field normalized journal impact and Norwegian weights for allocation of university research funds. *Scientometrics*, *92*(3), 767–780. doi:10.1007/s11192-012-0632-x.

Bentley, P., & Kyvik, S. (2011). Academic staff and public communication: A survey of popular science publishing across 13 countries. *Public Understanding of Science*, *21*(1), 48–63. doi:10.1177/0963662510384461.

Dansk Center for Forskningsanalyse. (2014). *Evaluering af den norske publiceringsindikator*. Aarhus: Dansk Center for Forskningsanalyse.

Engels, T. C., Ossenblok, T. L., & Spruyt, E. H. (2012). Changing publication patterns in the social sciences and humanities, 2000–2009. *Scientometrics*, *93*(2), 373–390. doi:10.1007/s11192-012-0680-2.

Hammarfelt, B., & de Rijcke, S. (2015). Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. *Research Evaluation*, *24*(1), 63–77. doi:10.1093/reseval/rvu029.

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.

Kyvik, S., & Sivertsen, G. (2013). Økende Forskningsformidling. *Forskningspolitikk, 4*(213), 16–17.

Ossenblok, T. L., Engels, T. C., & 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.

Schneider, J. W. (2009). An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway. *European Political Science*, *8*(3), 364–378. doi:10.1057/eps.2009.19.

Sivertsen, G. (2010). A performance indicator based on complete data for the scientific publication output at research institutions. *ISSI Newsletter*, *6*(1), 22–28.

Sivertsen, G. (2014). Scholarly publication patterns in the social sciences and humanities and their coverage in Scopus and Web of Science. In E. Noyons (Ed.), *Proceedings of the Science and Technology Indicators Conference 2014 Leiden* (pp. 598–604). Leiden: Centre for Science and Technology Studies.

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. doi:10.1007/s11192-011-0615-3.

Sivertsen, G., & Schneider, J. W. (2012). *Evaluering av den bibliometriske forskningsindikator*. Oslo: NIFU.

Technopolis Group. (2013). *Evaluation of the Flemish Academic Bibliographic Database for the social sciences and humanities (VABB-SHW)*. Technopolis Group: Executive summary. Amsterdam.

van Leeuwen, T. (2006). The application of bibliometric analyses in the evaluation of social science research. Who benefits from it, and why it is still feasible. *Scientometrics*, *66*(1), 133–154. doi:10.1007/s11192-006-0010-7.

Assessment of Journal & Book Publishers in the Humanities and Social Sciences in Spain

Elea Giménez Toledo

Abstract This chapter reflects on how journals and book publishers in the fields of humanities and social sciences are studied and evaluated in Spain, particularly with regard to assessments of books and book publishers. The lack of coverage of Spanish output in international databases is underlined as one of the reasons for the development of nationwide assessment tools, both for scholarly journals and books. These tools, such as RESH and DICE (developed by ILIA research team), are based on a methodology which does not rely exclusively on a citation basis, thus providing a much richer set of information. They were used by the main Spanish assessment agencies, whose key criteria are discussed in this chapter. This chapter also presents the recently developed expert survey-based methodology for the assessment of book publishers included in the system Scholarly Publishers Indicators.

1 Introduction

There is little doubt that scholarly communication, reading and citation habits among humanists and social scientists differ from those in other scientific disciplines (as has been studied by Glänzel and Schoepflin 1999; Hicks 2004; Nederhof 2006; Nederhof and Zwaan 1991; Thompson 2002, among many others). Considerable scientific evidence points to the following: in the social sciences and the humanities (SSH), (a) there is a stronger citation pattern in books and book chapters; (b) taking into account the more limited use of scholarly journals, the national-oriented ones are more relevant than the international-oriented ones; (c) this last attribute is related to the local/national character of the research topics covered by the SSH; and (d) the internationality of the research in these branches is conditioned by the research topics.

E. Giménez Toledo (✉)
Spanish National Research Council (CSIC) and Center for Human and Social Sciences (CCHS), C/Albasanz, 26-28, 28037 Madrid, Spain
e-mail: elea.gimenez@cchs.csic.es

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_8

As a brief profile of Spanish scholarly journals, Thomson Reuters Essential Science Indicators ranks Spain ninth for its scientific production and eleventh for the number of citations received. The number of scholarly journals produced in Spain is quite impressive (data from 2012): 1,826 in SSH, 277 in science and technology and 240 in biomedical sciences. Concerning SSH titles, 58 are covered by the Arts and Humanities Citation Index (AHCI), 44 by the Social Science Citation Index (SSCI), 214 by the European Reference Index for the Humanities (ERIH)—both in the 2007 and 2011 lists. These figures indicate an acceptable degree of visibility of Spanish literature in the major international databases, especially if compared with the undercoverage in these databases 15 years ago. Nevertheless, these percentages are not sufficient for dealing adequately with the evaluation process of researchers, departments or schools of SSH. Taking into consideration just the scholarly production included in Web of Science (WoS) or in Scopus, a type of scholarly output which is essential in SSH is underestimated: works published in national languages which have a regional or local scope.

As shown in Fig. 1, the number of Spanish journals not covered by any of these sources is enormous—a group too large to be dismissed. There are at least three reasons for this lack of coverage: (a) Perhaps there are too many journals published in these areas, which can be explained not only by the existence of different schools of thought but also because of the eagerness of universities to have their own reference publications, as another indicator of their status within the scholarly community; (b) in some of these journals, there is a lack of quality and professionalization; and (c) there are high quality journals which will never be covered by those databases due to their lack of internationality—they are specialized in local topics—because they are published in Spanish and because international databases need to define a limited

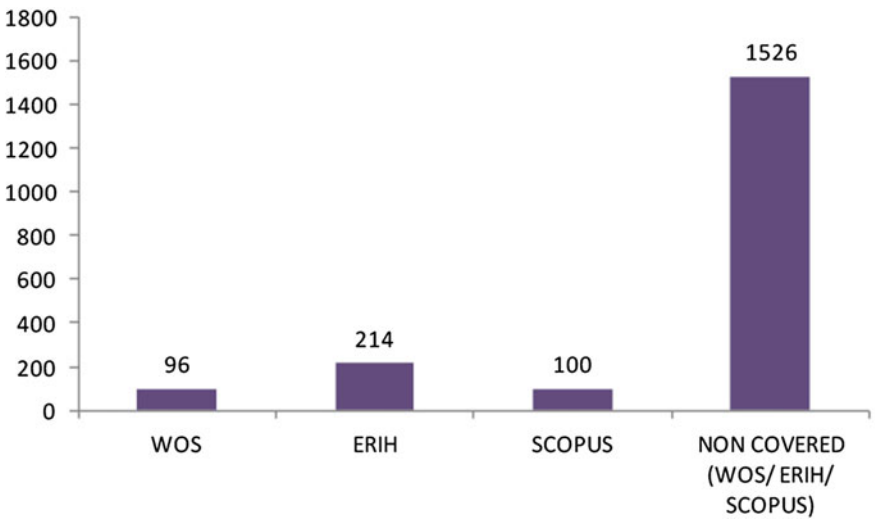


Fig. 1 Coverage of Spanish SSH journals in international databases/indexes

corpus of source journals. It is important to note, on the one hand, that indexing new journals is costly, and, on the other hand, the selective nature of these databases make them suitable for evaluation purposes.

Providing a solution to this problem has been a priority of different research groups in Spain. In the last two decades, several open indicators systems covering Spanish scholarly journals have been created especially for SSH. In all cases, the main motivation for doing so was to build national sources with indicators for journals in a way that complements international sources, to obtain a complete picture of scholarly output in SSH.

The construction of those tools constitutes the applied research developed by the aforementioned research groups, while the theoretical research has had as its object of study the communication and citation habits of humanists and social scientists, as well as the Spanish scientific policy and its research evaluation processes. Such work has drawn the following unequivocal conclusion: not only it is desirable to provide indirect quality indicators for the whole set of journals in a given country; for the successful development of research evaluation in those fields, it is necessary to pay attention to scholarly books, recognize their role as scientific output, increase their weight in assessment processes and develop and apply indicators which might help with assessment processes—but not provide the ultimate verdict (Giménez-Toledo et al. 2015).

2 Research Evaluation in Social Sciences and Humanities in Spain

Research evaluation in Spain is not centralized in a single institution. Several agencies have, among their aims, the assessment of higher education and research institutions, research teams, research projects and scholars. All these agencies are publicly funded and depend on the Spanish Public Administration; nevertheless, their procedures and criteria are not harmonized. This lack of coordination in procedures and criteria can be partially explained by the different objectives which each of these agencies has, but it puzzles scholars and causes confusion regarding the national science policy, which must be the sole one.[1]

The three main evaluation agencies in Spain are CNEAI, ANEP and ANECA. CNEAI (National Commission for the Assessment of Research Activity) is in charge of evaluating lecturers and research staff, through assessing their scientific activity, especially their scientific output. Every 6 years, each researcher may apply for the evaluation of his/her scholarly activity during the last 6 years. A successful result means a salary complement, but what is more important is the social recognition that

[1] At the time of writing this chapter, ANECA (National Agency for Quality Assessment and Accreditation) and CNEAI (National Commission for the Assessment of Research Activity) are in merger process and changes are announced in the evaluation procedures; these are specified in a more qualitative assessment and according to the characteristics of each area.

this evaluation entails: it enables promotions or appointment to PhD committees, or even having a lower workload as a lecturer (BOE 2009).

ANEP (National Evaluation and Foresigh Agency) assesses research projects. Part of its work includes evaluating the research teams leading research projects. Its reports are strongly considered by the Ministry in its decisions to fund (or not fund) research projects.

Finally, ANECA (National Agency for Quality Assessment and Accreditation) has the ultimate goal of contributing to improving the quality of the higher education system through the assessment, certification and accreditation of university degrees, programmes, teaching staff and institutions.

Although the Ministry of Economy and Competitiveness, which currently handles research policy matters,[2] performs ex ante and ex post assessments of its funded projects, and the executive channel for that assessment is ANEP. In addition, FECYT (the Spanish Foundation on Science and Technology) manages assessment issues, since it has the task of evaluating the execution and results of the Spanish National Research Plan. Nevertheless, its conclusions do not directly target researchers nor universities but the national science policy as a whole.

Unlike in other European countries, Spanish assessment agencies are not funding bodies. Each of them establishes its own evaluation procedures, criteria and sources from which to obtain indicators.

Over the past several years, all of these organizations progressively defined specific criteria for the different groups of disciplines, as a form of recognition of their differences. This occurred not only in the case of SSH but also in other fields, such as engineering and architecture. Some researchers regard this specificity as a less demanding subsystem for certain disciplines. Nevertheless, it seems obvious that if communication patterns differ because of the nature of the research, the research evaluation methods should not omit them. Moreover, research assessment by field or discipline is not unique to the Spanish context; a clear example of the extended use of such methodologies is the assessment system applied in the Research Excellence Framework (REF).[3]

The difference in the assessment procedures established by Spanish agencies can be clearly seen in the criteria for publications. With respect to SSH, the following points are worth mentioning:

- Books are taken into account. This might seem obvious, but, in other disciplines, they are not considered at all. In SSH, some quality indicators for books or book publishers are foreseen (see below).
- Regarding journals, and as a common pattern for all fields, WoS is the main source, that is, hierarchically it has much more value than the others. Nevertheless, there are two relevant differences in journal sources for SSH. On one hand, alternative international sources, such as ERIH, Scopus and Latindex, are also mentioned,

[2]From December 2011, and as a consequence of the change of government, the former socialist government created the Ministry for Science and Innovation, a more focused organization for research issues.

[3]http://www.ref.ac.uk/.

even if they appear to have a lower weight. On the other hand, national sources, such as DICE[4] or In Recs,[5] which provide quality and impact indicators for Spanish journals, are considered as well.

The fact that national or international sources are taken into account to obtain the quality indicators of journals (impact, visibility, editorial management, etc.) does not mean that all sources have the same status or weight. However, it does guarantee that a more complete research evaluation can be carried out, by considering most of the scholarly production of an author, research team, etc., and not only what is indexed by WoS. Since some national sources include all journals published in the country, expert panels consider the value of indicators (level of internationalization, peer reviewed journal, etc.), not just their inclusion in the information system.

This is not how it was 15 years ago. However, the appearance of various evaluation agencies, the development of national scientific research plans and the demands of the scientific community have caused the various evaluation agencies (ANECA, CNEAI and ANEP) to gradually refine their research evaluation criteria, and specifically those that refer to publications.

3 Spanish Social Sciences and Humanities Journals' Indicators

Similar to some Latin American countries, such as Colombia, Mexico or Brazil, Spain has extensive experience in the study of its scholarly publications, both in its librarian aspects, such as identification and contents indexation, and in bibliometric or evaluative dimensions.

The Evaluation of Scientific Publications Research Group (EPUC)[6]—recently transformed into *ÍLIA. Research Team on Scholarly Books*—is part of the Centre for Human and Social Sciences (CCHS) at the Spanish National Research Council (CSIC). It was created in 1997 in order to carry out the first systematic studies on the evaluation of scientific journals in SSH.

Shortly thereafter, Spain joined the Latindex system (journal evaluation system, at the basic level, for the countries of Latin America, Spain and Portugal), and this group took charge of representing Spain in this system until 2013.

The team is dedicated to the study of scholarly publications in SSH, particularly in the development and application of quality indicators for scholarly journals and books. One of the objectives of the research is to define the published SSH research so that the systems of research evaluation can consider the particularities of scholarly communication in these fields without renouncing the quality requirements. Another

[4] http://epuc.cchs.csic.es/dice.

[5] http://ec3.ugr.es/in-recs/. IN-RECS is a bibliometric index that offered statistical information from a count of the bibliographical citations, seeking to determine the scientific relevance, influence and impact of Spanish social science journals.

[6] http://ilia.cchs.csic.es.

objective is to improve, by means of evaluation, the average quality of Spanish publications.

During the last decade, the team developed the journal evaluation systems RESH[7] and DICE.[8] The former was built and funded within the framework of competitive research projects (Spanish National Plan for Research, Development & Innovation), while the latter was funded between 2006 and 2013 by ANECA. It is worth mentioning the issue of funding, since it is a crucial issue not only for creating rigorous and reliable information systems but also for guaranteeing the sustainability of those systems. Going even further, public institutions should support the production of indicators which can be used for evaluating research outputs, mostly developed under the auspices of Spanish public funds (METRIS 2012, p. 25). In this way, public funding generates open systems and makes them available, as a public service, to all researchers, guaranteeing transparency and avoiding extra-scholarly interests from non-public database producers. Furthermore, these systems are complementary to the information which can be extracted from the international databases.

Unfortunately, the production of indicators for Spanish publications has not had stable funding. Even the funding of DICE by ANECA, probably the most stable source, ended in 2013 due to budgetary cuts.

As regards RESH and DICE, although they are no longer updated, they are still available online, and they have influenced other Latin American systems. Both systems provided quality indicators for Spanish SSH journals and were useful for researchers, publishers, evaluators of scientific activity and librarians. In addition, they were an essential source of information for the studies carried out by EPUC, as they permitted the recognition, for each discipline, of publication practices, the extent of the validity of each indicator, the particular characteristics of each publication, the level of compliance with editorial standards, the kind of editorial management, etc.

The most complete of these is RESH (see Fig. 2), developed in collaboration with the EC3 group from the University of Granada. It includes more than 30 indirect quality indicators for 1,800 SSH journals.

Users can see all Spanish scholarly journals classified by field. For every single title, its level of compliance with the different indicators established by evaluation agencies (see Table 1 for a list of indicators) is provided (ANECA 2007). Some of them include peer review (refereed/non-refereed journal), databases indexing/abstracting the journal, features of the editorial/advisory board (internationality and represented institutions), percentage of international papers (international authorship) and compliance with the frequency of publication.

This kind of layout makes the system practical. In other words, agencies may check the quality level of a journal according to their established criteria; researchers may search for journals of different disciplines and different levels of compliance with quality indicators; and editors may check how the journals are behaving according to the quality indicators (Fig. 3).

[7] http://epuc.cchs.csic.es/resh.

[8] http://epuc.cchs.csic.es/dice.

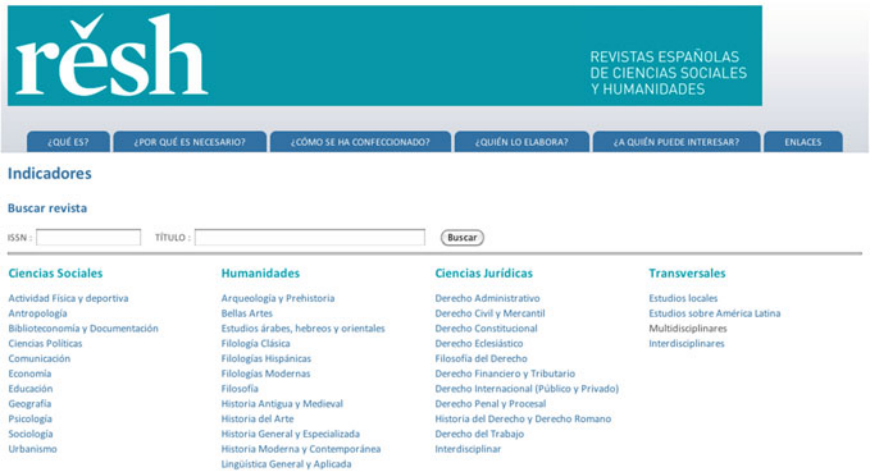


Fig. 2 RESH: a multi-indicator system for evaluating Spanish SSH journals (screenshot)

Table 1 CNEAI indicators of publishing quality

Presence of an Editorial and Advisory Board and Scientific Committee
Detailed guidelines for authors
Summary (Bilingual)
Details about the publishing process
Frequency fulfilled
Blind peer review
Institutional openness of the Advisory Committee
Institutional openness of the Editorial Board
Institutional openness of authors (regarding Editorial Board)
Rate of manuscripts accepted
Indexed in specialized databases
Identification of editorial members
Abstract
Peer review system
Frequency declaration
External reviewers
Justified communication of the editorial decision
Percentage of internationality of the Advisory Committee
Original research
Institutional openness of authors (regarding publishing institution)
Indexed in WoS/JCR and/or ERIH

Revistas

Comunicar. Revista Científica Iberoamericana de Comunicación y Educación

ISSN	1134-3478
AÑO COMIENZO - FIN	1994 -
PERIODICIDAD	SEMESTRAL
EDITOR	Grupo Comunicar, Colectivo Andaluz de Educación en Medios de Comunicación
LUGAR DE EDICIÓN	HUELVA
SOPORTE	IMPRESA
HISTORIA	Hasta marzo de 1994: Comunica. Revista de Medios de Comunicación y Enseñanza. AC-F: 1993-1994
URL	http://www.revistacomunicar.com
ÁREA TEMÁTICA 1	COMUNICACIÓN
ÁREA TEMÁTICA 2	CIENCIAS DE LA EDUCACIÓN
ÁREA DE CONOCIMIENTO 1	COMUNICACIÓN AUDIOVISUAL Y PUBLICIDAD
ÁREA DE CONOCIMIENTO 2	DIDÁCTICA Y ORGANIZACIÓN ESCOLAR
CNEAI	18
ANECA	22
LATINDEX	33
OPINIÓN DE LOS EXPERTOS 2009	79.54
BASES DE DATOS	ISOC ACADEMIC SEARCH COMPLETE CMMC DOAJ ERA FRANCIS MLA PAIS RED ALyC SA SCOPUS SSCI
FECHA DE ACTUALIZACIÓN	2012-06-07

Fig. 3 Databases indexing/abstracting the journal in RESH (screenshot)

RESH also included three more quality indicators not specifically mentioned by evaluation agencies:

- Number and name of databases indexing/abstracting the journal, as a measure of the journals dissemination (see Fig. 3). This information was obtained by carrying out searches and analysing lists of publications indexed in national and international databases.
- An indicator related to experts opinion, since scholars are the only ones who can judge the journals content quality. This indicator was obtained from a survey among Spanish SSH researchers carried out in 2009. The study had a response rate of over 50 % (more than 5,000 answers). By including this element in the integrated assessment of a journal, correlations (or the lack thereof) among different quality indicators may arise. This shall allow for a more accurate analysis of each journal.
- An impact measure for each journal, similar to the Thomson Reuters Impact Factor, but calculated just on the basis of Spanish SSH journals. These data will reveal how Spanish journals cite Spanish journals.

Since no single indicator may summarize the quality of a journal, it seems to be more objective to take into account all these elements in order to provide a clear idea of the global quality of each publication.

4 Book Publishers Assessment

On the one hand, as mentioned previously, books are essential as scholarly outputs of humanists and certain social scientists. Publishing books or using them as preferential sources of research are not erratic choices. On the contrary, books are the most adequate communication channel for the research carried out in the SSH fields.

On the other hand, SSH research should not be evaluated according to others fields patterns but according to their own communication habits. This is not a question of the exceptionality of SSH research but of the nature and features of each discipline. Therefore, an appropriate weight to books in the evaluation of scholarly output is needed to avoid forcing the humanist in the long run to research and publish in a different format, with subsequent prejudices to advance certain kinds of knowledge.

Scholarly publications are the main pillar of the scholarly evaluation conducted by the different assessment agencies.

During the last decade, Spanish evaluation agencies have provided details on journal evaluation criteria. Consequently, the rules are now clearer and more specific for scholars. However, in the case of book assessments, there is still a lot of work to be done. Evaluation agencies have mentioned quality indicators for books. Despite citation products, such as Book Citation Index, Scopus and Google Scholar, there were no sources offering data for making more objective the evalauation of a certain book.

Spanish evaluation agencies have mentioned the following indicators for assessing books in SSH: citations, editors, collections, book reviews in scholarly journals, peer review, translations to other languages, research manuscripts, dissemination in databases, library catalogues and publisher prestige. Nevertheless, generally speaking, the formulation of these criteria is diffuse, subjective or difficult for conducting an objective assessment.

5 Publisher's Prestige

One of the possible approaches to infer the quality of books is to focus on the publisher. In fact, a publishers prestige is one of the most cited indicators by evaluation agencies. Moreover, the methods for analysing quality at the publisher level seem to be more feasible and efficient than at the series or book level, at least if a qualitative approach is pursued. By establishing the quality or prestige of the publisher, the quality of the monographs published could be inferred somehow. The same actually

happens with scholarly papers: they are valued according to the quality or impact of the journal in which they have been published.

With the aim of going into more depth in the study of the quality of books, and mainly to provide some guideline indicators on the subject, the ILIA research team has been working on the concept of publishers prestige. In the framework of our last research projects,[9] we wondered about what publishing prestige is, how it could be defined, which publishers are considered prestigious or how we could make this concept more objective.

The main objectives of this research[10] have been (a) to know the indicators or features that are more valued and accepted by Spanish SSH researchers for evaluating books or book publishers, (b) to identify more relevant publishers according to expert opinion and (c) to analyse how these results could be used in evaluation processes.

In order to achieve these objectives, ILIA designed a survey, aimed at Spanish researchers working in the different disciplines of SSH. Their opinion is the closest expression to the quality of the monographs published by a publisher, as they are the specialized readers and authors who can judge the content of the works, although globally. As the results are opinions, there is always room for bias. Bias nevertheless becomes weaker when the population consulted is wide and the response rate is high.

The survey was sent by e-mail to 11,000 Spanish researchers and lecturers. They had at least a 6-year research period approved by CNEAI. In total, 3,045 completed surveys were returned, representing a 26 % response rate.

One of the questions asked the experts to indicate the three most important publishers in their disciplines. The Indicator of Quality of Publishers according to Experts (ICEE) was applied to the results obtained:

$$ICEE = \sum_{i=1}^{3} n_i * \frac{N_i}{N_j} \tag{1}$$

where n_i is the number of votes received by the publisher in position i (1st, 2nd or 3rd), N_i is the number of votes received by all the publishers in each position (1st, 2nd or 3rd) and N_j is the total number of votes received by all publishers in all positions (1st, 2nd or 3rd).

The weight applied to the votes received by a publisher in each position is the result of dividing the mean of the votes received in that position (in (1st, 2nd or 3rd)) by the sum of the mean of the three positions. In the results, the weight is always bigger for the first position than for the second, and the second bigger than the third.

This indicator has allowed ILIA to produce a general ranking of publishers as well as different rankings by each of the SSH disciplines. The results indicate that there are vast differences between the global ranking and the discipline-based one.

---

[9] *Assessment of scientific publishers and books on humanities and social sciences: qualitative and quantitative indicators HAR2011-30383-C02-01* (2012–2014), funded by Ministry of Economy and Competitiveness. R&D National Plan and *Categorization of scholarly publications on humanities & social sciences* (2009–2010), funded by Spanish National Research Council (CSIC).

[10] Some details on the first project may be found in Giménez-Toledo et al. (2013), p. 68.

Therefore, they also highlight the convenience of using both rankings in the frame of any given research assessment process, as each of them can provide different and relevant information.

5.1 Scholarly Publishers Indicators

These rankings were published for the very first time on the Scholarly Publishers Indicators (SPI) website[11] in 2012. This information system is aimed at collecting the indicators of a different nature for publishers (editorial processes, transparency, etc.), not with the intention of considering them as definitive but as a guide of the quality of the publishers. Indicators and information included are to inform not to perform. In order to avoid the temptation of using them automatically, it is necessary to promote a responsible use of the system.

Since 2013, SPI has been considered by CNEAI as a reference source, albeit not the only one, for the evaluation exercises in some fields of the humanities (history, geography, arts, philosophy and philology). This represents a challenge for further research and developments on this issue. It would be very interesting, for example, to extend the survey to the international scientific community, in order to consolidate and increase the robustness of the results.

6 Conclusions

The aforementioned evaluation tools are a way to improve or at least obtain more information on SSH research evaluation processes. If experts can provide their judgements on the research results, indicators for publications offer objective information on the channel of communication, providing a guide for evaluation processes.

Complementary sources for journals as well as indicators for books or book publishers are needed at the national level if a fair and complete research evaluation is pursued. Although quality indicators for publications may be improved, refined or adapted to special features of certain disciplines, three more complex problems have to be tackled: (a) gaining the acceptance of the scientific community for these kind of indicators, (b) the *formula* for funding these systems and (c) the relationship between large companies devoted to scientific information and selection of information sources for evaluation purposes in evaluation agencies at the national and international level. All of them should be studied in detail in order to handle the underlying problems regarding evaluation tools. Without such a research, any of the evaluation systems will remain limited, biased or unaccepted.

[11] http://ilia.cchs.csic.es/SPI/.

References

ANECA. (2007). *Principios y orientaciones para la aplicación de los criterios de evaluación*. Madrid: ANECA. Retrieved from http://www.aneca.es/content/download/11202/122982/file/pep_criterios_070515.pdf.

BOE. (2009). *Resolución de 18 nov. 2009, de la Presidencia de la Comisión Nacional Evluadora de la Actividad Investigadora, por la que se establecenlos criterios especificos en cada uno de los campos de evaluación*. Retrieved from https://www.boe.es/boe/dias/2009/12/01/pdfs/BOE-A-2009-19218.pdf.

Giménez-Toledo, E., Mañana-Rodriguez, J., Engels, T. C., Ingwersen, P., Pölönen, J., Sivertsen, G., & Zuccala, A. A. (2015). The evaluation of scholarly books as research output. Current developments in Europe. In A. A. A. Salah, Y. Tonta, A. A. Akdag Salah, C. Sugimoto, & U. Al (Eds.), *Proceedings of the 15th International Society for Scientometrics and Informetrics Conference, Istanbul, Turkey, 29th June to 4th July, 2015* (pp. 469–476). Istanbul: Bogazici University. Retrieved from http://curis.ku.dk/ws/files/141056396/Giminez_Toledo_etal.pdf.

Giménez-Toledo, E., Tejada-Artigas, C., & Mañana-Rodriguez, J. (2013). Evaluation of scientific books' publishers in social sciences and humanities: Results of a survey. *Research Evaluation*, *22*(1), 64–77. doi:10.1093/reseval/rvs036.

Glänzel, W., & Schoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences. *Information Processing & Management*, *35*(1), 31–44. doi:10.1016/S0306-4573(98)00028-4.

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.

METRIS. (2012). *Monitoring European trend in social sciences and humanities. Social sciences and humanities in Spain* (Country Report). Retrieved from http://www.metrisnet.eu/metris//fileUpload/countryReports/Spain_2012.pdf.

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.

Nederhof, A. J., & Zwaan, R. (1991). Quality judgments of journals as indicators of research performance in the humanities and the social and behavioral sciences. *Journal of the American Society for Information Science*, *42*(5), 332–340. doi:10.1002/(SICI)1097-4571(199106)42:5<332:AID-ASI3>3.0.CO;2-8.

Thompson, J. W. (2002). The death of the scholarly monograph in the humanities? Citation patterns in literary scholarship. *Libri*, *52*(3), 121–136. doi:10.1515/LIBR.2002.121.

European Educational Research Quality Indicators (EERQI): An Experiment

Ingrid Gogolin

Abstract 'European Educational Research Quality Indicators (EERQI)' was a research project funded under the EU 7th Framework Programme from 2008 to 2011. The mission of this project was to develop new approaches for the evaluation of quality of educational research publications. Traditional methods of assessing quality of scholarly publications are highly depended on ranking methods according to citation frequency and journal impact factors. Both are based on methodologies that do not reflect adequate coverage of European scientific publications, namely in the social sciences and humanities. Hence, if European science or institutions are exposed to these evaluation methods, not only individual researchers and institutions are widely ignored, but also complete subject domains and language areas. The initiators of the EERQI project, as well as numerous researchers and evaluation bodies within the European Union, recognized the need to remedy the inadequacies of this situation.

According to our hypotheses, educational research served as a model case for research in the social sciences and humanities. EERQI aimed to

- develop a prototype framework for the intelligent combination of new indicators and methodologies for the assessment of quality in educational research texts,
- make this framework operational on a multilingual basis (starting with English, German, French and Swedish),
- test the transferability of the EERQI framework to another field of social sciences and the humanities.

The contribution[1] focuses on the design of the project and its general aims and basic ideas. In brief, the EERQI-prototype framework is sketched: what is it about? How is it composed? What is its scientific and practical value?

[1] This article is based on a contribution to the Conference 'Research Quality in the Humanities', Zurich, October 2010. My thanks go to Virginia Moukouli for her support of the presentation.

I. Gogolin (✉)
University of Hamburg, School of Education, Institute for International Comparative and Intercultural Education, Von-Melle-Park 8, 20146 Hamburg, Germany
e-mail: Gogolin@uni-hamburg.de

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_9

1 General Outline: The EERQI Project

In order to understand the scope and aims of the EERQI project, a brief excursion to the context of the endeavor should be helpful: why was it felt necessary to start the EERQI-project?

1.1 Motivation

All across the world the structures and control mechanisms of publicly funded research projects have changed dramatically in the last decade. There are many widely discussed causes of these developments. The set of causes on which we concentrate here is based on the evocation of the 'ability to compete internationally'—a request that is expressed vis-à-vis national research landscapes in Europe as well as the European research space.

A metaphor that is either explicitly used or implicitly resonates in the existing discourses, in the decisions on new governance mechanisms and in new modes of research funding is *quality*. The discovery, improvement and promotion of research quality are the driving motives for the tendency to re-evaluate and redevelop structures for the research area, for redesigning the funding of research institutions and projects, and for instituting control and legitimization systems that are (or intend, or pretend to be) helpful for decision-makers.

In the framework of these developments the questions of *how quality is interpreted* and *how it is measured* are of fundamental importance. Analyses dealing with this question supplied the starting point for the development of the research project 'European Educational Research Quality Indicators (EERQI)'. The project was developed by a truly interdisciplinary European research consortium, a unique composition of experts from Educational Science, Biblio- and Webometrics, Information and Communication Technology, Computational Linguistics and Publishing Houses. It received funding under the Social Sciences and Humanities Funding Scheme of the European Union's 7th Research Framework until March 2011.[2]

The focus of the analysis prior to the project was on particular questions such as: What constitutes and marks the current quality control systems that are applied in contexts of governance and funding, irrespective of the genre and type of research that is at stake? And what are possible effects of these systems on research that is conducted in the European Research Area, especially in the domains of Social Sciences and the Humanities?

According to our assumptions, educational research is especially privileged for considerations and research on such questions because it can be considered as prototypical for vast areas of the whole field of social Sciences and Humanities.

[2]For details have a look at the EERQI website: http://www.eerqi.eu; see also Gogolin (2012) and Gogolin et al. (2014).

This legitimates as follows: Education science and research combine a wide spectrum of theoretical and methodological approaches—from a primarily philosophical-historical methodologies as used in the humanities to psychologically or sociologically based empirical observations of individual development, education, training or *Bildung*; from hermeneutical interpretation, single case studies to the generation and statistical analysis of great amounts of survey data. This manifests relevant characteristics of knowledge production which are also found in other disciplines in the Social Sciences and Humanities.

The EERQI review on the appropriateness of instruments and strategies for quality assessment that are actually applied to educational science resulted in a generic judgment that can briefly be articulated as follows: The existing instruments do not lead to valid results because they do not measure what they claim to measure. An example for the illustration of this statement is quality assessment based on citation indices and journal rankings. This is, at least as yet, the most common approach in vast areas of quality assessment.

The central criterion that is used in these instruments is 'international visibility' of research findings. This is expressed by the placement of the publication, namely in journals with a good reputation, and by the number of citations of a publication. This approach is characteristic of the Social Science Citation Index, a commercial instrument owned by the US-American publishing group Thomson Reuters. Its results often play an important role in reporting systems on research achievement. A closer look at the documentation of the journals represented by this index reveals (for 2009 and the field of educational science according to the 'Journal Citation Report'[3]) the following:

A total of 201 educational research journals were incorporated in the rankings in 2009. Approximately 52 % of these journals were published by US-American publishers. An additional 24 % derived from British publishing houses. The next 'largest' nations in this ranking were the Netherlands (with 4 % of cited journals) and Germany (with 3 % of cited publications). All together 15 nations across the world were represented in the ranking of the Journal Citation Report. A slightly different perspective reveals that 89 % of the publications were in English. The next 'largest' languages with 2.5 % and 2 % respectively were in German, Spanish and Turkish. Eleven languages in total were represented by the index. A language such as French was not included.

We have to admit that the Thomson Reuters-Group itself recently started with a revision of their policies of including journals into the rankings. The Group has incorporated additional journals from other areas of the world into their system—this may be a reaction on international criticism of the instruments, and EERQI may have played a modest role in this. But nevertheless the findings illustrate that these

[3] Journal Citation Reports are a commercial product offered by the US-American publishers' group Thomson Reuters, see http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/ [November 2014]. The products can be linked with ISI Web of Knowledge and Web of Science.

kinds of approaches do not produce valid information in the sense they pretend to do, because the intended international relevancy of the included publications cannot be proven. The rankings are still heavily biased: they essentially refer to US-American or UK-publications and publications in English. International visibility as a quality criterion must be translated here to: the visibility of products from a selection of national research spaces to the rest of the world. The provided information is perfectly suitable to substantiate the powerful dominance of a 'minority' of regional and linguistic research areas.

It is unfortunate that other regional and linguistic research areas, which do not have the benefit of this reinforcement of dominance, participate actively in cementing and safeguarding the existing pattern. This is not least the case in Europe. Prominent research funding institutions affirmatively employ methods that lead to the illustrated result and thus fortify their importance. An example: Calls in the framework of European Research Council's ERC Grant Schemes include the following advice that implies, as we may assume, criteria for the evaluation of proposals. Applicants are asked for 'A list of the top 10 publications, as senior author (or in those fields where alphabetic order of authorship is the norm, joint author), listing all authors, in major international peer-reviewed multidisciplinary scientific journals and/or in the leading international peer-reviewed journals and/or peer-reviewed conferences proceedings of their respective research fields, also indicating the number of citations (excluding auto-citations) they have attracted and possibly the h-index (if applicable)' (European Research Council 2011).[4]

Negotiations about possible alternatives for the assessment of quality in research areas that are not appropriately mirrored in these kinds of methodologies have as yet not been overwhelmingly successful. An example for this is the British Research Excellence Framework—the system for assessing the quality of research in the UK Higher Education system.[5] The 2011 Higher Education Funding Council in England (HEFCE) Report on a pilot exercise to develop bibliometric indicators for the Research Excellence Framework, a review that was used for the preparation of the British Research Excellence Framework, stated: 'The pilot exercise showed that citation information is not sufficiently robust to be used formulaically or as a primary indicator of quality; but there is considerable scope for it to inform and enhance the process of expert review'.[6] Hence, whilst fully aware of the constraints of these methodologies, the respective instruments and data deriving from them, they are extensively in demand and applied by the bodies that conduct processes of research assessment and governance (for the development of this see Oancea 2014).

[4]In recent calls, requirements are described in less detail, but still insist on publications in 'the leading international peer-reviewed journals' (see for example ERC-CoG-2015 on http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/9063-erc-cog-2015.html, accessed 15th December 2014).

[5]See http://www.ref.ac.uk, accessed 9th December 2015.

[6]See http://www.ref.ac.uk/about/background/bibliometrics/, accessed 9th December 2015.

1.2 The EERQI-Project

The motivation for the development of the EERQI-project, in a nutshell, was the observation that the strategies of assessment that were developed in 'hard science'-contexts are heavily criticized for their methodological weakness and lack of validity—not only from a social sciences and humanities point of view (Bridges 2009; Bridges et al. 2009; Gogolin and Lenzen 2014; Mocikat 2010).[7] At the same time there is a serious desire to dispose of approaches that can serve better for the aim of detecting research quality. This desire unites the research community as well as relevant stakeholders from other spheres, such as publishing houses, research funding, political decision making.

The initiators of the EERQI-project never had the idea to take up a battle and try to compete with the economically powerful suppliers of approaches like the Thomson-Reuter's, Scopus (Elsevier) or similar players. Our general intention was to develop useful tools that support the *process* of quality detection. An intelligent combination of such tools—that was our assumption—could be able to assist the readers in the task of determining the class and value of a single text or a series of research texts, be it for assessment purposes or for information in a research process. The application of these process-oriented tools should meet two aims:

1. It should raise the transparency and quality of the process of quality detection itself;
2. It should make the task better manageable and less time-consuming.

In order to meet these aims, EERQI's objective was *not* to develop one single method, such as an index. Instead we aimed to develop and test a *set* of tools that can be applied in different stages of an assessment process, as single methods or in intelligent combinations. These tools should be based on explicit criteria that make the assessment process and result more transparent. In other words: EERQI did not aim at replacing the human decision making in evaluation and assessment procedures, but at maintenance for the individual actors in the procedures or for groups of actors, such as assessment boards. The set of tools we developed is what we call the EERQI Prototype Framework.

2 What EERQI Achieved

The EERQI Prototype Framework is based on the products that were developed in the course of the project. It consists of the following

- a content base with educational research texts in the four European languages that were included in the EERQI project as examples for European multilingualism: English, German, French and Swedish.

[7] See also http://www.adawis.de/index.php?navigation=1, accessed 22nd May 2011.

- a multilingual search engine that includes query expansion: an effective tool dedicated to educational research in general, capable of finding educational research texts in the web in the four EERQI languages.
- automatic semantic analysis for the detection of key sentences in a text. This method is applicable to educational research publications (for a start) the four EERQI languages.
- a combination of bibliometric/webometric approaches for the detection of 'extrinsic' quality indicators (a tool named aMeasure).
- first tests of a citation analysis method that has the potential to be further developed for the application to educational research (and other SSH) texts.
- a set of text-immanent (intrinsic) indicators for the detection of quality in educational research publications that has been presented to the research community and was positively evaluated.
- an accompanying peer review questionnaire that was tested for reliability and feasibility of the instrument.
- a set of use case-scenarios that advice on how to use which resp. combination of the above mentioned tools.
- first attempts to detect interrelations between 'extrinsic' and 'intrinsic' quality indicators.
- and last not least: a successful test of transferability of the approaches developed in EERQI to political science, another areas of social sciences and humanities.

All products are accessible via the EERQI web site (http://www.eerqi.eu). Figure 1 illustrates the Prototype Framework and its elements.

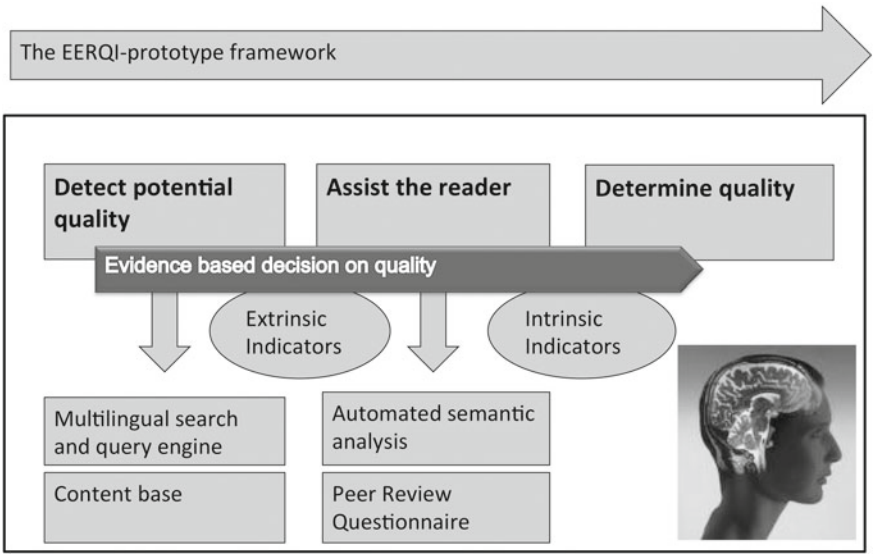


Fig. 1 The EERQI prototype framework

The elements of the EERQI Prototype Framework can be used during the process of quality detection. After detecting and identifying relevant texts, different approaches to consolidate a judgment on quality can be applied. The EERQI project distinguished two different types of indicators that are relevant in these approaches: one type that is external to the text, such as bibliometric and webometric features; and another type that is internal in the text—namely the signals that are given within the words, graphs, metaphors of which the text is composed. The application of EERQI tools in the process can be illustrated as follows:

1. detection of potential quality via the identification of relevant educational research texts in different sources. In this step, the EERQI content base (educational research texts provided by the EERQI publisher partners) and the multilingual search and query engine can be applied.
2. gathering information on extrinsic features of a text. For this purpose, an instrument called 'aMeasure' was developed (by EERQI partner Humboldt University). This is a stack of tools and programs which indicate extrinsic characteristics of research publications (such as citations, webmentions) by using different sources (e. g. Google Scholar, Google Web Search, MetaGer, LibraryThing, Connotea, Mendeley and citeulike) and combining their results, thus providing more comprehensive and less (but still!) biased information.
3. supported transverse reading, allowing for quick information on the usefulness or quality potential of a text. For this step, a linguistic technology in order to provide automatic support for evaluating the quality of a text was developed (by EERQI partner XEROX). The method allows for the automatic identification of key sentences to indicate parts of documents to which peer reviewers should pay particular attention (automated semantic analysis). The respective tests in the EERQI-project showed that this method is especially efficient for the identification of the *bad* quality of a text. It can reduce the time that has to be spent on a text in a review process considerably (up to two thirds of reading time).
4. support of a peer review process. For this step, the EERQI project developed a questionnaire containing items that operationalize five generic indicators of research quality (EERQI Peer Review Questionnaire). The indicators as well as their operationalization in the questionnaire have been tested for reliability, practicality and acceptance in the education research community—with very satisfactory results.

The elements of the EERQI Prototype Framework can either be applied as single methods for specific parts of an assessment process; or they can be applied consecutively, leading to a final judgment on the basis of intense reading of selected texts.

3 Conclusion

The approaches that were generated and tested in the EERQI project open up prospects for future developments that can meet the practical needs in accelerating assessment processes and make them better manageable as well as more transparent. Both are necessary, not least because the number and aspiration of such processes are continuously growing. The intelligent combination of qualitative and quantitative approaches, and the multilingual functionalities of the EERQI products, open up the vision that sets of tools can be made available, allowing for well-informed, evidence based judgments on research quality that are supported by technical tools. The application of the tools can accelerate the process and increase transparence—but cannot replace the human judgment. There cannot be any doubt that the EERQI experimental approach had some methodological limitations (see for example the contributions by Mooij (2014) or by Severiens and Hilf (2014) in Gogolin et al. (2014)). Nevertheless, the present empirical outcomes of the project are promising for future EERQI developmental and research activities, which could, for example, also integrate semantic latent factors and indicators. The approaches that were developed and tested in EERQI show encouraging possibilities to appraise Europe's multicultural and multilingual heritage in research, especially in the social sciences and humanities.

References

Bridges, D. (2009). Research quality assessment. Impossible science, possible art? *British Educational Research Journal*, *35*(4), 497–517. doi:10.1080/01411920903111565.

Bridges, D., Smeyers, P., & Smith, R. (Eds.). (2009). *Evidence-based educational policy. What evidence? What basis? Whose policy?*. Oxford: Wiley Blackwell.

European Research Council. (2011). ERC Grant Schemes Guide for Applicants for the Advanced Grant 2011 Call. Retrieved from http://ec.europa.eu/research/participants/portal/doc/call/fp7/erc-2011-adg_20110406/30876-final_gfa_adg2011_en.pdf.

Gogolin, I. (2012). Identificación de la calidad en las Publicaciones de Investigación Educativa. Proyecto Europeo sobre los Indicadores de Calidad en la Investigación Educativa (EERQI). *Revista de Investigación Educativa*, *30*(1), 13–27. doi:10.6018/rie.30.1.140812

Gogolin, I., Åström, F., & Hansen, A. (Eds.). (2014). *Assessing quality in European educational research. Indicators and approaches*. Wiesbaden: Springer VS.

Gogolin, I., & Lenzen, D. (Eds.). (2014). *Qualitätsmessung im Bildungs- und Wissenschaftssystem*. Wiesbaden: Springer VS.

Mocikat, R. (2010). Qualitätsbewertung in den Naturwissenschaften mithilfe quantitative Parameter. *Journal der Sächsischen Akademie der Wissenschaften*, *5*, 90–102. Retrieved from http://repo.saw-leipzig.de:80/pubman/item/escidoc:16291/component/escidoc:16290/denkstroeme-heft5_90-102_mocikat.pdf.

Mooij, T. (2014). A prototype empirical framework of intrinsic and extrinsic EERQI indicators. In I. Gogolin, F. Åström, & A. Hansen (Eds.), Assessing quality in European educational research. Indicators and approaches (pp. 121–138). Wiesbaden: Springer.

Oancea, A. (2014). Research assessment as governance technology in the United Kingdom. Findings from a survey of RAE 2008 impacts. *Zeitschrift für Erziehungswissenschaft*, *17*(6 Supplement), 83–110. doi:10.1007/s11618-014-0575-5

Severiens, T., & Hilf, E. R. (2014). A scientific editors's support tool: Design, analysis and value. In I. Gogolin, F. Åström, & A. Hansen (Eds.), Assessing quality in European educational research. Indicators and approaches (pp. 139–164). Wiesbaden: Springer VS.

Part III
Bibliometrics in the Humanities

Beyond Coverage: Toward a Bibliometrics for the Humanities

Björn Hammarfelt

Abstract In this chapter, the possibility of using bibliometric measures for evaluating research in the humanities is pondered. A review of recent attempts to develop bibliometric methods for studying the humanities shows that organizational, epistemological differences as well as distinct research practices in research fields ought to be considered. The dependence on colleagues, interdisciplinarity and the 'rural' nature of research in many humanistic disciplines are identified as factors that influence the possibilities of applying bibliometric methods. A few particularly promising approaches are highlighted, and the possibility of developing a 'bibliometrics for the humanities' is examined. Finally, the intellectual characteristics of specific disciplines should be considered when quality indicators are constructed, and the importance of including scholars from the humanities in the process is stressed.

1 Introduction

In this chapter, I argue that bibliometric research on the humanities is now slowly maturing. It appears as if the field is gradually moving from analyzing coverage to a new line of inquiry that tries to understand the humanities on its own terms: looking at specific fields rather than a large heterogeneous collection of disciplines gathered under the label of 'the humanities' or 'the social sciences and the humanities' (SSH). This new line of research refrains from the familiar, but sometimes unfortunate, distinction between the humanities and the natural sciences, and in doing so abandons the common practices of portraying the social sciences and the humanities as the 'other' that does not fit into the bibliometric universe.

The additional focus on the actual characteristics of disciplines has led to attempts to develop bibliometric approaches that are sensitive to the organization of research fields in the humanities. Examples of such attempts include the use of non-source items in established citation databases such as Web of Science (Hammarfelt 2011; Linmans 2010), the use of alternative databases like Google

B. Hammarfelt (✉)
University of Borås, Allégatan 1, Borås, Sweden
e-mail: Bjorn.Hammarfelt@hb.se

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_10

Scholar (Kousha and Thelwall 2009; Koshua et al. 2011) and the recent exploration of the possibilities that the new Book Citation Index offer (Gorraiz et al. 2013; Leydesdorff and Felt 2012). These efforts include exploration of local databases (Engels et al. 2012), references in grant applications (Hammarfelt 2012b), book reviews (Zuccala and van Leeuwen 2011) as well as inclusion in library catalogues (White et al. 2009). Recently, the possibilities that altmetrics offer for the humanities have also been investigated (Hammarfelt 2014; Holmberg and Thelwall 2013; Mohammadi and Thelwall 2013).

The broadening of quality criteria as well as the inclusion of many different types of approaches and materials appear promising. However, this chapter highlights aspects other than methods, materials and coverage as it emphasizes the purpose and organization of research. Thus, I claim that coverage is not the only issue, and maybe not even the most problematic one when discussing the use of bibliometrics on research fields gathered under the heading 'humanities'.

I begin by outlining the background of bibliometric research on the humanities. I do not claim this overview—which is partly adopted from my dissertation (Hammarfelt 2012a)—to be an extensive review of previous research; instead, I sketch out some of the main findings on the topic. Following this short overview, I discuss recent attempts to develop bibliometric methods that are in tune with research practices in the humanities. These include novel databases, new sources and methods as well as already implemented evaluation systems. In the subsequent section, I introduce theoretical concepts for relating the organization of research fields to publication and citation patterns. Whitleys (2000) theory on the intellectual organization of research as well as Becher and Trowlers (2001) characterization of academic tribes are explicated in this context. I then use these concepts to explain the organization of research in the humanities and its implications for bibliometric measures. Finally, I examine the possibilities of establishing a bibliometrics for the humanities and propose a few suggestions for future research.

1.1 The Humanities

The definition of research fields as either social science or humanities is governed by institutional and epistemological considerations, which further depend on the organization of research in countries or regions. The lists of fields defined as the humanities differ between contexts and countries. The Organization for Economic Co-operation and Development (OECD) lists history, archaeology, genealogy, literature, languages, philosophy, arts, history of arts, religion and theology (OECD 2002, p. 68) while The European Reference Index for the Humanities (ERIH) distinguishes fifteen fields in the humanities (including educational research as well as gender studies and psychology). In the United States, however, the Humanities Resources Center includes eleven fields (Leydesdorff et al. 2011).[1]

[1]These fields are English language and literature, foreign languages and literature, history, philosophy, religion, ethnic-, gender- and cultural studies, American studies and area studies, archeology, jurisprudence, selected arts and selected interdisciplinary studies.

Due to the blurry boundaries of the humanities and the ever-changing disciplinary landscape, no definite collection of fields in the humanities can be given. However, a core of fields—that are on all 'lists'—can be distilled: art, philosophy, music, language, literary studies and religious studies. These fields are also the ones discussed in this chapter with an additional focus on literary studies. The humanities is a heterogeneous collection of disciplines, and major differences exist between journal-based fields such as linguistics and more book-based fields such as literary studies and religious studies. The conclusions drawn in this chapter concern the latter disciplines rather than more journal-oriented fields such as linguistics and philosophy. I take the liberty of using the term 'the humanities' as the topic of enquiry, and this is in line with the majority of previous research on this theme. At the same time, I recognize and discuss the problems that such an approach entails.

2 Bibliometric on the Humanities: A Short Recapitulation

Historically, bibliometric research on the humanities has focused mainly on the inadequate coverage of publications by humanities scholars in available citation databases.[2] Several reasons for the scant coverage are mentioned in the literature on the topic: diverse publication channels, the importance of 'local' languages as well as the wide-ranging audience of research.

The heterogeneous audience of research is an often-asserted characteristic of scholarship in the humanities. A basic division is often made between publications directed toward fellow researchers and writings directed to a public audience. Nederhof distinguishes the audience further (2006, p. 96) into three groups: international scholars, researchers on the national or regional level and a non-scholarly audience. Another often-cited division is the one suggested by Hicks (2004), in which she separates journal articles, books, national and non-scholarly literature. Her categorization—although originally used to characterize scholarly literature in the social sciences—is also used for describing the humanities. The main difference between these two schemes for describing the varied publications channels and the heterogeneous audience of research is that Nederhof focuses on the 'target audience' while Hicks discusses 'types of literatures'. I propose that focusing on the audience rather than the publication channel allows for a discussion that places the role and purposes of the humanities at the forefront. The three groups suggested by Nederhof also have the advantage of not being clearly separated, as a publication potentially could target all three groups. The categories proposed by Hicks, on the other hand, demand a separation between scholarly and non-scholarly literature. It is also unclear

[2]For an orientation in the wider literature on the evaluation of the humanities, the reader can consult the Arts and Humanities Research Assessment *Bibliography* (Peric et al. 2013), which currently has a little over a thousand publications indexed, Nederhof (2006) provides a review of issues regarding bibliometric evaluation, and recently a bibliography of research on the humanities and bibliometrics covering the years 1940–2010 was provided by Ardanuy (2013).

how these groups relate to each other; a book directed to a national and public audience could in theory be categorized as 'book', 'national' and 'non-scholarly' at the same time.

2.1 Publication Patterns

Of special interest in the discussion regarding publication practices in the humanities is the role of the monograph (Lindholm-Romantschuk and Warner 1996; Thompson 2002). The monograph reaches all three audiences to a greater extent than the journal article, and has been deemed especially efficient in targeting non-scholarly readers. Publications directed to a popular audience play an important role, and writing monographs can be seen as an effort to target a scholarly and a popular audience.

However, articles in journals and books are the publication channels most frequently used by researchers in the humanities. Kyviks (2003) study of publication practices among Norwegian scholars in the humanities showed that articles—in books or in periodicals—are the most common output. Articles or chapters in books are also frequent in the social science and the humanities, and a small increase in international (English) and co-authored publications was detected. The recent exploration of publication patterns in the social sciences and humanities in Flanders (Belgium) shows that journal publishing is increasing in the social sciences but decreasing in the humanities. A general increase in the production of publications and especially English language publications was also detected, but no major shift toward publishing in journals was discerned (Engels et al. 2012). Similar results—an increase in the number of international publications (including publications in German or French)—were found in a recent study of publication patterns at the faculty of Arts at Uppsala University in Sweden. Notable from this study was that researchers perceived major changes in publication patterns while the actual changes in publication patterns were small (Hammarfelt and de Rijcke 2015).

2.2 Citing of Sources

A sweeping generalization is that scholars in the humanities mostly publish journal articles and book chapters but cite monographs. Thus, the overlap between citing and cited documents is small in many fields, and it is often reported that scholars in the humanities use older literature as well as primary sources. However, there are notable differences within the humanities in the citing of sources, and the percentage of references to books and edited books varies from 88 % in religion to only 49 % in linguistics (Fig. 1).[3]

[3]Data collected from several previous studies: religion (Knievel and Kellsey 2005), philosophy (Cullars 1998), music (Knievel and Kellsey 2005), literature (Thompson 2002), arts

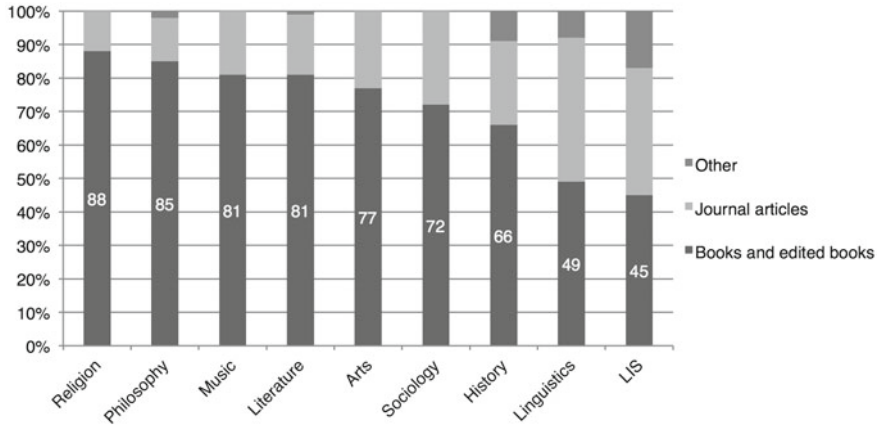


Fig. 1 Percentage of cited books and journal articles in selected fields in the humanities and the social sciences (data from 1995 to 2005). Figure from Hammarfelt (2012a, p. 31)

The earlier findings summarized in the Fig. 1 show that religion, philosophy and literature are book-based disciplines, while journals play an important role in history and linguistics. The overview also shows that books are often cited in social science fields such as sociology and library and information science (LIS). Thus, the problem with counting only citations of journal articles is not restricted to research fields in the humanities.

The extent to which fields in the humanities are adopting referencing practices from the natural sciences has been debated. Larivière et al. (2006) compared the humanities, the social sciences, engineering and the natural sciences in terms of journal publication. The authors found a general increase in journal citations between 1981 and 2000, and this finding applied to the natural sciences and engineering as well as to the social sciences and the humanities. However, when specific fields, such as history, law and literary studies, were examined, a decrease in journal citations during the period was detected.

2.3 The Language and Age of Cited Sources

The language of sources is rarely an issue in the natural sciences since English is the lingua franca. The situation is different in the humanities as many fields in the social sciences and the humanities have a strong regional or national orientation. This is the case especially in fields such as literary studies, sociology and political science (Nederhof 2006 citing Luwel et al. 1999). Databases that predominately

(Footnote 3 continued)
(Knievel and Kellsey 2005), history (Lowe 2003), sociology (Lindholm-Romantschuk and Warner 1996), LIS (Chung 1995) and linguistics (Georgas and Cullars 2005).

index English-language sources cannot adequately cover these fields, and this is a major issue when using established databases such as Web of Science or Scopus to study research fields in the humanities.

Literary studies are a field in which non-English sources play a major role. The influence of English-language sources is moderate: Less than 15 % of the cited sources in German literature and only 9 % of the cited sources in French literature are in English (Cullars 1989). Swedish literary studies has a higher percentage of citations of English-language sources (between 43 and 54 %), but Swedish as well as German and French sources are frequently cited (Hammarfelt 2012b). Consequently, studies of these fields must incorporate non-English sources, and the same applies to many other countries and research fields.

Scholars in the humanities use sources that cover a wide age span. The age of the sources used in research is related to the search for literature, and the pressure to keep up with current research is less pronounced. Thus, a research front is hard to discern, and long time windows are needed when conducting bibliometric analyses. De Solla Price explained the difference in the 'consumption' of sources by using a metaphor of digestion: 'With a low index one has a humanistic type of metabolism in which the scholar has to digest all that has gone before, let it mature gently in the cellar of wisdom, and then distill forth new words of wisdom about the same sort of questions' (de Solla Price 1970, p. 15). This characterization disregards the diversity of research in the humanities, although the metaphor of digestion is illustrative. Furthermore, Price overlooked that many sources in the humanities are primary sources (for example, historical sources and literary works), which increases the median age of the sources considerably.

Bibliometric studies of the humanities show that the type of publication most frequently cited is the monograph, the age span of the cited references is broad and languages other than English play a significant role in many fields (Hammarfelt 2012a). These characteristics are agreed upon by many, but several matters remain unresolved. One question is whether the publication practices of scholars in the humanities are adapting to the norms that prevail in the natural sciences. A few studies (Butler 2003; Kyvik 2003) suggest that this might be the case, while others emphasize the constancy of cited and published material (Hammarfelt and de Rijcke 2015; Larivière et al. 2006). How the increasing importance of 'research outputs' across research fields will influence publication practices in the humanities has not been determined. However, implementing publication-based performance measures will undoubtedly put further focus on this issue, and perhaps this will lead to in-depth studies of the effect that evaluation systems have on scholarship in the humanities.

3 In Pursuit of a Bibliometric for the Humanities

In this section, I briefly present several recent attempts to apply bibliometric methods to the humanities. In addition to being current, the selected studies also have a further sensitivity to the characteristics of research in the humanities in common.

Thus, these studies are not only examples of bibliometrics applied to the humanities but also to some extent examples of bibliometric methods developed 'for the humanities'. A general feature of these attempts are an effort to introduce new sources for bibliometric analysis, sources that go beyond journals indexed in citation databases such as Thomson Reuters' Web of Science or Elsevier's Scopus.

3.1 Book Citation Index

An obvious solution to the problem of low coverage of non-journal publications in citation indexes is to start indexing books. The launch of the Book Citation Index in 2011 is an attempt to improve the coverage of the humanities, and it could open up for analysis of how the journal literature and the book literature relate to each other. However, the index still has a very limited scope, mainly English-language sources are included (Gorraiz et al. 2013), and problems remain when distinguishing between different types of books. Initial studies have also found that the citation rates of books are low in many research fields (Leydesdorff and Felt 2012). Thus, the current Book Citation Index is of little use for evaluating research but might provide valuable knowledge regarding the relation between journal literature and books.

3.2 Non-source Items

It was possible to track citations of books that are not indexed in citation databases, before the launch of the Book Citation Index. Citation of so-called 'non-source' items has been used for studying impact and interdisciplinarity (Hammarfelt 2011; Linmans 2010). However, this method involves limitations on the size of the material used, and considerable data cleaning is needed, since the cited sources are not standardized. Another constraint of this method is that it gathers citations only from a small portion of the literature in many research fields in the humanities. The approach is in principle restricted to English-language publications, and the analysis of 'non-source' items is limited to small data sets due to the manual work involved.

3.3 Google Scholar, Google Book Search

An alternative to the use of traditional citation indexes is options such as Google Scholar (GS) or Google Book Search (Kousha and Thelwall 2009; Koshua et al. 2011). The main constraints of GS are that analyses cannot be automatized and the data is hard to process. Every post has to be checked, and new searches for each publication are required. The benefit of Google Scholar is greater coverage—which includes books—and that everyone is free to use the database (with limitations

on what you can do). The reliability of the data is a concern since inflated citation counts as well as ghost authors and 'phantom authors' limit the usability of the data for bibliometric analysis (Jacso 2010).

3.4 Ad Hoc Databases

A response to the limits of existing data sources is to build your own citation database. When targeting specific contexts—Catalan literature (Ardanuy et al. 2009) or Swedish literary studies (Hammarfelt 2012b)—this method might be viable. The building of 'ad hoc databases' allows analyses of materials that usually are not indexed in citation indices such as grant applications (Hammarfelt 2012b), and small local studies can provide valuable contrast to larger studies of citation patterns. However, the amount of labor involved in harvesting references by hand and then indexing them in a database inherently limits the size of the datasets used.

3.5 Library Catalogues

Several authors have suggested that library catalogues might be a possible data source for evaluating the impact of books (Linmans 2010; Torres-Salinas and Moed 2009; White et al. 2009). The basic idea is simple: The more libraries that stock a book, the more influential it is deemed to be. The inclusion of a book in a catalogue indicates that the book is judged important. However, implementing the model on a larger scale would be difficult: Libraries do not always make informed judgments when buying books; they often buy bundles of books. The model does not include open access or e-books, and an evaluation system based on this approach would put the librarians making the buying decisions in a delicate position. Furthermore, one could imagine that authors and publishers could easily manipulate such a system.

3.6 Book Reviews

Book reviews have an important gatekeeping function in the humanities, and reviews are often seen as an important merit and indicator of influence for the author writing the review. Book reviews have also been proposed as an important unit of analysis when it comes to book-oriented fields. Zuccala and van Leeuwen (2011) proposed that the number of book reviews produced by a researcher can be seen as a measure of success. One problem though is that already established and older researchers often are those invited to review books. Thus, a system that counts written reviews could disadvantage younger and less renowned scholars. Another alternative

is to view book reviews as 'mega-citations' that indicate the quality of a book (Zuccala et al. 2014). This approach has many advantages, especially since book reviews play an important function in the humanities; however, many books are never reviewed, and the overall coverage is possibly too low for systematic assessment.

3.7 Counting and Weighing Publications

An alternative of course is to not use citations at all and instead count publications. This system makes it possible to evaluate research in all fields independently of publication channel and language. A qualitative aspect can be introduced in order to circumvent the flourishing of low-quality publications. The idea of weighing publication according to type and channel has been proposed by Finkenstaedt (1990) and Moed et al. (2002). However, the most well-known and influential system for counting and weighing publications is the Norwegian one (Schneider 2009; Sivertsen 2010). This system is used for allocating resources among universities in Norway. The main benefits of the system are the coverage of publications, transparency and the adaptability of the system (Ahlgren et al. 2012). However, many publications in the humanities are still not included due to the definition of 'scholarly literature', and monographs at prestigious 'non-academic' publishers are seldom counted. The consequence is that a lower share of the total publications by humanities scholars is covered by the system. This disadvantage is partly compensated by publications being fractionalized over authors, which has shown to benefit scholars in the humanities compared to disciplines where co-authorship is common (Piro et al. 2013).

3.8 Altmetric Approaches

Altmetrics—metrics based on data from the social web—is a promising approach in the efforts to find appropriate methods for assessing the humanities (Tang et al. 2012). These new, 'altmetric' measures propose not only to solve problems with established methods but also to measure impact beyond citations from academic journals. One of the most popular data sources used for altmetric analysis is Twitter. Holmberg and Thelwall (2013) found that scholars in the history of science were less likely to use Twitter for scholarly purposes compared with other fields, and across all fields, few tweets contained links or mentions of scholarly literature. Another common source of altmetric data is the social reference manager Mendeley, but the coverage for humanities articles was also quite low (28 %) when compared to the social sciences (58 %) (Mohammadi and Thelwall 2013). The inclusion of many different types of sources, the ability to study impact beyond the scholarly realm, as well as the openness of many services appear promising for the humanities. However,

limitations remain with the dominance of English-language journal articles the most significant (Hammarfelt 2014).

There is no shortage of approaches for studying the humanities with bibliometric methods, and the brief orientation given here is not exhaustive. Still, the overview illustrates that bibliometric research depends on the availability of data sources, especially citation indices, and the content, availability and coverage of these data sources dictate how research is conducted. Thus, many of the studies mentioned were influence by the introduction of new services such as Google Book Search, Google Scholar, or The Book Citation Index. The research field of bibliometrics can be duly criticized for its dependence and focus on available data sources, even more as these services are provided by private companies and, thus, are not easily adapted to the fields needs by scholars themselves. However, the main purpose of bibliometric research is not to study databases or coverage, but to further our understanding of communication structures in science and research. In this effort, we have to go beyond issues of database content and coverage and focus on the organization and characteristics of research in different disciplines. Accordingly, in the following chapter I reflect on publication patterns and referencing practices in relation to the social and intellectual organization of research fields.

4 Intellectual Organization of Research Fields and Its Bibliometric Consequences

In the following section, I describe how publication practices and citation patterns can be understood from a disciplinary perspective where the use of references depends on how a research field is organized. The characterization of research fields in the humanities suggested by Whitley (2000) and Becher and Trowler (2001) is briefly reviewed, and related to publication patterns and referencing practices. However, the vast difference between research fields and subfields gathered under the umbrella of the humanities should be acknowledged, and the generalizations made here apply foremost to literary studies and similar book-based disciplines.

4.1 Fragmented and Rural Research Fields

The majority of disciplines within the humanities are in Whitleys characterization defined as fragmented adhocracies. These fields are intellectually varied as well as heterogenic since research in *fragmented adhocracies* is personal and poorly coordinated, and the degree of specialization is limited. The dominant attribute of these fields is the lack of a stable configuration; tasks are not specialized; co-ordination is weak, and when it occurs, it is based on personal relations (Whitley 2000). Subgroups form around specific topics and discrete methodological approaches. Audiences are

varied, and so are the methods used. There is considerable disagreement on which topics to study as well as on how these topics should be approached, and the lack of standards makes it difficult to resolve disputes.

Another useful characterization for understanding the organization of research fields is the one between rural and urban fields (Becher and Trowler 2001). The distinction between rural and urban concerns the 'density' of a discipline or a research area; if many researchers are working on the same problem, then the research area is described as urban, while a less populated discipline is deemed rural. Strong competition for positions and resources can be observed in an urban research area (for example, biomedicine), whereas there are fewer struggles for resources and recognition (as well as fewer rewards) in rural fields.

4.2 Referencing Practices and Citation Patterns

I propose that referencing practices and citation patterns are further understood by the intellectual characteristics of the research field: A less demarcated discipline lacking a central core is heavily influenced by other research fields and therefore more interdisciplinary in referencing practices. Citation patterns are also determined by the number of researchers engaged on a specific topic: In an urban field, it is important to keep up with the 'research front' and cite recent literature, while the age of sources plays less of a role in rural fields. This is also connected to the speed of publication, which is considerably faster in an urban field (biomedicine) than in a rural one (literary studies) (Table 1).

Another variable that influences referencing practices is the audience. In fields where a non-academic audience plays an important role, scholars may choose a referencing style—the footnote is an example—that serves a scholarly and a popular audience. The degree of dependence between researchers and the definition of originality also affect the use of references. It is important to cite colleagues in a field where researchers depend on each other for recognition and rewards, but in fields where originality is highly valued, referencing serves other purposes as well (Hellqvist 2010).

Table 1 Characteristics of the humanities and influence on publication and citation patterns

Field characteristics	Publication patterns	Referencing practices
Low dependence on colleagues	Various publication channels; importance of public audience	Interdisciplinary references common
Rural organization	The pace of publications is slow	Citations gather slowly; number of 'possible citations' is low

Thus, two main characteristics that influence referencing practice and citation patterns in the humanities can be discerned: *low dependence on colleagues* and the *rural organization of the field*. The varied audience, rural organization and low dependence on colleagues are related. A diverse audience makes it possible for individual researchers to find readers outside their own field, with the consequence that scholars depend less on peers for recognition. The high task uncertainty of many fields in the humanities and the low dependence on colleagues give the individual scholar great freedom in pursuing a unique research profile, which results in researchers being scattered across many different topics with little communication between them. Thus, scholars in the humanities enjoy many possibilities when selecting topics, publication channels and whom to cite, but this in turn limits the potential of receiving 'rewards' in the form of citations. The low coverage of publications in citation databases is therefore not the most important reason why citation scores are less applicable as an indicator of impact in the humanities. Instead, I propose that the social and intellectual organization of the humanities is the main reason to why citation-based approaches are less applicable in these fields.

5 Conclusions

The bibliometric community has rightly discouraged the use of conventional bibliometric methods for evaluating the humanities. Especially, citation analysis using journals indexed in citation databases is less applicable in these fields. This conclusion is firmly based on several studies showing that the coverage of the humanities in databases such as Web of Science or Scopus is insufficient for evaluation and not representative of research in the humanities. Research assessment systems, such as the one used in Norway, amend this by including all scholarly publications. The publications are then given points depending on the publication channel (monograph, anthology, or journal) and the 'quality level' of the journal or the publisher. However, the definition of what should count as a 'scholarly publication' is still a matter of debate. There is no consensus on what an important research output is in the humanities; a peer-reviewed journal article in an international journal, a book chapter in an anthology edited by a renowned scholar, or a monograph at a prestigious non-academic publisher can all be seen as important outputs, and publications directed toward a popular audience are often highly rated. Consequently, the choice of publications that should be valued in assessing research depends on our view of the humanities and its overall purpose in society.

A recurrent problem in evaluating the humanities is the long time span needed for measuring the impact of research. The lifetime, as well as the distribution of citations to a publication over time, must be considered. Research by humanities scholars may be used in twenty, fifty, or even a hundred years, but sustainability is seldom measured in research assessment exercises. Thus, a considerable part of research in

the humanities—such as the preservation and translation of cultural heritage—might be valuable for future generations, but it is invisible in the limited perspective of research evaluation.

The development of bibliometric methods that fairly capture the 'impact' of research involves understanding how research is organized in these fields. This is confirmed by the findings recapitulated that point to differences in intellectual organization, and in the actual use of references as major reasons for why citation-based approaches are less applicable to the humanities. Thus, in developing bibliometric methods that accurately depict the humanities, we must go beyond the issue of coverage and focus on the social and intellectual organization of the fields involved. However, there are vast differences in research practices within the humanities, and differences are also evident among specialties within the same discipline. Furthermore, research practices are constantly changing due to technical developments (digitalization), external demands (research evaluation, open access) and internal negotiations on the purpose of research. Research on scholarly communication—including bibliometric approaches—is needed in order to follow these developments. Furthermore, when studying scholarly practices, we must be careful not to be caught in old dichotomies that portray 'two cultures', but acknowledge that research across all disciplines shares many similarities. The need for fair and reliable assessment methods cuts across all research fields, and constructing indicators that properly capture the quality and impact of research is challenging for academia at large.

Constructing appropriate indicators involves actively engaging the researchers being evaluated. Recent attempts at identifying quality indicators in the humanities show that the 'notion' of quality is not easily captured, and several conflicting norms were found (Ochsner et al. 2013). The construction of general and all-encompassing indicators is hindered by the heterogeneous nature of research as well as differences in how quality is perceived. However, alternatives to the use of peer review, which not only is time-consuming but also prone to reinforce established hierarchies, are needed in the humanities. Here I believe evaluations that use bibliometrics might provide a valuable complement to traditional peer review, but only if the indicators used are carefully constructed in a dialog with the researchers being evaluated.

5.1 Challenges

Bibliometrics may play an important role in future attempts to study the wider impact of research in the humanities, and citation analysis could be used to further our understanding of the organization and development of research in these fields. Approaches such as using citations to 'non-source items', introducing new databases and services, and using altmetric measures all appear promising but are far from utilizable on a general level. These and several other innovative techniques for studying the humanities have been identified in this chapter, and one argument made is that bibliometric

research on the humanities has become more attuned to the scholarly tradition of humanistic scholarship. Still, much must be done to study and assess the humanities, and I identify a few areas that are particularly interesting for future research.

First, I suggest that it is time to devote attention to more detailed and restricted areas of research. It is less complicated to define fields and delineate 'subfields' in the natural sciences, and this might be one reason for using a broad and inclusive definition when studying the humanities. Extensive interdisciplinary citing might be another reason for adopting 'the humanities' as the object of study. However, I propose that focusing further on specific fields and specialties would yield a better understanding of publication and citation patterns in the humanities. I also envision that developing new and more accessible bibliometric tools and approaches will result in further application of bibliometric methods by humanist scholars themselves.

Altmetric methods that are in tune with the organization of the humanities is an additional area for research. Attempts at actually systematically measuring social impact—impact outside academia—are promising. Such measures would be an important contribution not only for assessing the humanities but also for measuring the general influence of research in society. Exploring sources, mainly books and non-English language publications that are seldom covered by traditional bibliometric approaches is another exciting vein of research. Altmetrics is a very novel phenomenon and its ability to measure quality or impact is still debated, but the general ambition of including many different types of sources that measure impact in a multifold of ways is encouraging for the efforts to develop 'metrics' for the humanities.

Finally, the meeting of a 'metric culture' with scholarship in the humanities is a particularly important area of study. For a long time, the natural sciences have lived with impact factors, and researchers in these fields often calculate their own H-index. However, scholars in the humanities are less familiar with bibliometric measures, and many researchers not only fear unfair rankings and evaluations but also often see them as alien to humanistic scholarship. Thus, a crucial topic is how the organization and character of the humanities will respond to additional measurement and assessment attempts. The answer to this question is important not only for the bibliometric community but also for the future of scholarship in the humanities.

Acknowledgments This chapter builds on findings from my dissertation, *Following the Footnotes: A bibliometric analysis of citation patterns in literary studies* (2012), and segments of the text are redrafted and shortened versions of arguments found there.

References

Ahlgren, P., Colliander, C., & Persson, O. (2012). Field normalized rates, field normalized journal impact and Norwegian weights for allocation of university research funds. *Scientometrics*, *92*(3), 767–780. doi:10.1007/s11192-012-0632-x.

Ardanuy, J. (2013). Sixty years of citation analysis studies in the humanities (1995–2010). *Journal of the American Society for Information Science and Technology*, *64*(8), 1751–1755. doi:10.1002/asi.22835.

Ardanuy, J., Urbano, C., & Quintana, L. (2009). Citation analysis of Catalan literary studies (1974–2003): Towards a bibliometrics of humanities studies in minority languages. *Journal of the American Society for Information Science and Technology*, *81*, 347–366. doi:10.1007/s11192-008-2143-3.

Becher, T., & Trowler, P. R. (2001). *Academics tribes and territories: Intellectual enquiry and the cultures of disciplines*. Buckingham: Open University Press.

Butler, L. (2003). Explaining Australia's increases share of of ISI publications—the effects of a funding based on publication counts. *Research Policy*, *32*, 143–155. doi:10.1016/S0048-7333(02)00007-0.

Chung, Y.-K. (1995). Characteristics of references in international classification systems literature. *The Library Quarterly*, *65*(2), 200–215.

Cullars, J. M. (1989). Citation characteristics of French and German literary monographs. *Library Quarterly*, *59* (305–325).

Cullars, J. M. (1998). Citation characteristics of English-language monographs in philosophy. *Library and Information Science Research*, *20*(1), 41–68. doi:10.1016/S0740-8188(98)90005-6.

de Solla Price, D. J. (1970). Citation measures of hard science, soft science, technology, and non-science. In C. E. Nelson & D. K. Pollock (Eds.), *Communication among scientists and engineers* (pp. 3–22). Lexington, MA: Heath.

Engels, T. C., Ossenblok, T. L., & Spruyt, E. H. (2012). Changing publication patterns in the social sciences and humanities, 2000–2009. *Scientometrics*, *93*(2), 373–390. doi:10.1007/s11192-012-0680-2.

Finkenstaedt, T. (1990). Measuring research performance in the humanities. *Scientometrics*, *19*(5–6), 409–417. doi:10.1007/BF02020703.

Georgas, H., & Cullars, J. M. (2005). A citation study of the characteristics of the linguistics literature. *College and Research Libraries*, *66*(6), 496–515. doi:10.5860/crl.66.6.496.

Gorraiz, J., Purnell, P. J., & Glänzel, W. (2013). Opportunities for and limitations of the book citation index. *Journal of the American Society for Information Science and Technology*, *64*(7), 1388–1398. doi:10.1002/asi.22875.

Hammarfelt, B. (2011). Interdisciplinary and the intellectual base of literature studies: Citation analysis of highly cited monographs. *Scientometrics*, *86*(3), 705–725. doi:10.1007/s11192-010-0314-5.

Hammarfelt, B. (2012a). Following the footnotes: a bibliometric analysis of citation patterns in literary studies. (Doctoral dissertation. Skrifter utgivna vid institutionen för ABM vid Uppsala Universitet, Vol. 5). Uppsala: Uppsala Universitet. Retrieved from http://www.diva-portal.org/smash/get/diva2:511996/FULLTEXT01.pdf.

Hammarfelt, B. (2012b). Harvesting footnotes in a rural field: citation patterns in Swedish literary studies. *Journal of Documentation*, *68*(4), 536–558. doi:10.1108/00220411211239101.

Hammarfelt, B. (2014). Using altmetrics for assessing research impact in the humanities. *Scientometrics*, *101*(2), 1419–1430. doi:10.1007/s11192-014-1261-3.

Hammarfelt, B., & de Rijcke, S. (2015). Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. *Research Evaluation*, *24*(1), 63–77. doi:10.1093/reseval/rvu029.

Hellqvist, B. (2010). Referencing in the humanities and its implications for citation analysis. *Journal of the American Society for Information Science and Technology*, *61*(2), 310–318. doi:10.1002/asi.21256.

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.

Holmberg, K., & Thelwall, M. (2013). Disciplinary differences in Twitter scholarly communication. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger,& H. F. Moed (Eds.), Proceedings of ISSI 2013 Vienna: 14th International Society of Scientometrics and Informetrics Conference, Vol. 1, pp. 567–582. Vienna: Austrian Institute of Technology GmbH.

Jacso, P. (2010). Metadata mega mess in Google Scholar. *Online Information Review*, *34*(1), 175–191. doi:10.1108/14684521011024191.

Knievel, J. E., & Kellsey, C. (2005). Citation analysis for collection development: A comparative study of eight humanities fields. *The Library Quarterly*, *75*(2), 142–168.

Koshua, K., Thelwall, M., & Rezaie, S. (2011). Assessing the citation impact of books: The role of Google Books, Google Scholar and Scopus. *Journal of the American Society for Information Science and Technology*, *62*(11), 2147–2164. doi:10.1002/asi.21608.

Kousha, K., & Thelwall, M. (2009). Google book search: Citation analysis for social science and the humanities. *Journal of the American Society for Information Science and Technology*, *60*(8), 1537–1549. doi:10.1002/asi.21085.

Kyvik, S. (2003). Changing trends in publishing behavior among university faculty, 1980–2000. *Scientometrics*, *58*(1), 35–48. doi:10.1023/A:1025475423482.

Larivière, L., Archambault, È., Gringas, Y., & Vignola-Gagné, È. (2006). The place of serials in referencing practices: Comparing the natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, *57*(8), 997–1004. doi:10.1002/asi.20349.

Leydesdorff, L., & Felt, U. (2012). 'Books' and 'book chapters' in the book citation index (BKCI) and science citation index (SCI, SoSCI, A&HCI). *In Proceedings of the ASIS&T Annual Meeting*, Vol. 49. doi:10.1002/meet.14504901027.

Leydesdorff, L., Hammarfelt, B., & Salah, A. A. A. (2011). The structure of arts and humanities citation index: A mapping on the basis of aggregated citations among 1157 journals. *Journal of the American Society for Information Science and Technology*, *62*(12), 2414–2426. doi:10.1002/asi.21636.

Lindholm-Romantschuk, Y., & Warner, J. (1996). The role of monographs in scholarly communication: An empirical study of philosophy, sociology and economics. *Journal of Documentation*, *52*(4), 389–404. doi:10.1108/eb026972.

Linmans, J. A. M. (2010). Why with bibliometrics the humanities does not need to be the weakest link. Indicators for research evaluation based on citations, libarary holdings and productivity measures. *Scientometrics*, *83*(2), 337–354. doi:10.1007/s11192-009-0088-9.

Lowe, S. M. (2003). Reference analysis of the American Historical Review. *Collection Building*, *22*(1), 13–20. doi:10.1108/01604950310457168.

Luwel, M., Moed, H. F., Nederhof, A. J., De Samblanx, V., Verbrugghen, K., & van der Wurff, L. J. (1999). *Towards indicators of research performance in the social sciences and humanities. An*

exploratory study in the fields of law and linguistics at Flemish universities. Leiden: Centre for Science and Technology Studies.

Moed, H. F., Luwel, M., & Nederhof, A. J. (2002). Towards research performance in the humanities. *Library Trends*, *50*(3), 498–520.

Mohammadi, E., & Thelwall, M. (2013). Assessing the Mendeley readership of social sciences and the humanities. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. F. Moed (Eds.), *Proceedings of ISSI 2013 Vienna: 14th International Society of Scientometrics and Informetrics Conference*, Vol. 1, pp. 200–214. Vienna: Austrian Institute of Technology GmbH.

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. (2013). Four types of research in the humanities: Setting the stage for research quality criteria in the humanities. *Research Evaluation*, *22*(4), 79–92. doi:10.1093/reseval/rvs039.

OECD. (2002). *Frascati manual: Proposed standard practice for surveys on research and experimental development*. Paris: OECD.

Peric, B., Ochsner, M., Hug, S. E., & Daniel, H.-D. (2013). *Arts and Humanities Research Assessment Bibliography (AHRABi)*. Zürich: ETH Zurich. doi:10.3929/ethz-a-010610785.

Piro, F. N., Aksnes, D. W., & Rorstad, K. (2013). A macro analysis of productivity differences across fields: Challenges in the measurement of scientific publishing. *Journal of the American Society for Information Science and Technology*, *64*(2), 307–320. doi:10.1002/asi.22746.

Schneider, J. W. (2009). An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway. *European Political Science*, *8*(3), 364–378. doi:10.1057/eps.2009.19.

Sivertsen, G. (2010). A performance indicator based on complete data for scientific publication output at research institutions. *ISSI Newsletter*, *6*(1), 22–28.

Tang, M., Wang, C., Chen, K., & Hsiang, J. (2012). Exploring alternative cyber bibliometrics for evaluation of scholarly performance in the social sciences and humanities in Taiwan. In: *Proceedings of the ASIS&T Annual Meeting*, Vol. 49. doi:10.1002/meet.14504901060.

Thompson, J. W. (2002). The death of the scholarly monograph in the humanities? Citation patterns in literary scholarship. *Libri*, *52*(3), 121–136. doi:10.1515/LIBR.2002.121.

Torres-Salinas, D., & Moed, H. F. (2009). Library catalogue analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in economics. *Journal of Informetrics*, *3*(1), 9–26. doi:10.1016/j.joi.2008.10.002.

White, H. D., Boell, S. K., Yu, H., Davis, M., Wilson, C. S., & Cole, F. T. H. (2009). Libcitations: a measure for comparative assessment of book publications in the humanities and social sciences. *Journal of the American Society for Information Science and Technology*, *60*(6), 1083–1096. doi:10.1002/asi.21045.

Whitley, R. (2000). *The intellectual and social organization of the sciences* (2nd ed.). Oxford: Oxford University Press.

Zuccala, A., & van Leeuwen, T. (2011). Book reviews in humanities research evaluations. *Journal of the American Society for Information Science and Technology*, *62*(10), 1979–1991. doi:10.1002/asi.21588.

Zuccala, A., van Someren, M., & van Bellen, M. (2014). A machine-learning approach to coding book reviews as quality indicators: Toward a theory of megacitation. *Journal of the American Society for Information Science and Technology*, *65*(11), 1643–2330. doi:10.1002/asi.23104.

Quotation Statistics and Culture in Literature and in Other Humanist Disciplines

What Citation Indices Measure

Remigius Bunia

Abstract The humanities display a strong skepticism toward bibliometric evaluations of their quotation practices. This is odd, since their citations partly serve the same purpose as they do in the sciences: They can indicate a beneficial influence on one's own work. In Literature, a still-stranger observation asks for an explanation: Even in the most important journals, the articles receive only an astonishingly few citations. This paper presents some facts about the quotation culture, the low levels of citation and the databases involved. It shows that the low numbers are not a product of deficiencies in data, but should be subject to analysis. In the final discussion, this paper offers two explanations: Either Literature is, in fact, no discipline that should be treated as academic; or Literature is a discipline facing its own imminent intellectual death. Yet it is hoped that other explanations will be found; however, this issue requires further research on the practices in Literature and related fields.

1 Introduction

We face a fascinating, yet strange contradiction in the humanities: On the one hand, they disapprove of any bibliometric assessments of academic performance, and, on the other hand, they cherish quotations as a core component of their academic culture. Their dissatisfaction with quantitative bibliometrics may seem to be a mere matter of principle: The humanities are supposed to avoid numbers wherever they can. But this would be an explanation much too simple to account for the intricacies of the quotation culture in the humanities. What is odd is the fact that many disciplines in the humanities quote but do so very rarely. Particularly, Literature[1] shows a strong dislike for a systematic compilation of references. Literature is an

[1] I use the term *Literature* (uppercase) for all related disciplines: Literary Studies, German Literature, English Literature, Comparative Literature, and so on.

R. Bunia (✉)
Leo-Spitzer-Institut, Zillestraße 81, 10585 Berlin, Germany
e-mail: e-mail@web.bunia.de

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_11

extreme case within the spectrum of humanities but, as such, is characteristic of a specific academic condition. Literature's aversion to bibliometrics seems partly legitimate because statistics can be meaningful only if they rely on sufficiently large numbers. But at the same time, this antipathy raises questions about the academic culture itself. The contradiction could be located in the self-perception of certain disciplines—rather than in a conflict between citational practices and quantitative methods.

In the second section, I will bring forth a historical and systematic argument. It follows the epistemic patterns of the humanities. I will outline the traditions of quoting other works in Literature. These may be compared to the practices in the sciences; and these have to be related to the common critique of quantitative methods. In the third section, I will present some statistical data; I do not create new data but simply use existing information. My focus will be on the small numbers involved, that is, I will show how few quotations actually occur in Literature.

Since I need to combine results from both sections, I will only then proceed to the discussion and reserve for it a section of its own. I will consider possible explanations, some that approve of the citational practices in the humanities and others that are in disfavour of their academic culture. After all, if my initial claim about the intrinsic operational contradictions within the humanities proves true, more research must be undertaken to understand the present-day tense situation.

2 Quotation Culture in the Humanities

2.1 Characteristics of Quotations in the Humanities

Quotations have always been part of the core techniques in Literature. Let me give a short historic overview (for a more detailed version and for references, see Bunia 2011b). Even before the surge of modern science, all philosophical disciplines quoted the 'authorities' and, thus, worshipped canonized authors. Book titles were even invented because Aristotle needed to quote himself (cf. Schmalzriedt 1970). With the advent of the rationalist and empiric movements in the 17th century and their icons, René Descartes and Francis Bacon, respectively, in all disciplines, novelty became prestigious, and both scholars and scientists started quoting their peers rather than Ancient authorities. Not until the late 19th century did quoting that *completely* covers the field become a moral obligation. Before, it was sufficient to cite what lay at hand; it was not the researcher's task to show *blatantly* that he was up to date. The increase of publications led to new worries and, finally, caused the need for citation analysis as pioneered by Eugene Garfield.

In Literature, it has always been mandatory to quote as much as possible to prove that one is well read. In fact, 'monster footnotes' (Nimis 1984) are particularly popular in the humanities: they consist of lengthy enumerations of papers related to the topic of the citing paper (see also Hellqvist 2010, pp. 313–316). As Hüser

(1992) notes, an impressively long list of references is one of the most important prerequisites for a doctoral dissertation to be accepted in Literature. These observations are not in conflict with the (very debatable) claim that humanities, in general, do not aim to convey pieces of 'positive' knowledge (MacDonnald 1994), since it does not matter whether one quotes to present knowledge or more obscure forms of excellence. Since the broad usage of references came up in the 19th century, when humanist disciplines tried to become 'scientific' (Hellqvist 2010, p. 311), the difference between the humanities and the sciences should not be taken to be very strong. In brief, literary scholars are usually expected to quote one another extensively, not to omit any possible reference, and to provide comprehensive lists of preceding publications.

Many disciplines limit the obligation to quote comprehensively to recent years and choose other forms of worship for their great minds (e.g. name of theorems in mathematics, see Bunia 2013). Contrary to this practice, literary scholars often cite old canonical works, thus evoking the very roots of their approach. Even more frequent is the practice of using quotations to signal the in-group the scholar belongs to (see Bunia and Dembeck 2010). This is why publications in Literature (in fact, in all disciplines in the humanities) tend to include large lists of old texts.

Two practices challenge my short outline. First, literary scholars also quote the objects of their investigation, e.g. literary, philosophical, or other texts. These appear in the references, too, thus complicating the analysis (see Sect. 3.3). Second, in very conservative circles—and, fortunately, such circles are not numerous—highly established professors are no longer expected to quote unknown young scholars; they restrict their open quotations to peers equal in rank and to classic authors such as Aristotle (see Bunia 2013).

Reputation is highly important (see Luhmann 1990 [Reprint 1998], p. 247; Ochsner et al. 2013, pp. 83, 84, in particular, item 14 'Research with reception'). As is the case in most disciplines, literary scholars hold intellectual impact on their own community in high esteem (Hug et al. 2013, pp. 374 and 382, for English Literature and German Literature). This is one of the criteria to be used to judge young researchers' performance. Intellectual influence becomes manifest due to quotations. In sum, citation analysis should be a method adequate to the disciplinary traditions of Literature.

2.2 Disapproval of Bibliometrics and of 'Quantities' Per se

The most widespread criticism advanced by scholars in the humanities attacks bibliometric analysis for its inability to measure quality. Unfortunately, this attack suffers from a basic misconception. First, it neglects the circumspection that fuels much of the bibliometric debate. For instance, bibliometric research papers are replete with doubts, questionings and reservations about using bibliometric parameters to rate an individual researcher's intellectual performance (e.g. Bornmann 2013). The central misapprehension, however, is the product of a more fundamental skepticism

that asks: How is it possible that quantitative analysis can account for qualitative evaluations? Consequently, bibliometric analyses are thought to be structurally inadequate to express qualitative judgments.

This deduction is a misconception of citation analysis because it ignores the abstract separation of qualitative judgments and their mapping on quotations. When we look at the impact system prevalent in many disciplines, such as Medicine, we see that the qualitative assessment takes place in peer review. This process is not influenced or even compromised by the impact factor culture (see also Bornmann 2013, p. 3). Of course, the impact factor culture produces, stabilizes and usually boosts the differentiation between journals. The effect is that some journals receive the most attention and the best submissions because these journals have the biggest impact. This eventually means that these journals can have the most rigorous selection process. The decisive factors within the selection process remain 'qualitative', that is, they are not superseded by mathematical criteria. This is why all peer review systems have been repeatedly demonstrated to be prone to failure (see the editorial by Rennie 2002; see also Bohannon 2013).

For review processes to warrant optimal evaluation, it is mandatory that the review process rely on accepted and mutually intelligible criteria. The problems with peer review result from the imperfections of the process: careless reviewers, practical limits of verifiability, or missing criteria. Slightly neglectful reviewers do not impair the review process to a dramatic degree; the review process must no longer, as has been previously done, be mistaken for a surrogate of replications. The combination of peer review and bibliometrics provides a suitable technique to map qualitative evaluations on quantities.

However, the situation is the inverse if disciplinary standards of assessment are deficient. If shared criteria of evaluation are weak and if parochialism prevails, peer review can have negative effects on the *average* quality of evaluations (Squazzoni and Gandelli 2012, p. 273). As a consequence, the humanist disciplines that oppose bibliometrics might be right in doing so—but for the wrong reasons: The only sensible reason to object to bibliometric assessment is to admit an absence of *qualitative* criteria.

2.3 The European Reference Index for the Humanities

The disciplines in the humanities feel increasing pressure from funding agencies and governments to expose their strategies of evaluation (cf. Wiemer 2011). Due to the widespread and virtually unanimous refusal to participate in common ranking systems as those provided by bibliometric analysis, the European Science Foundation (http://www.esf.org) initiated the European Reference Index for the Humanities (ERIH) project. The project decisively dismisses all statistical approaches as inadequate for the humanities and replaces them by a survey conducted among highly distinguished scholars who were asked to name the most prestigious journals in their respective fields. The result is a list grouped into three categories: 'INT1', 'INT2'

and 'NAT'. This order indicates the (descending) importance of the journals in the respective category. Again, quite resolutely, the list is meant to be no ranking: '[Question:] Is ERIH a ranking system? [Answer:] ERIH is not a billiometric [*sic*] tool or a reanking [*sic*] system. The aim of ERIH is to enhance the global visibility of high-quality research in the Humanities across all of Europe and to facilitate access to research journals published in all European languages; it is not to rank journals or articles' (European Science Foundation 2014). Compiled by only four to six European scholars per discipline, the list is not undisputedly acknowledged; as far as I know, it is not even widely known.

2.4 Rigor and Quotations

Garfield himself has always pointed out that the citation analysis of journals refers only to the *usage* of a published text; it does not say anything about approval or disapproval, nor does it assess the quality of a paper (Garfield 1979, p. 148). He then notices that the citation network allows its users to know what new developments emerge. It thus enables them to focus on prevalent trends. This idea can be put differently: High quotation rates and dense subnets show a strong *cohesion* of the group.

There may be two main reasons for the cohesion that becomes visible because of the quotation network. (1) First, it can derive from shared convictions about scientific rigor. Only publications that comply with the methodological demands of the respective discipline will have a chance to be cited. Regardless of the quality, originality and importance of the paper, cohesion makes the author belong to the specific group. Anecdotally, Kahneman reports that his success in Economics is due to only one improbable and lucky event: one of his articles being accepted in an important economic (rather than psychological) journal (Kahnemann 2011, p. 271). In this first case, cohesion warrants at least minimal standards of scientific procedure. (2) Then again, cohesion can simply result from a feeling of mutual affection and enthusiasm. In this second case, the cohesion comes first and stabilizes itself. It relies on the well-known in-group bias, i.e. the preference for one's own group. For example, members of pseudoscientific communities will cite one another (such as followers of homeopathy). If such a group is large enough, it will produce high quotation levels.

As a consequence, impressive quotation rates do not say what *kind* of agreement or conformity a respective group chooses as its foundation. It *can* be scientific rigor; but it can *also* be anything else. This conclusion is not new and not important for my argument. However, its reverse is. If a group shows low quotation levels, it necessarily lacks cohesion. It possesses neither clear standards of methodological rigor nor a feeling of community.

3 Low Quotation Frequencies in Literature

3.1 Materials and Methods

To analyse citation rates in Literature, I am going to use citation indices provided by commercial services. Among the available databases, only the *Scopus* database (run by Elsevier B.V.) covers a sufficient number of Literature journals to calculate journal rankings. Therefore, this database is my only resource. For its ranking, *Scopus* uses the indicator SJR2, which depicts not only the frequency of its articles being cited but also the prestige of each journal (Guerrero-Botea and Moya-Anegón 2012). Despite certain differences, this database is comparable to the Impact Factor. The indicator, however, will not play a major role in my argument; it will be used only to find journals that are supposed to be cited at an above-average rate.

As of 2012, the *ISI Web of Knowledge*, provided by Thomson Reuters, does not include any journals that belong to the 'hard-core' disciplines within the humanities. Although the *Web of Science*—also operated by Thomson Reuters and the company's main trademark which also includes the *ISI Web of Knowledge*—lists Literature journals, it does not provide any rankings or helpful statistics. Likewise, *Google Scholar*, run by Google Inc., does not allow any inferences from its data. Unlike its competitors (cf. Mikki 2009), *Google Scholar* browses all kinds of research publications (including books) and retrieves quotations by analyzing the raw text material. It thus covers books—this being an advantage over Elsevier and Thomson Reuters. However, *Google Scholar* is so unsystematic that the data contain artifacts and detect fewer quotations than *Google Scholar's* competitors (as of 2013).

My analysis focuses on two aspects. On the one hand I am interested in the absolute numbers of citations. They are the cause of the methodological difficulties in citation analysis; but, at the same time, they are an important fact that deserves attention of its own. On the other hand, I concentrate on the ratios of cited and uncited articles across different disciplines. For the sake of simplicity, I limit my analysis to Medicine. I choose to compare the aforementioned ratios (despite the problem of validity) because this is the only parameter that at least *can* be examined.

3.2 Results

Let us examine the citation analysis provided by *Scopus* for the subject category Literature and Literary Theory and the year 2012 (see Table 1). The absolute numbers of the top five most influential journals are strikingly low. The top journal, *Gema Online Journal of Language Studies*, which, by the way, I had never heard of before, does not appear in the ERIH ranking at all (Sect. 2.3). This journal is ranked first with regard to the SJR2 indicator implemented by *Scopus*. The strange phenomenon is easily explained: The journal focuses on linguistics; in the respective ranking ('Language and Linguistics'), it holds only position 82. Since it sometimes publishes

Table 1 The five highest ranking publications in the subject category Literature and Literary Theory in 2012 (citation data by Scopus)

Title	SJR	H-index	Total Docs.[a]	Total Docs.[b]	Total Refs.	Total Cites[b]	Citable Docs.[b]	Cites Doc.[c]	Ref. Doc.
Gema Online Journal of Language Studies	0.470	6	72	71	1,870	85	67	1.25	25.97
New Literary History	0.416	9	38	142	1,659	68	132	0.61	43.66
Shakespeare Quarterly	0.366	7	16	68	940	21	61	0.21	58.75
College Composition and Communication	0.353	12	30	160	1,006	59	138	0.42	33.53
Journal of Biblical Literature	0.343	8	38	143	2,762	34	141	0.22	72.68

Note SCImago Journal and Country Rank, JOURNAL RANKING: Subject Area: All, Subject Category: Literature and Literary Theory, Country: All, Year: 2012

[a]In 2012. [b]During 3 years. [c]During 2 years

articles in Literature, too, it is included in both lists; since the SJR2 indicator does not detect disciplinary boundaries, a comparatively mild impact in Language and Linguistics can make it the most prestigious journal in Literature and Literary Theory. Presumably, this effect must follow from the small numbers involved in quotations in Literature and Literary Theory so as to allow an interdisciplinary journal to move to the first position.

The second journal might be worth a closer look. *New Literary History* belongs to the highest ERIH category ('INT1'); personally, I would have guessed it might be among the top journals. This prestigious periodical, however, does not seem to be quoted very often, if one inspects the numbers provided by *Scopus* (see Table 2). For the 142 articles published between 2009 and 2011, only 68 citations were found. If one takes the small ratios between cited and uncited documents into account, viz. 26 % for this time window, the hypothesis seems acceptable that these few citations concentrate on few articles. The only undisputable inference is the mean citation frequency per article: We find two citations per article on average.

It is possible to compare these numbers to those of the most influential journal in Medicine (as ranked by the SJR2 indicator again), the *New England Journal of Medicine*. In the same time window (i.e. 2009–2012), we find 5,479 articles and 65,891 citations; on average, an article garnered 12 citations, and 46 % of these articles were cited within the time window.

As for the *New Literary History*, I discuss one of the journals that at least do receive some attention (in terms of citation analysis). Let us turn to *Poetica*, one of the most prestigious German journals. Within the ERIH ranking, *Poetica*, too, belongs to the highest category, 'INT1'. Yet it ranks only 313th in the *Scopus* list. The more detailed numbers are disconcerting (see Table 3). Between 2009 and 2011, the journal published altogether 48 articles, among which only three received at least one citation (within this time window). In the long run, the quotation ratio never exceeds 16 %; but the 6 %, which can be found in three columns (2006, 2007, 2012), is not an exception. More astonishingly, only four citations were found. This is to say that two articles garnered exactly one citation, and one article can be proud to have been cited twice.

The problems that I mention apply to all entries in the ranking. On the one hand, the absolute numbers are so low that small changes affect the position of journals; on the other hand, interdisciplinary journals automatically move up (this effect could be dubbed 'cross-listing buoyancy'). The ranking does not reflect the 'qualitative' assessment of the European Science Foundation. These figures have significance only as they show that quotations in Literature are rare.

3.3 Possible Objections

My approach may face three major objections. First, absolute numbers have limited value. They are not embedded in a statistical analysis, and, therefore, they cannot characterize the phenomenon in question. I will not deny the cogency of

Table 2 Development of citations between 2004 and 2012 for the high ranking international journal *New Literary History* (data by Scopus)

Indicators	2004	2005	2006	2007	2008	2009	2010	2011	2012
SJR	0.166	0.129	0.236	0.218	0.202	0.112	0.255	0.361	0.416
Total Docs.	39	38	49	41	43	55	47	40	38
Total Docs. (3 years)	76	115	117	126	128	133	139	145	142
Total references	1,359	946	1,360	1,572	1,506	1,642	1,857	1,386	1,659
Total Cites (3 years)	26	21	36	31	29	17	40	62	68
Citable Docs. (3 years)	74	110	110	114	116	120	127	134	132
Cites/Docs. (4 years)	0.35	0.19	0.32	0.30	0.22	0.20	0.31	0.46	0.48
Cites/Doc. (2 years)	0.35	0.15	0.33	0.31	0.24	0.19	0.19	0.57	0.61
References/Doc.	34.85	24.89	27.76	38.34	35.02	29.85	39.51	34.65	43.66
Cited Docs.	22	19	29	25	24	13	29	39	37
United Docs.	54	96	88	101	104	120	110	106	105
Ratio (cited/uncited docs) (%)	*41*	*20*	*33*	*25*	*23*	*11*	*26*	*37*	*35*

Note SCImago Journal and Country Rank, JOURNAL CLOSE-UP: *New Literary History*, Publisher: Johns Hopkins University Press, ISSN: 00286087, 1080661X. Italics indicate my own calculations

Table 3 Development of citations between 2004 and 2012 for the high ranking German language journal *Poetica* (data by Scopus)

Indicators	2004	2005	2006	2007	2008	2009	2010	2011	2012
SJR	0.101	0.101	0.100	0.151	0.154	0.123	0.111	0.111	0.100
Total Docs.	16	17	16	14	12	19	13	16	0
Total Docs. (3 years)	23	39	48	49	47	42	45	44	48
Total Cites (3 years)	3	2	3	4	5	5	5	8	4
Cites/Doc. (4 years)	0.14	0.05	0.07	0.11	0.11	0.10	0.13	0.19	0.07
Cites/Doc. (2 years)	0.14	0.07	0.00	0.06	0.13	0.15	0.13	0.16	0.03
References/Doc.	55.88	70.71	52.31	110.07	104.08	51.89	63.38	56.56	0.00
Cited Docs.	2	2	3	3	5	4	5	7	3
Uncited Docs.	21	37	45	46	42	38	40	37	45
Ratio (cited/uncited docs) (%)	*9*	*5*	*6*	*6*	*11*	*10*	*11*	*16*	*6*

Note SCImago Journal and Country Rank, JOURNAL CLOSE-UP: *Poetica*, Publisher: Wilhelm Fink Verlag, ISSN: 03034178. Italics indicate my own calculations

this objection. However, the point is that the low numbers themselves are the phenomenon to be explained. My analysis also comprises the comparison of relative quantities. By contrasting the ratios of uncited and cited papers across disciplines, I can increase the plausibility of my claims. I am confident that the synopsis of all data corroborates the hypothesis that literary scholars' quotation rates are altogether marginal.

The second possible objection concerns the available data about research in the humanities. Currently, the most widespread attempt to remedy the tiny absolute numbers is the inclusion of books. The idea is that the databases are deficient—not the citation culture (e.g. see Nederhof 2011, p. 128). The inclusion of monographs is Hammarfelt's (2012, p. 172) precept. In 2011, Thomson Reuters launched its *Book Citation Index* covering books submitted by editors from 2005 onward and continuously has worked on improving the *Book Citation Index* ever since. However, the inclusion of monographs will not provide an easy solution. There are three obstacles:

(1) *Primary versus secondary sources*. In the humanities, some books are *objects* of analysis, and some provide *supporting arguments*. In the first case, we speak of *primary*, in the latter case of *secondary sources*. In many contexts, the distinction between both types is blurry (see Hellqvist 2010, p. 316, for an excellent discussion).[2] Hammarfelt's (2012) most radiant example, Walter Benjamin's *Illuminationen*, which he states to have spread across disciplines (p. 167), is a compilation of essays from the 1920s and 1930s. The book is cited for very different reasons. The quotations in computer science and physics (Hammarfelt 2012, p. 167) will probably have an ornamental character; Benjamin is a very popular supplier of chic epigraphs. Within the humanities, Benjamin is one of the authors whose works are analysed rather than used, that is, he is a primary source. So are other authors whom (Hammarfelt 2012, p. 166) counts among the canonized: Aristotle, Roland Barthes, Jacques Derrida, etc. Even more, some of his canonized authors wrote just fiction (Ovid and James Joyce). Hence, these monographs must be primary sources.

An algorithm that distinguishes between primary and secondary sources is difficult to implement. The software has to discriminate between different kinds of arguments, which requires semantic analysis. As is well known, we are far away from any sensible linguistic analysis of texts without specific ontology (in the sense of semantics); so even the effort will be futile. The only reliable possibility would be a systematic distinction between primary and secondary sources in the bibliographies, a practice common in many scholarly publications, but far from ubiquitous. With this problem realized, it is difficult to implement an automatic analysis.

Recent publications, of course, can be counted as secondary sources per convention. This would be reasonable and useful, even if we know that the transition from 'secondary scholar' to 'primary author' is what scholars in the humanities dream of and what they admire (cf. Ochsner et al. 2013, pp. 83–85). Quite often this happens late, often after the scholar's death (and his reincarnation as 'author'), as was the case

[2]This is why Zuccala's (2012) similar—and barely novel—distinction between vocational and epistemic misses the point. This article tends to overlook many problems I discuss here.

with Benjamin, too, who was even refused a university position during his lifetime. The usage of recent publications remains only a possibility.

The inclusion of books would not change the whole picture. The absolute numbers would remain low. In a more or less systematic case analysis, Bauerlein (2011) shows that scholars do not cite books either (p. 12). Quite on the contrary, Bauerlein (himself a professor of English Literature, by the way) concludes that the production of books is an economic waste of resources and should be stopped. *Google Scholar* confirms that literary scholars quote but do so rarely. As stated above, the service includes books. Since Google has scanned and deciphered incredibly many books, including those from the past decade, for its service *Google Books* (prior to the service's restriction on account of massive copyright infringements), it has a pretty good overview of the names dropped in scholarly books. Nonetheless, Google's services show that books are quoted as rarely as articles (if not even less frequently). We thus count the documents cited. Scholars quote numerous sources; at least nothing indicates that lists of references are shorter in the humanities than they are in other disciplines. But all signs point at the possibility that only a few scholars can hope to be quoted by their peers. The fact remains that literary scholars quote each other but do so rarely.

(2) *Reading cycles*. Another remedy being discussed involves larger time windows. Literary scholars are supposed to have 'slower reading cycles', to stumble upon old articles and to unfold their impact much later than the original publication. Unfortunately, there is little evidence for this myth. Of course, there are many 'delayed' quotations in the humanities. But the problem is that they do not change the whole picture. In the vast majority of cases, their distribution is as Poisson-like as the 'instantaneous' quotations, and they are as rare. Again, the sparse data Google provides us with do not indicate any significant increase of citations caused by a need for long-lasting contemplation. Nor does Bauerlein find any hint of boosting the effects of prolonged intellectual incubation periods. Nederhof (1996) claims that in some humanist disciplines, the impact of articles reaches a peak in the third year; hence, the chosen citation window appears adequate and meaningful.

(3) *What quotations stand for*. The third obstacle is different in kind. Since the figures show small numbers, citations that do not refer to the content of the cited articles may distort the results of the statistical analysis to a significant extent. As recently demonstrated by Abbott (2011), a considerable percentage of citations does not relate in any conceivable way to the cited article, which could indicate that this article has never been actually read. Examples are easily at hand. In one of the top journals in Literature, *Poetics Today* ('INT1'), the Web of Science records two citations of an article of mine. Unfortunately, these citations come from scholars who use my article to introduce a notion established by Plato around 400 B.C. With two citations, my text belongs to the very small cohort of highly cited articles, but the actual quotations are disastrously inappropriate. This problem cannot be ruled out in other disciplines either. There is no clue whatsoever indicating that inappropriate quotations occur more often in the humanities than in other disciplines. Nonetheless, we have to consider the possibility that even the small numbers found in the figures

are not the result of attentive reading, but of the need to decorate an article with as many references as possible.

We eventually have to reconcile two apparently contradictory observations. On the one hand, scholars present us with long lists of references and are expected to quote as much as possible. On the other hand, each scholar can expect only little attention and very few (if any) citations by peers. This miracle can be easily resolved: Partly, scholars quote from other disciplines, partly, quotations cluster around certain few 'big names', who are quoted abundantly. There is no contradiction between long lists of references and few citations, that is, between many incidents of citing and only a few of being cited.

4 Discussion

As we have seen, the disciplinary culture of Literature requires scholars to quote one another extensively, but only few citations can be found. How can this be explained? Although I have expressed my doubts about the importance of coverage, first, more data must be obtained: Books must be extensively included in the analysis, and the citation windows must be enlarged, maybe up to decades. Such an improvement of the databases does not add to the bibliometric assessment of individual scholarly performance; instead, it adds to the understanding of the intellectual configuration of Literature and of other related fields in the humanities. Before we start understanding the criteria of excellence and develop a means of mapping qualitative judgments on quantities, we must first understand why citations occur so rarely.

Perhaps publications in Literature do not contain pieces of positive information that can be used to support one's own argument straightforwardly. Publications present the scholar with helpful or dubious opinions, useful theoretical perspectives, or noteworthy criticisms, but, possibly, a publication cannot be reduced to a simple single result. If this is the case, the question is which (societal) task Literature is committed to. If this is not case, the lack of quotations raises the question of why so many papers are written and published that do not attract any attention at all.

I can conceive of two explanations. (1) The first explanation concerns a possible 'archival function' of Literature (and related fields in the humanities). As Fohrmann (2013) recently put it, the disciplines may be responsible for the cultural archive (pp. 616, 617). Indeed, scholars count 'fostering cultural memory' among the most important factors that increase excellence in the humanities (Hug et al. 2013, pp. 373, 382). Teaching and writing in the humanities do aim to increase knowledge and to stabilize our cultural memory. As a consequence, seminars and scholarly publications are costly and ephemeral, but still are necessary byproducts of society's wish to uphold and to update its cultural heritage.

At first glance, this may sound sarcastic, but, in fact, this explanation would imply that the current situation might harm both the humanities and the university's sponsors (in Europe, these are mostly the governments and, therefore, the taxpayers). In the 1980s, the humanities had to choose whether they would adapt to the institutional

logic of the science departments, or to move out of the core of academia and to become cultural institutions, such as operas and museums. The humanities chose to remain at the heart of the university and thus accepted the slow adoption of mechanisms such as the competition for third-party funding and the numerical augmentation of publications. Now, the humanities produce texts that no one reads, that the taxpayer pays for and that distract the scholars from their core task: to foster the cultural archive, to immerse oneself in old books for months and years, to gain erudition and scholarship, and to promote the cultural heritage to young students and to society as a whole. (This is maybe why scholars are reluctant to cherish the scholars' impact on society, as Hug et al. (2013, pp. 373, 382) also show. In the scholars' view, their task is to expose the impact of the cultural heritage on society. In a way, giving too much room to the scholars seems to be a kind of vanity at the expense of the actual object of their duties.) Maybe, releasing the humanities from the evaluations and structures made for modern research disciplines would free the humanities from their bonds, reestablish their own self-confidence and decrease the costs their current embedding in the universities impose on the sponsors. It would be a mere question of labeling whether the remaining and hopefully prosperous institutions could still be called 'academic'.

(2) The second explanation, however, is less flattering. It could also turn out that low citation frequencies indicate the moribund nature of the affected disciplines. When I recall that citations and debates have been core practices in the humanities for centuries, another conclusion pushes itself to the foreground: Scholars in the affected fields feel bored when they have to read other scholars' publications.

In the 1980s and the early 1990s, there were fierce debates, and the questions at stake could be pinpointed (see Hüser 1992). Today, the very questions vanish; scholars have difficulties stating what they are curious about (Bunia 2011a). If no scholar experiences any intellectual stimulation instilled by a peer's publication, she will tend to read less, to turn her attention to other fields and to quote marginally. With regard to cohesion (see Sect. 2.4), such a situation would also imply that the scholars in the affected fields no longer form a community that would identify itself as cohesive; one no longer feels responsible for the other and for the discipline's future. If all debates have ended, the vanishing quotations simply indicate a natural death that no one has to worry about.

Both explanations will easily provoke contestations. As for the first one, one would have to ask why scholars have never realized that they had been cast in the wrong movie. As for the second one, there are only few hints at a considerable change in the past 20 years. Did scholars cite each other more fervently in the 1970s and 1980s than today? I do not know. Therefore, we need more research on the scholars' work. For instance, we need to know why they read their peers' work and if they enjoy it. It is good that researchers, namely, Hug, Ochsner and Daniel, began asking scholars about their criteria to understand how the scholars evaluated their peers' performance. But we also have to take into account the deep unsettledness reigning in Literature and related fields (see Scholes 2011; see again Bauerlein 2011; Bunia 2011b; Lamont 2009; Wiemer 2011). We have to thoroughly discuss a 'criterion', e.g. 'rigor', which is a virtue scholars expect from others (Hug et al. 2013,

pp. 373, 382). But 'rigor' is characterized by 'clear language', 'reflection of method', 'clear structure' and 'stringent argumentation', which are virtues the humanities are not widely acclaimed for and are qualities that may be assessed differently by different scholars. In brief, these self-reported criteria have to be compared to the actual practice. It may be confirmed that a criterion such as rigor is being consistently applied to new works; but it may equally well turn out that the criterion is a passe-partout that conceals a lack of intellectual cohesion in the field. Again, this means that we first must understand what the humanities actually do before we start evaluating the outcome of their efforts by quantitative means.

References

Abbott, A. (2011). Varianten der Unwissenheit. *Nach Feierabend*, *6*, 15–33. doi:10.4472/9783037342671.0002.

Bauerlein, M. (2011). Literary research: Cost and impact. Washington, DC: Center for College Affordability and Productivity. Retrieved from http://centerforcollegeaffordability.org/uploads/Literary_Research_Bauerlein.pdf.

Bohannon, J. (2013). Who's afraid of peer review? *Science*, *342*(6154), 60–65. doi:10.1126/science.342.6154.60.

Bornmann, L. (2013). Evaluations by peer review in science. *Springer Science Reviews*, *1*(1–2), 1–4. doi:10.1007/s40362-012-0002-3.

Bunia, R. (2011a). Das Handwerk in der Theoriebildung: Zu Hermeneutik und Philologie. *Journal of Literary Theory*, *5*(2), 149–162. doi:10.1515/JLT.014.

Bunia, R. (2011b). Die Ökonomien wissenschaftlichen Zitierens: Das Verbot von Plagiaten als Indiz für eine gestörte Balance in der Wissenschaft. In T. Rommel (Ed.), *Plagiate - Gefahr für die Wissenschaft?* (pp. 17–31). Münster: LIT.

Bunia, R. (2013). Zitieren. In U. Frietsch, & J. Rogge (Eds.), Über die Praxis des kulturwissenschaftlichen Arbeitens (pp. 485–489). Bielefeld: transcript.

Bunia, R., & Dembeck, T. (2010). Freunde zitieren: Das Problem wissenschaftlicher Verbindlichkeit. In G. Stanitzek & N. Binczek (Eds.), *Strong ties/weak ties* (pp. 161–195). Heidelberg: Winter.

European Science Foundation. (2014). European Reference Index for the Humanities (ERIH). Frequently Asked Questions. Retrieved from http://web.archive.org/web/20131203155642/http://www.esf.org/hosting-experts/scientific-review-groups/humanities-hum/erih-european-reference-index-for-the-humanities/frequently-asked-questions.html.

Fohrmann, J. (2013). Weltgesellschaft und Nationalphilologie. *Merkur*, *770*, 607–618.

Garfield, E. (1979). *Citation indexing. Its theory and application in science, technology, and humanities*. New York, NY: Wiley.

Guerrero-Botea, V. P., & Moya-Anegón, F. (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of Informetrics*, *6*(4), 674–688. doi:10.1016/j.joi.2012.07.001.

Hammarfelt, B. (2012). *Following the footnotes: A bibliometric analysis of citation patterns in literary studies*. (Doctoral dissertation. Skrifter utgivna vid institutionen för ABM vid Uppsala Universitet, Vol. 5). Uppsala: Uppsala Universitet. Retrieved from http://www.diva-portal.org/smash/get/diva2:511996/FULLTEXT01.pdf.

Hellqvist, B. (2010). Referencing in the humanities and its implications for citation analysis. *Journal of the American Society for Information Science and Technology*, *61*(2), 310–318. doi:10.1002/asi.21256.

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.

Hüser, R. (1992). *Kommissar Lohmann*. Bielefeld: typescript.

Kahnemann, D. (2011). *Thinking, fast and slow*. London: Allan Lane.

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

Luhmann, N. (1990 [Reprint 1998]). *Die Wissenschaft der Gesellschaft*. Frankfurt a. M.: Suhrkamp.

MacDonnald, S. P. (1994). *Professional academic writing in the humanities and social sciences*. Carbondale, IL: Southern Illinois University Press.

Mikki, S. (2009). Google Scholar compared to Web of Science. *Nordic Journal of Information Literacy in Higher Education*, *1*(1), 41–51. doi:10.15845/noril.v1i1.10.

Nederhof, A. J. (1996). A bibliometric assessment of research council grants in linguistics. *Research Evaluation*, *6*(1), 2–12. doi:10.1093/rev/6.1.2.

Nederhof, A. J. (2011). A bibliometric study of productivity and impact of modern language and literature research. *Research Evaluation*, *20*(2), 117–129. doi:10.3152/095820211X12941371876508.

Nimis, S. (1984). Fussnoten: Das Fundament der Wissenschaft. *Arethusa*, *17*, 105–134.

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*(4), 79–92. doi:10.1093/reseval/rvs039.

Rennie, D. (2002). Fourth international congress on peer review in biomedical publication. *The Journal of the American Medical Association*, *287*(21), 2759–2760. doi:10.1001/jama.287.21.2759.

Schmalzriedt, E. (1970). *Peri Physeos: Zur Frühgeschichte der Buchtitel*. München: Fink.

Scholes, R. (2011). *English after the fall: From literature to textuality*. Iowa City, IA: University of Iowa Press.

Squazzoni, F., & Gandelli, C. (2012). Saint Matthews strikes again. An agent-based model of peer review in the scientific community structures. *Journal of Informetrics*, *6*(2), 265–275. doi:10.1016/j.joi.2011.12.005.

Wiemer, T. (2011). Ideen messen, Lektüren verwalten? Über Qualitätskriterien literaturwissenschaftlicher Forschung. *Journal of Literary Theory*, *5*(2), 263–278. doi:10.1515/JLT.2011.024.

Zuccala, A. (2012). Quality and influence in literary work: Evaluating the 'educated imagination'. *Research Evaluation*, *21*(3), 229–241. doi:10.1093/reseval/rvs017.

Part IV
Evaluation of Research in the Humanities in Practice

Peer Review in the Social Sciences and Humanities at the European Level: The Experiences of the European Research Council

Thomas König

Abstract In this article, I outline the evaluation process established by the European Research Council (ERC) and present results of the ERC's funding calls between 2007 and 2012. Because of its European added value, the ERC is a unique funding organization in the European research landscape. Based on a rigorous evaluation process, the ERC dedicates a considerable share of its budget to the social sciences and humanities.

1 The European Research Council's Mission

The European Research Council (ERC) was established in 2007 as part of the European Commission's 7th Framework Programme (namely, the 'Ideas' Specific Programme); under the new framework program, Horizon 2020, it has been extended until 2020. Since inception, the ERC has filled a gap in the European funding landscape. The council's principle is to make decisions on the criterion of 'excellence only'. Although RD&I funding has become a major policy issue of European integration during the last 20 years, cutting-edge basic research remained largely underdeveloped at the European level (Dosi et al. 2009, pp. 233, 234). There are several reasons for this delay. One is the initial mandate to the European Commission to fund research under framework programs to the extent it supports the competitiveness of European industry. Consensus on the need to fund frontier research at the European level was not reached until the negotiations for FP7.

In the initial reasoning for setting up the ERC, frontier research was perceived as the (necessary) counterpart to a top-down approach in research funding, because frontier research is an investment in the European knowledge base and the innovation cycle (Schibany and Gassler 2010). Equally important, however, the ERC makes genuine competition among research institutions and researchers at the European level possible for the first time. The previous framework programs (FPs) lacked a specific

T. König (✉)
IHS, Josefstädter Straße 39, 1080 Vienna, Austria
e-mail: thomas.koenig@ihs.ac.at

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_12

drive to integration (Banchoff 2002). It turns out that, with the bottom-up approach and simple funding instruments, the ERC contributes significantly to a 'European added value' (Andrèe 2009; Stampfer 2008). Under the FP7 framework, the ERC received 15 % of the entire budget dedicated to research funding, totaling EUR 7.5 billion over 7 years, which makes the ERC a powerful instrument for funding research at the frontier of knowledge. Together with well-established national research funding organizations in European countries (although endowed with unequal budgets), the ERC now contributes decisively to fostering the European Research Area, the backbone of the European knowledge society. Under Horizon 2020, the ERC's budget will increase considerably, to approximately EUR 13.1 billion.

1.1 How Does the ERC Work?

The governing body of the ERC is the Scientific Council, which is responsible for developing the ERC's strategy. The Scientific Council represents the ERC to the scientific community, establishes the annual Work Program and in general ensures the ERC's high profile. The Scientific Council is composed of 22 highly distinguished members of the European scientific community, acting in a personal capacity. The governing structure of the ERC will change under the new legislation of Horizon 2020 (Nowotny 2013); however, the main principle will remain the same: Committed only to the principle of scientific excellence, the Scientific Council members are independent from political, economic, or other interests. To administratively support the Scientific Council, the Executive Agency (ERCEA) was created in 2009. Located in Brussels, the ERCEA currently has a staff of approximately 380, and the number is rising.

Exclusively committed to funding curiosity-driven, bottom-up frontier research by individual principal investigators (PIs) in EU member states or associated countries host institutions, the ERC is open to applications from all fields and to researchers from all over the world. At the moment, three funding mechanisms have been established. For talented post-docs and early-stage researchers (between 2 and 7 years after PhD), the Starting Grant scheme offers funding for 5 years and a project budget of up to EUR 1.5 million. The Consolidator Grant scheme, implemented since 2013, is a breakout from the Starting Grant call; this scheme covers the subsequent scientific career steps for more advanced scientists (seven to 12 years past PhD). Finally, well-established, senior researchers can apply under the Advanced Grant scheme, which offers funding for 5 years and a project budget of up to EUR 2.5 million. Advanced Grant applicants must have a distinguished track record over the past 10 years and present an innovative, ambitious research project. In 2012, the Scientific Council implemented a fourth grant programme for research groups, called the Synergy Grant. In addition, the Proof of Concept Scheme provides an opportunity for current ERC grantees to receive top-up funding for commercializing their research results. Each grant call is usually published annually.

Projects are funded based on proposals presented by individual researchers on subjects of their choice, with a clear emphasis on interdisciplinary and high-risk projects. Proposals are evaluated on the sole criterion of scientific excellence. Since there are no thematic or other priorities preselecting among the ideas and projects that applicants wish to pursue, evaluation of the project proposals relies heavily on the expertise of the reviewers. The ERC evaluation process is carried out by 25 panels for each funding mechanism with alternate panels put in place every other year—adding up to 75 panels annually (not including the extra panels in the Synergy Grant, which follows a different evaluation procedure). Each panel consists of approximately 12 to 16 panel members, all international experts in their field. They are supported by approximately 1,600 external (remote) reviewers per call.

1.2 European Added Value

Within a very short period, the ERC has become an undisputed success story. With its simple funding instruments, the ERC responds to the expectations of the younger generation of researchers who seek to break out of academic hierarchies and their national systems to obtain early scientific independence. And the ERC encourages advanced researchers to pursue riskier ideas that might lead to new breakthroughs and discoveries. However, beyond providing trustworthy and fair funding opportunities for the European scientific community exclusively based on scientific merit, the ERC carries European 'added value' (Nedeva and Stampfer 2012).

This 'added value' can be demonstrated on two levels. The first is related to the evaluation process. The ERC's evaluation process has won such high acclaim and reputation that high-level experts are willing to participate in the lengthy evaluation process, knowing that the ERC upholds its promise of the highest professionalism and, at the same time, allows them to witness the newest developments in their field. One of the most significant results of the ERC is the completely international set-up of its evaluation panels. On average, no more than two experts from the same country are represented on one panel, and on average, seven to ten countries are represented on one panel. Thus, the ERC has the most international evaluation procedure in place. At the same time, the panels are an excellent breeding ground for establishing a truly European academic culture that profits from the diverse cultural background of members, but is nevertheless focused on intrinsically scientific values.

The second level is related to the stimulation ERC grants provide to research institutions in Europe. It is based on a quite simple but nevertheless very effective equation: Countries and host institutions (universities and other research centres) can compare how many ERC grants they have won. With ERC grants distributed all over Europe, we start to see certain patterns. In terms of absolute numbers, related to the size of the population, the biggest winners of ERC grants thus far have been the United Kingdom, Switzerland and Israel. Comparisons like this that make policy makers and scientists demand more efficient infrastructure and support, in order to

achieve better results in the ERC grant competition. By and large, the ERC has become a quality threshold for the European research community.

The success story of the ERC has been critically acclaimed in evaluations (Vike-Freiberga et al. 2009; Annerberg et al. 2010, pp. 34–37) and public statements. As a role model for institution building, the ERC has already raised the interest of independent researchers (Gross et al. 2010; Hummer 2007; Nedeva 2009) and students (Haller 2010; Tan 2010). Members of the Scientific Council, when presenting the ERC to the academic community, continuously stress that the ERC is a learning institution and that improvements, particularly regarding the governance structure and the long-term funding of the ERC, are still needed (Antonoyiannakis and Kafatos 2009; Fricker 2009; Gilbert 2010; Nowotny 2010, 2013; Winnacker 2008).

2 Why Social Sciences and Humanities?

It goes without saying that the panels and reviewers follow the highest standards of peer review, as established and monitored by the ERC. The 25 panels are divided into three domains: physics and engineering (PE), life sciences (LS) and social sciences and humanities (SH). According to an interview with Helga Nowotny, ERC president from 2010 to 2013, the ERC was initially planned to cover only life sciences and physics, and it took some effort to convince politicians and representatives of the 'hard sciences' that social sciences and humanities must be included. Now the ERC's agenda is clear, as Nowotny, a sociologist by training, emphasizes: 'We fund research in the 19th century, German conception of *Wissenschaft*, which includes everything' (Enserink 2011, p. 1135).

Under FP7, the share of social sciences and humanities in the ERC's overall budget of EUR 7.51 billion was approximately 17 %. This was a much higher share than any other programme dedicated to social sciences and humanities. For example, in the 'capacities' special program, the socio-economic sciences and the humanities accounted for only 2 %. What is interesting, however, is that the social sciences and humanities were slower in recognizing the ERC as a source of funding. After a weak start in the first calls in 2007 and 2008, the number of applications rose more sharply in the SH domain than in the other domains. And, as we shall see, in the SH domain the popularity of the ERC still differs remarkably between disciplines and fields.

2.1 An Inclusive Approach

We live in a time when 'innovation' has almost gained the status of a buzzword in the European political discourse. Public spending for research is often evaluated along the (promised) impact on economic development. However, there is more to innovation. Whether it is a result of the financial crisis that asks for a critical validation of our understanding of capitalism, or the general question how to support societies

abroad, struggling to find a just and democratic society: Every time questions on societal and cultural foundations arise, in-depth analysis and expertise are required from the social sciences and humanities.

Unfortunately, the very disciplines and fields usually subsumed under the label of social sciences and humanities, thus far, cannot take advantage of this. An analysis of previous efforts by the European Commission showed that, although these programs were received very well by the community, the influence on 'the strategies and practices [...] has been limited' (Watson et al. 2010, p. 17). Whether the ERC's inclusive approach will have a more stimulating effect on elevating social sciences and humanities on the European level in the future remains to be seen. But it deserves our close attention here to clarify what lies behind the inclusive meaning of *Wissenschaft*. Clearly, in the sense of spanning all scientific fields, it avoids the danger of limiting the success of new approaches and the possibility of projects not being fundable because of a lack of expertise. Since the ERC actively encourages scientists to reach beyond disciplinary borders and to implement interdisciplinarity as a fundamental principle in European research, the number of cross-panel and cross-domain projects is increasing.

The ERC funds not merely basic research but also *frontier research*. This distinction is crucial for the role of the social sciences and humanities in the ERC, and therefore needs more explanation. According to a now famous classification, research can be divided along two different motivating factors: the role of applications and the use and the depth of understanding of causes, phenomena and behaviour. From the four possible combinations, frontier research can be understood as that 'of applications-oriented research with the pursuit of fundamental understanding' (Whitley 2000, p. xxi). This kind of research is often also represented by the reference to Louis Pasteur (Stokes 1997), but it drives not only parts of the 'hard sciences' as genetics, for example. Indeed, as has been noted, this motivating combination can be 'found in most of the human sciences' (Whitley 2000, p. xxi), because these fields of knowledge are concerned with societal and human affairs. Thus, the social sciences and humanities are particularly well suited for the type of research that the ERC aims to fund.

Social sciences and humanities have always played a distinctive role in the European Commission's research programs (Kastrinos 2010, pp. 300–304). Nevertheless, due to the austerity principles established in the aftermath of the financial crisis, concerns have been growing over the past few years that the social sciences and humanities programs will be severely cut in the European Commission's next multi-annual funding program, Horizon 2020. On December 8, 2010, social scientists published a memorandum warning of 'alarming developments' (Risse et al. 2010). Since then, the debate on the role of social sciences and humanities in Horizon 2020 has taken many turns, and dominated the EU Presidency Conference in Vilnius in September 2013 (Mayer et al. 2014).

That there is a widespread feeling of threats to funding for social sciences and humanities within the communities is not so much because politicians disregard these fields, as the common belief goes. Instead, it is a consequence of the fact that the social sciences and humanities have only weak institutional forms of advocating on

the European level. For example, there is no equivalent to the well-organized and powerful European Molecular Biology Organization (EMBO) that participates in many important events and represents the interests of its field in many respects.

For the social sciences and humanities, this lack of representation has its reasons. Most research funding in these fields still comes from national sources, and it is on this level for which knowledge is produced and on which representation is focused. In an integrated Europe with new funding opportunities, however, orientation along national aspects becomes detrimental. To compensate the lack of institutional representation, members of disciplines and fields in the social sciences and humanities therefore often resort to an alarmist rhetoric. Since the ERC will continue to follow its inclusive approach, the council is becoming an important point of reference for the social sciences and humanities.

2.2 ERC Evaluation in the Social Sciences and Humanities

Based on an excellence-only approach, the ERC evaluation follows a well-established, rigid process. Two aspects are particularly important:

(A) The process is the same over all three domains. There is no special treatment for any discipline or research field regarding the evaluation process, simply because of two reasons. Cross-panel proposals are distributed to members of other panels; in order to incorporate these evaluations, the procedure must be consistent. Additionally, the Scientific Council believes that proposals from all fields can be assessed under the same premise, namely, excellence. Of course, there are huge differences in what excellence means in different disciplines, fields and paradigms. However, there can be no doubt that excellence exists in each case, and that the focus on excellence as the only criterion for selection helps to foster the intrinsic values of *Wissenschaft* across all domains.

(B) The ERC focuses on individual, bottom-up research projects with one PI. Since the proposal and the PI's track record are crucial for the success of the funded project, they are thoroughly assessed by multidisciplinary panels. This approach distinguishes between the originality of the proposal and the PI's capability to actually carry out the proposal.

What makes the ERC so special in Europe is not that the council funds research based on this notion of excellence, nor that the ERC relies on a rigid peer review system. This is nothing new, since the most prominent funding organization, the U.S. National Science Foundation, was founded in 1950. Other organizations in industrialized countries either followed this model or set up variants. All over Europe, funding organizations rely on decision-making procedures similar to those described by the European Science Foundation (2011). In many respects, therefore, the ERC is simply absorbing well-established procedures and patterns, particularly in the evaluation process. Nevertheless, within this reliable structure, the ERC has also developed remarkable new features. The most important aspect is the fruitful combination of

the internationality of the ERC peer review process with the rigid process put in place. This combination creates a diversified approach to excellence.

The proposal evaluation follows a two-step procedure. In the first step, after proposals have been submitted and eligibility has been checked, panel members evaluate the proposals and the track record of the grant applicants. These are the only two criteria for the evaluation process. An original project proposal and an excellent career path are required to reach the second step of the evaluation. In preparation for the second step, the applicant's proposal and CV are again evaluated, this time not only by at least three panel members assigned to the proposal but also by remote (external) reviewers, specifically from the research field of the proposal. This is also a very important undertaking with cross-panel and cross-domain proposals. In the case of such proposals, a streaming takes place, using appropriate experts from other panels. Thus, the ideal mix of expertise can be achieved, also with an interdisciplinary proposal.

The second step of the evaluation process is different in the Starting and Consolidator Grant schemes and the Advanced Grant scheme. In the latter, where it is assumed that the PI has already gained a recognizable position in his/her field, the final funding decision is based on a second, thorough assessment of the proposals that made it into step two. In the Starting and Consolidator Grant schemes, where young researchers competing for large sums, the panels are required to get a better impression of the PI. Thus, every Starting and Consolidator Grant applicant who made it to step two is invited to an interview with the panel. The interview serves two purposes: It shows whether the PI is really committed to his/her research proposal and if he/she is really capable of doing it. At the same time, the interview gives the PI the opportunity to engage in a discussion with the panel in order to convince its members of the PI's intellectual strength and his/her commitment to the proposed research.

Peer review is a well-established procedure. When assessing the intrinsic scientific value of a research project proposal, peer reviewing is the only valid selection procedure. Nevertheless, peer review has its flaws, particularly in terms of the novelty of approaches, concepts and methodologies. If panels decide according to conventional wisdom and are not prepared to choose risky but promising research projects, the panels fail to achieve the ERC's main target. In the case of social sciences and humanities, a particularly broad range of different conceptual approaches exists. Lamont (2009, p. 57) distinguishes different types of epistemological styles (constructivist, comprehensive, positivist, utilitarian), and all panels must respect each style as scholarly valuable.

There are several ways on which the ERC relies in order to achieve a fair evaluation procedure focused on excellence, and all are centered on the evaluation panels. To begin, the ERC Scientific Council sets up the panels in a broad, interdisciplinary way. Only 25 panels cover all fields of science, scholarship and engineering. Let's take a closer look at the six panels that are assembled under the two letters SH. Fields and disciplines range from economics and management (SH1), sociology, anthropology, political science, law (SH2), geography, demography, migration, environmental and

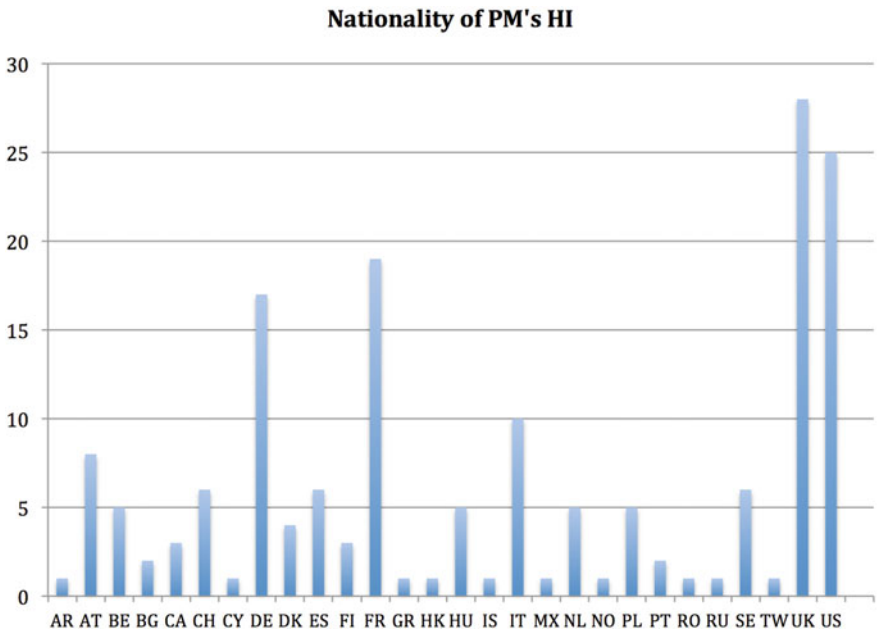


Fig. 1 Nationality of panel members' host institution

urban studies (SH3), linguistics, philosophy, education, psychology (SH4), literature and philology, art history, musicology (SH5) to history and archaeology (SH6).

Panel members are selected based on their scientific reputation; usually they have specialist as well as generalist competence, since they have to be open to multidisciplinary research perspectives. Diversity is not, as some may expect, a contradiction to excellence. In the case of the ERC, a diversified panel is considered a strength in the evaluation process. To take but one example, the approximately 170 panel members for the 12 SH evaluation panels in 2011 were situated at host institutions in 28 different countries worldwide (see Fig. 1). Experts from Anglo-American countries (the United Kingdom and the US) made up about 30 % of the total, thus presenting the largest group. Other large academic communities, such as the Germanic and the Francophone, constituted about 15 % and 11 %, respectively, of the total.[1]

The ERC Scientific Council, responsible for selecting and nominating panel members, has committed to a gender equality plan (ERC 2011), aiming at representation of female panelists of about 40 %. In the 2011 SH panels, this target was almost met; approximately 37 % of the experts on the six panels were female. Finally, panel members are advised to look for unconventional career paths and take them into consideration during decision-making. If we take the rising reputation of ERC grants and the huge acceptance that the ERC receives from the European academic community, this mix of strategies seems to be successful.

[1]The panel composition may change slightly during the course of an evaluation circle.

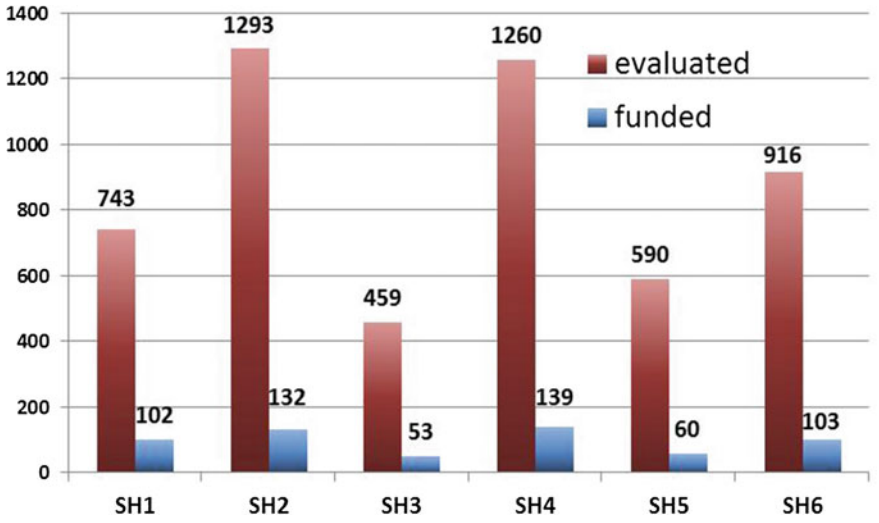


Fig. 2 Applications and granted projects submitted per panel, 2008–2012

3 What Are the Results?

Although the goal of this volume is the humanities (*Geisteswissenschaften*), distinguishing between social sciences and humanities does not make sense in the case of the ERC. Actually, there is only one domain (SH) in which the approaches are combined and intertwined.

If we look at the accumulated results from all 10 ERC calls for individual PIs from 2007 to 2012, there are interesting patterns in the SH-related project proposals.[2]

The success rate of the proposals submitted to the SH panels in the ERC is on average the same as in the two other domains. SH-related project proposals constitute about 17 % of the ERC budget spent on proposals submitted in these calls—or 600 projects in total.[3] The number of applications is rising more sharply in the SH domain.[4] Maybe even more significant, the number of applications to the panels is quite uneven. Thus, we can assume that certain fields (such as the social sciences in SH2 and the cognitive sciences in SH4) are more responsive to the ERC than others (such as the core humanities panel, SH5) (see Fig. 2, also the next paragraph).

[2]Data from the ERC Executive Agency website, http://erc.europa.eu. In 2007, only the Starting Grant call was announced; in 2008, only the Advanced Grant call. From 2009 onwards, both funding streams were carried out annually. When this contribution was being completed, data on these calls carried the most accurate information. The overall trend described in the following paragraphs did not change with the results of the three calls in 2013.

[3]This does not necessarily include so-called cross-disciplinary proposals, which were regarded as a separate 'fourth domain' in the earliest ERC calls.

[4]The initial ERC funding call, the Starting Grant Call of 2007, is not included here for two reasons: With a success rate of only 2 %, it was heavily over-subscribed, and the panel structure was different.

Since the budget of one call for each domain is distributed to the panels along the number of applications that each panel initially received, this difference also determines the number of fundable projects per panel. Thus, this results in a striking variation in how many projects are funded by each panel. Since the panels SH3 and SH5 receive few submissions, only 53 and 60 projects, respectively, were funded during the nine calls. On the other side, SH2 and SH4 are large panels in terms of submissions, and funded 132 and 139 projects, respectively. The SH1 and SH6 panels received fewer applications, but since the project budgets for these panels were on average smaller, approximately the same number of projects was funded as in the largest panels.

If we examine the country distribution of the submitted and granted SH proposals in all 10 calls, we see that the submitted proposals and granted projects are evenly distributed throughout Europe. The largest number of applications came from the UK (1,343), followed by Italy (878), Netherlands (590), Germany (577), Spain (474) and France (422). If we look at the grants funded, British host institutions lead the field with 208, followed by Dutch (79) and French institutions (68), German (57), Italian (52) and Spanish (37).[5]

4 Outlook

We know that the way research funding is set up affects the way research is carried out in the social sciences and humanities (Marton 2005, p. 184). Not even 10 years after the ERC's inception, the question if the ERC has already shaped the way research in the social sciences and humanities is carried out remains unanswered. We can assume, however, that the ERC has had an impact on two levels (Nowotny 2009, p. 3). First, particularly young grantees achieve early independence that, thus far, is widely unknown in the European university and research systems. Since the dependency of young researchers always had a particularly crippling impact on the social sciences and humanities, we may expect new, unconventional and highly innovative knowledge from Starting and Consolidator Grantees within the next few years.

Second, these young researchers may develop a new form of non-hierarchical collaboration from which the entire range of disciplines may profit. As a result, we can assume that there is a new visibility on social sciences and humanities, since more than ever they are working on transnational, comparative topics.

Given the ERC's budget in relation to the sums spent in other programs, the ERC is still a small player. Its reputation stems from its rigid evaluation process, its strict focus on excellence and its broad, pan-European approach. For the social sciences and humanities, the ERC offers a great opportunity to strengthen frontier research in an almost unprecedented manner. Nevertheless, some issues remain critical. One

[5]Because an ERC-funded project is portable and can be shifted to a host institution in another country, we cannot calculate a success rate per country of host institutions with the data available.

of the general problems the ERC has to deal with is the gender quota, particularly in the Advanced Grant scheme. The ERC Scientific Council therefore adopted the Gender Equality Plan (ERC 2011), and commissioned a study dedicated to gender and excellence in relation to ERC-funded projects.

Even more troubling to some is the participation of certain countries, and the looming fear that these countries may not be integrated in the emerging European Research Area. Certainly, there is a need to foster independent research in these countries. The ERC cannot deviate from its core mission, namely, focus on excellence; the ERC must support research facilities and infrastructure in these countries to create an environment such that researchers at these sites become competitive.

In SH in particular, another concern is the balance of panel member composition. In some respect, the SH panels represent the strong. There are more experts from different countries, but the difficulty here is the language. In the humanities, excellent researchers sometimes do not publish in English, and therefore remain 'invisible' as potential reviewers. Although the diversity of experts regarding country distribution is actually quite good, more experts should be invited from countries with such well-established traditions in the humanities.

In some fields, the ERC has witnessed a steady growth of applications, while in others, the number of applications is stagnant. This often goes hand in hand with the misunderstanding that projects primarily concerned with classificatory research are submitted. Undoubtedly, this is an important field of research; however, it is not within the ERC's funding policy, and therefore, projects with this background will be turned down. It seems that, particularly in the humanities (*Geisteswissenschaften*), communication of what the ERC can do for these disciplines and fields must be strengthened.

To a large extent, the ERC's high reputation among scholars and scientists comes from the fact that the evaluation process is admired and trusted by the research community. In this regard, again, diversity is crucial, because understanding excellence in a multi-dimensional way is a necessary prerequisite for research proposals from different fields and academic cultures. This understanding is already growing among the evaluation panels; one of the most fascinating aspects of the ERC is that it has created, perhaps for the first time in history, a truly transnational, that is, European, evaluation culture. In this setting, 'excellence' is understood not as exclusive but open to the unexpected.

The ERC involves reviewers from the entire world. Between 2007 and 2013, more than 4,000 distinguished scientists have reviewed more than 40,000 ERC applications. The panels and remote reviewers constitute the most precious asset of the ERC. The ERC has also contributed to raising the evaluation standards among national funding organizations throughout Europe and facilitates best practice by demonstrating a model of an exclusively merit-based evaluation culture, in particular for countries that, for historical reasons, lack such a culture.

References

Andrèe, D. (2009). Priority-setting in the European Research Framework (Vinnova Analysis 17). Retrieved from http://www.vinnova.se/upload/EPiStorePDF/va-09-17.pdf.

Annerberg, R., Begg, I., Acheson, H., Borrás, S., Hallén, A., Maimets, T., & Ylihonko, K. (2010). Interim evaluation of the Seventh Framework Programme. Report of the expert group. Retrieved from http://ec.europa.eu/research/evaluations/pdf/archive/other_reports_studies_and_documents/fp7_interim_evaluation_expert_group_report.pdf.

Antonoyiannakis, M., & Kafatos, F. C. (2009). The European Research Council: A revolutionary excellence initiative for Europe. *European Review*, *17*(3–4), 511–516. doi:10.1017/S1062798709000921.

Banchoff, T. (2002). Institutions, inertia, and European Union research policy. *Journal of Common Market Studies*, *40*(1), 1–21. doi:10.1111/1468-5965.00341.

Dosi, G., Llerena, P., & Labini, M. S. (2009). Does the European Paradox still hold? Did it ever? In H. Delange, U. Muldur, & L. Soete (Eds.), European science and technology policy. Towards integration or fragmentation? (pp. 214–236). Cheltenham: Edward Elgar.

Enserink, M. (2011). Keeping Europe's basic research agency on track. *Science*, *331*(6021), 1134–1135. doi:10.1126/science.331.6021.1134.

ERC. (2011). ERC Scientific Council gender equality plan 2007–2013. Retrieved from http://erc.europa.eu/sites/default/files/document/file/erc_scc_gender_equality_plan_2007_2013.pdf.

European Science Foundation. (2011). European peer review guide. Integrating policies and practices into coherent procedures. European Science Foundation. Strasbourg. Retrieved from http://www.esf.org/fileadmin/Public_documents/Publications/European_Peer_Review_Guide_01.pdf.

Fricker, J. (2009). Putting Europe on the scientific map. *Molecular Oncology*, *3*, 386–388. doi:10.1016/j.molonc.2009.06.005.

Gilbert, N. (2010). The labours of Fotis Kafatos. *Nature*, *464*, 20. doi:10.1038/464020a.

Gross, T., Karaalp, R. N., & Wilden, A. (2010). *Regelungsstrukturen der Forschungsförderung. Staatliche Projektfinanzierung mittels Peer-Review in Deutschland, Frankreich und der EU*. Baden-Baden: Nomos.

Haller, M. (2010). *The European Research Council. An agency governed by scientists?* (Unpublished master's thesis). College of Europe, Brugge.

Hummer, W. (2007). The European Institute of Technology and the European Research Council: Two instruments of European Excellence. *Integration*, *30*(22), 150–165.

Kastrinos, N. (2010). Policies for co-ordination in the European research area: A view from the social sciences and humanities. *Science and Public Policy*, *37*(4), 297–310. doi:10.3152/030234210X496646.

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

Marton, S. (2005). Academics and the mode-2 society: Shifts in knowledge production in the humanities and social sciences. In I. Bleiklie, & M. Henkel (Eds.), *Governing knowledge. A study of continuity and change in higer education* (pp. 169–188). Amsterdam: Springer.

Mayer, K., König, T., & Nowotny, H. (2014). *Horizons for social sciences and humanities. Conference report*. Vilnius: Mykolas Romeris University Press.

Nedeva, M. (2009). Conceiving institutions: the establishment of the European Research Council. In *Proceedings of the Atlanta Conference on Science and Innovation Policy 2009*. Retrieved from http://hdl.handle.net/1853/32364.

Nedeva, M., & Stampfer, M. (2012). From 'Science in Europe' to 'European Science'. *Science*, *336*(6084), 982–983. doi:10.1126/science.1216878.

Nowotny, H. (2009). *Frontier research in the social sciences and humanities: What does it mean, what can it mean?* Paper presented at the ISSCICPHS workshop, Bergen. Retrieved from http://www.helga-nowotny.eu/downloads/helga_nowotny_b1.pdf.

Nowotny, H. (2010). What the ERC is looking for in applications. *Bioessays*, *32*(7), 545–548.

Nowotny, H. (2013). Preserve the European Research Council's legacy. *Nature*, *504*(7479), 189. doi:10.1038/504189a.

Risse, T., Sprungk, C., & Börzel, T. (2010). *Against the downsizing of social sciences in the EU. Memorandum*. Retrieved from http://www.polsoz.fu-berlin.de/en/v/transformeurope/news/allgemeines/2010_Against_Downsizing_Social_Sciences.html.

Schibany, A., & Gassler, H. (2010). Costs and benefits of (basic) research. *Technologie Innovation Politik*, *2010*(6), 1–9. Retrieved from http://www.joanneum.at/uploads/tx_publicationlibrary/TIP06_Policybrief_GLF_engl.pdf.

Stampfer, M. (2008). *European added value of community research activities. Expert analysis in support of the ex post evaluation of FP6*. WWTF-Vienna Science and Technology Fund. Wien. Retrieved from https://ec.europa.eu/research/evaluations/pdf/archive/fp6-evidence-base/expert_analysis/m.stampfer_-_european_added_value_of_community_research_activities.pdf.

Stokes, D. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: Brookings.

Tan, G. (2010). *Why do we have a European Research Council? (Unpublished master's thesis)*. Salzburg: University of Salzburg.

Vike-Freiberga, V., Sainsbury, D., Mény, Y., Schioppa, F. K. P., Röller, L.- H., & Zerhouni, E. (2009). *Towards a world class Frontier Research Organisation. Review of the European Research Council's structures and mechanisms*. Retrieved from http://ec.europa.eu/research/evaluations/pdf/archive/other_reports_studies_and_documents/erc_review_panel_report_-_23_july_2009.pdf.

Watson, J., Kitchener, M., Gutheil, M., Ward, B., Zadrozny, T., Ackers, L., & Harrap, K. (2010). *Evaluation of the impact of the Framework Programme on the formation of the ERA in social sciences and humanities (SSH). Final report*. European Commission. Brussels. Retrieved from https://www.kowi.de/Portaldata/2/Resources/fp7/coop/fp-ssh-evaluation.pdf.

Whitley, R. (2000). *The intellectual and social organization of the sciences* (2nd ed.). Oxford: Oxford University Press.

Winnacker, E.-L. (2008). On excellence through competition. *European Educational Research Journal*, *7*(2), 124–130. doi:10.2304/eerj.2008.7.2.124.

The Four 'I's: Quality Indicators for the Humanities

Wilhelm Krull and Antje Tepperwien

Abstract In a period, in which many things seem uncertain and yet everything is calculated and measured, the humanities can hardly avoid the evaluative quality measurement. However, a look into the world of benchmarks, ratings and rankings reveals that the oftentimes culture-specific performances of humanities research and teaching are almost immaterial therein. From the perspective of a private research funder, among others the following questions are traced: To what extent do international standards of quality exist in the humanities? Which criteria are suitable? Do assessment methods exist that allow for an adequate evaluation of performances in the humanities? To what extent should the humanities get involved with the construction of a publication and citation industry? What chance of survival do the humanities have in a world predominantly characterized by science and engineering?

1 Ranking Fever in Germany

A new era in German and European academic activities was launched on June 23, 2003, when the first Academic Ranking of World Universities (ARWU) was published by the Center for World-Class Universities (CWCU) at the Graduate School of Education (formerly the Institute of Higher Education) of Shanghai Jiao Tong University, China. It has been updated on an annual basis ever since.[1] The methods and criteria upon which the ranking is based are disputed, as the chosen indicators yield a strong bias favouring universities in English-speaking countries that focus

[1] On the Shanghai Ranking, see 'Academic Ranking of World Universities' at http://www.arwu.org.

W. Krull · A. Tepperwien (✉)
VolkswagenStiftung, Kastanienallee 35, 30519 Hannover, Germany
e-mail: tepperwien@volkswagenstiftung.de

W. Krull
e-mail: krull@volkswagenstiftung.de

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_13

on science and engineering.[2] Nevertheless, since the first 'Shanghai ranking' was published, Germany, like most other European countries, has been caught up in a ranking fever. This is evident not only in the nearly hysterical reaction to each new update of the ranking but also in the growing number of more or less 'homemade' national ranking lists that have appeared in recent years in diverse newspapers and periodicals.[3]

A quick look at these ranking lists shows how great the current demand for quantifiable assessment of the quality of teaching and research at German universities apparently is: The news magazine *Focus*, for instance, publishes an annual ranking of German universities that seeks to find out where in Germany the best research and higher education can be found based on surveys among professors, citation analyses and data from the German Federal Statistical Office. The news magazine *Der Spiegel* turned the tables, so to speak, and produced a ranking together with AOL and McKinsey that uses an online survey to assess the excellence of a university not based on the performance of its professors but on the achievement level of its student body (grades on school-leaving examination and university intermediate examinations). The business newspaper *Handelsblatt*, for the interests of its target group, reports on the top researchers and top faculties in the field of economics. The *Hochschulanzeiger* [higher education gazette] in the newspaper *Frankfurter Allgemeine Zeitung* compares the career success of graduates of private business schools in German-speaking countries. The newspaper *Karriere* chooses the best universities in the fields of economics, law, media sciences, mechanical engineering, electrical engineering, industrial engineering and computer science based on a survey of graduates, personnel managers and data from the German Federal Statistical Office. The *Wirtschaftswoche* business magazine publishes the results of a survey of 200 researchers on 'where Germany's best researchers in the 12 most important future technologies work' and also surveys personnel managers on the quality of graduates in economics, law, engineering sciences and computer science.

The Centre for Higher Education Development (CHE) in Germany would like its ranking to stand out among the others: First published in 1998, the CHE University Ranking covers study programs and is multidimensional.[4] However, this ranking,

[2]The following six indicators are decisive for a positioning in the Shanghai ranking: the number of alumni winning Nobel Prizes in physics, chemistry, medicine, or economics and Fields Medals in mathematics (10 %); the number of staff winning a Nobel Prize or Fields Medal (20 %); the number of articles written and co-authored by staff and published in the journals Nature and Science (20 %); the number of published articles written by staff and indexed in Science Citation Index - Expanded and Social Sciences Citation Index (20 %); the number of highly cited researchers at the university in 21 different fields (20 %); and per capita academic performance with respect to the size of the university (10 %). On the Shanghai ranking criteria and criticism of the criteria, see, for example, http://www.che-ranking.de/cms/?getObject=108\&getLang=d. Accessed 2 May 2014.

[3]An overview of the rankings is provided at the website of the Centre for Higher Education Development (CHE) at http://www.che-ranking.de/cms/?getObject=47\&getLang=de. Accessed 2 May 2014.

[4]On the CHE University Ranking, see http://www.che-ranking.de.

too, does not find favour with all universities and with all disciplines. For instance, the *Verband der Historikerinnen und Historiker Deutschlands* [Association of Historians in Germany] published a statement in 2009 refusing participation in ratings or rankings such as those conducted by the CHE (Historikerverband 2009).

2 The Reaction of the Humanities to the Ranking Fever

The historians' association's disapproval of rankings and ratings is an example of the difficulties that arise from the increasing demand for quantifiable evaluation of research in the humanities and social sciences. The historians' association not only rejects the larger and smaller forerunners, offshoots and competitors of the Shanghai ranking but also does not support the efforts of the *Wissenschaftsrat* (German Council of Science and Humanities) to put forward a differentiated research rating as an alternative to the overly simple and often methodologically unsound rankings by private providers: After lengthy discussions, the historians' association refused in 2009 to participate in a research rating conducted by the *Wissenschaftsrat*, which had previously conducted ratings in sociology and chemistry (Historikerverband 2009).

The historians' association acknowledged the intention of the *Wissenschaftsrat* to rate different fields in a differentiated manner and according to a catalogue of criteria negotiated upon by representatives of the fields themselves, in contrast to the procedures by other rankings. But fundamental doubts as to whether it makes sense to create such a rating and to submit to the demand for quantifiable data led to disapproval by the association. In a statement on April 4, 2009, the then president of the association, Werner Plumpe, said that the opponents of a research rating in the historical disciplines doubt the sense and meaning of such a rating. Plumpe (2009, p. 123) summed up the position of the rating opponents as follows:

> Hier könne es allein aufgrund der Unmöglichkeit, ein dynamisches Fach wie die Geschichtswissenschaft parametrisch gleichsam in einer Momentaufnahme abzubilden und wertend zu erfassen, zu keinen sinnhaften Resultaten kommen. Was dabei herauskomme, seien teilweise quantifizierte, immer aber parametrisierte Informationen für politische Diskussions- und Entscheidungsprozesse, die gemessen an der Realität des Faches unterkomplex seien, der Politik aber das Gefühl des Informiertseins durch die Wissenschaft selbst vermittelten. Auf diese Weise bediene der *Wissenschaftsrat* letztlich die politische Illusion, Wissenschaft lasse sich parametrisch durch das Setzen bestimmter Anreize steuern, und fördere damit die Herausbildung und Verfestigung strategischer Verhaltensweisen, die zumindest in den Geisteswissenschaften die akademische Kultur zerstörten. Das Fach habe es aber weder nötig noch sei es im eigenen Interesse verpflichtet, die gefährlichen Illusionen der derzeit politisch hegemonialen Strömungen zu bedienen.
>
> [Here there can be no sensible results, due already to the impossibility of portraying a dynamic discipline like history parametrically in a snapshot, so to speak, and capturing it in a rating. The result would be partly quantified but always parameterized information for policy discussions and decision processes; the information would be under-complex compared to the reality, but it would give the politicians the feeling of being informed by science itself. In that way, the *Wissenschaftsrat* would ultimately serve the political illusion

> that science and scholarship can be steered parametrically by setting certain incentives, and this would thus promote the development and hardening of strategic behaviours that at least in the humanities would destroy the academic culture. But the discipline does not find it necessary, nor does it feel obligated in its own interest to serve the dangerous illusions of the current politically hegemonic trends.] (Plumpe 2009, p. 123)

In addition to these fundamental concerns, Plumpe (2009) reported that in the opinion of the rating opponents, it was also questionable how a rating could produce meaningful results unless it were continuously repeated—and would thus cost so much time and work that expenditure would be disproportionate to yield and would devour so much capacity (in the reporting and evaluation process) that it would run counter to the intention to improve the quality of research and teaching.

When the historians' association finally decided in the summer of 2009 not to participate in the rating—to boycott it essentially—their press release stated that it supported the concern of the *Wissenschaftsrat* to actively participate with the professional associations in reaching agreements on standards in the disciplines and in jointly developing discipline-specific criteria for research quality, but that it had fundamental reservations against the usefulness and feasibility of the rating being planned. In its statement, the association emphasized clearly that German historians were conscious of their responsibility to be accountable to the public and also signaled its willingness to participate in an appropriate form in the search for suitable concepts and in an open-ended discussion on the possibility of developing and measuring quality standards in the humanities (Historikerverband 2009).

This much is certain: In a time when so many things seem uncertain and yet everything is calculated and measured, the humanities can hardly avoid evaluative measurement of quality. A look at the world of benchmarks, ratings and rankings shows, though, that the often culture-specific achievements of humanities teaching and research do not really play a role in them at all. And the instruments used to create rankings do not do justice to the disciplines in the humanities.

3 Quantity Instead of Quality: Current Methods of 'Quality Assessment'

Just how unsuitable current methods, such as making the number and impact of publications measurable and verifiable as quality standards, are for quality assessment in the humanities can be shown by a look at the database of Thomson Reuters (originally called the Institute for Scientific Information and still later Thomson Scientific).[5] Its data analyses can only work in disciplines where the database contains not only the citing works but also the majority of the cited works. Whereas this is so for up to 100 % of the cases in the big disciplines in the natural sciences, this congruence is only 40–60 % in mathematics and economics. In the social sciences and humanities,

[5] See http://science.thomsonreuters.com.

the percentage is even much lower. For instance, in literary studies, only 11 % of the works cited are also contained in the database.

This example of the difficulties in assessing quality in the humanities and social sciences using instruments that are geared to the natural sciences was pointed out by Christoph Schneider, who for many years headed the department of scientific and scholarly affairs at the German Research Foundation (*Deutsche Forschungsgemeinschaft*, DFG). In an article in the *Frankfurter Allgemeine Zeitung* in October 2009 titled 'Zauberlehrlinge im Rate- und Ränkespiel' [Sorcerer's apprentices in the rating and ranking game], Schneider wrote on the new measurement madness that just as Midas in the Greek myth turned everything that he touched into gold and thus starved to death, evaluators that are obsessed with ranking lists turn everything into numbers, which soon distorts their reality (Schneider 2009).

It is now sufficiently well-known that quality assessment methods that have in part proved their worth in the natural sciences cannot be applied 1:1 to the humanities and social sciences. The differences between the two in their publication and communication cultures are too great. Often there is very little understanding of or knowledge about the 'other side'.

In a 2006 article in *Die ZEIT*, social psychologist Harald Welzer wrote about his experiences collaborating with a neurophysiologist in an interdisciplinary research project supported by the Volkswagen Foundation. Welzer felt that the often mentioned speechlessness between the disciplines is not it at all; instead it is cultural differences between the disciplines that make it difficult to engage in exchange. Welzer (2006, p. 1, par. 4) asked in *Die ZEIT*:

> Wer hätte sich je Gedanken darüber gemacht, dass die disziplinären Vorstellungen von einer "wissenschaftlichen Veröffentlichung" so voneinander abweichen, dass es fast unmöglich ist, gemeinsam einen Text zu verfassen? Für mich als Sozialwissenschaftler war es höchst befremdlich, noch die stumpfesten Hauptsätze, zu denen ich fähig war, von den Gutachtern eines Fachbeitrags als "episch breit" kritisiert zu finden, während im umgekehrten Fall Gutachter sozial- und geisteswissenschaftlicher Journale Phänomene wie die "zunehmende Reaktionsgeschwindigkeitsverminderung" für ziemlich absonderlich hielten.
> [Who ever thought that the disciplinary notions of a 'scientific or scholarly publication' would differ so greatly that it is nearly impossible to jointly write a text? For me as a social scientist it was highly disconcerting to have reviewers of a scientific article criticize even the dullest substantive clauses that I was capable of for being 'epically broad', whereas in the opposite case, reviewers for social sciences and humanities journals deemed phenomena such as 'increasing reaction rate reduction' quite peculiar.] (Welzer 2006, p. 1, par. 4)

Whereas in the natural sciences ground-breaking research findings are published in a handful of international journals known to all members of the scientific community in a given discipline, the main form of publication in the humanities continues to be the monograph, which is almost always written in the author's native language. Whereas in the natural sciences people argue about which author of a journal article should be listed in what position, the concept of 'first author' is hardly known in the humanities. In the humanities, excellence is still based mainly on the research achievements of individual scholars and not on the joint efforts of a research team. Current methods of quantitative assessment only very insufficiently take into account these different forms of knowledge creation and publication.

The amount of third-party funding is another example. Naturally, the natural sciences and engineering play in a very different league here, for their work requires in part expensive equipment and materials as well as support by technical personnel. In addition, they pay their research assistants, at least those with doctorates, full salaries. A researcher in the humanities, in contrast, requires mainly time, a good library and possibly money for trips to archives or for field research. For the humanities scholar, the time to conduct research gained by the funding of his position or of a temporary stand-in for his position is as valuable as the costly laboratory equipment is for the natural scientist. But for the third party, this type of research is of course considerably less costly, and in the third-party funding statistics it makes up approximately one-tenth of the amount of third-party funding that is customary in engineering and medicine.[6] If management boards of universities look only at the amount of external funding granted to researchers, they are in essence comparing apples and oranges. And they are also in danger of taking mere activity measures for evidence of achievement.

At present, therefore, the comparatively recent drive to assess quality in numbers puts the humanities rather at a disadvantage. At least they feel pressured and once again pushed into a corner. But it is clear even to critics of the current rankings and ratings that in the long term they cannot evade this trend towards assessment and evaluation. So the question is how to evaluate quality in the humanities appropriately.

4 Quality Assessment within a Discipline: The Evaluation Culture in the Humanities

Within the academic community assessment takes place constantly: when positions are filled, appointments are made, scientific or scholarly works are accepted by publishers, and third-party funding is granted. This quality assessment is based for the most part on criteria recognized within the community that are not measurable in numbers and that adhere to performance criteria.

A look at peer reviewers' reports provides deep insight into customary quality evaluation methods within a discipline. The Volkswagen Foundation, which funds research in all disciplines, is dependent on peer review of the research grant applications submitted by applicants. Some general things hold for all peer reviewers' reports, such as, for instance, that reports that recommend not funding a project tend to be longer than reports that recommend approving a project for a grant. But also in the peer review and assessment culture there are some fundamental differences between the humanities and the natural sciences. The Volkswagen Foundation sends a leaflet to all peer reviewers asking them to assess the following general criteria in their written reports (VolkswagenStiftung 2013, p. 2):

[6] See here, for example: Berghoff et al. (2009). *Das CHE-Forschungsranking deutscher Universitäten 2009*. Gütersloh, Germany: Centrum für Hochschulentwicklung.

1. Contribution to the further development of research:
   What place does the proposal take within the framework of the scientific or scholarly development in the respective area? What is new and original in the approach? What will be the benefit in terms of new knowledge to be acquired?
2. Clear-cut description and consistency:
   Does the project proposal reflect the present state of the art? Are the objectives clearly defined and attainable? Are the proposed methods and the working scheme adequate in order to achieve the project goals?
3. Personal qualification:
   What about the competence of the project staff, their publication record (also in consideration of their biography, e.g. family phase) and the preparatory work for the project?
4. Adequate extent of time, staff and consumables:
   Are the estimated time, staff and consumables really required to achieve the proposed objectives? On which budget items could savings be made or funds be reallocated?
5. Recommendations on the realization:
   Does the peer reviewer have helpful suggestions for conducting the project that, should the grant be approved, should be communicated anonymously to the grant applicant?

The Volkswagen Foundation lists these same aspects for the review of grant applications in all disciplines. The standards applied are of course the standards that are valid in the respective scientific or scholarly community from which the grant application comes. The peer reviews of grant applications are usually considerably longer in the humanities than in the natural sciences and engineering, and as to content—depending on the particular culture in the respective discipline—are more critical in their examination. In the humanities, even grant applications that in the end are unreservedly recommended for funding by the peer reviewer are often analysed in detail and criticized. Sometimes, there is an amazing discrepancy between the accompanying assessment sheet, on which the peer reviewer rates the applicant on criteria such as qualifications in the specific field, interdisciplinary potential and research chances for the future, and rates the project on quality, originality and complexity on the one hand, and the peer reviewer's lengthy written report on the other. Even if the peer reviewer gives an overall rating of 'excellent' on the assessment sheet, that does not mean that the grant application will not be taken apart point by point by the peer reviewer in the written report. In-depth examination of a grant application by an esteemed peer is seen as a 'token of love', so to speak, or for the peer reviewers, who see themselves as equals, as a kind of 'matter of honor'. This type of evaluation may work within the discipline, but where humanities scholars are competing with natural scientists for funding, this culture has a negative impact on the humanities' chances of winning. In the Volkswagen Foundation, for instance, this can be seen with the Lichtenberg Professorships, which are open to applicants in all disciplines.

And also in the context of the Excellence Initiative, as well as in several multistage review and selection processes, this difference in the evaluation cultures has all too often had a negative effect on the humanities' chances of success.[7]

But within the humanities, quality assessment functions more or less smoothly. It is usually not difficult for a peer or an editor at an academic publishing company to determine the quality of the work of an individual researcher. But how do humanities scholars communicate their evaluation culture, which is so frequently accompanied by fundamental criticism of the proposed research questions and methods, to colleagues in the natural sciences and engineering? And how do they handle it when they are expected to measure the quality of a department or an entire faculty and have to explain their evaluation results using numbers and facts in a way that the public can understand and verify?

Up to now, the humanities still owe an answer to the question of how quality can be 'measured' in the respective disciplines appropriately. There is no doubt that the instruments for quality assessment used in the natural sciences and engineering cannot be applied to the humanities. Those instruments are also not appropriate for several other disciplines, because often—as it appears, at least—today's rankings and ratings use quantitative and quantifiable criteria and disregard non-quantifiable criteria, as non-quantifiable criteria can be determined only at considerably greater expense. But if there is a demand for reference to qualitative criteria, the following question has to be answered: What is quality in the humanities?

5 What Is Quality in the Humanities? Looking Back

A central topic in humanities research is the analysis of past times, or more precisely, recording, revealing and conveying cultural material as an important part of our cultural heritage. Perhaps to answer the questions as to what quality is in the humanities and how it can be measured we need to look not only at the present and at other countries but also at the past, at the heyday of humanities research in Germany. Why are the late 1800s and early 1900s characterized as a kind of heyday? This is because of the then international impact of German humanities research, the great attractiveness of the German universities for students and scholars from abroad, and the transfer to other countries of forms of teaching and research methods developed in Germany.

What about that impact today? Whereas the natural sciences and engineering have settled on a more or less good laboratory English as the lingua franca, the vast majority of the humanities disciplines remain bound to national languages. The decline of

[7] On funding decisions in the Excellence Initiative, see http://www.dfg.de/foerderung/programme/exzellenzinitiative/allgemeine_informationen/index.html. Accessed 2 May 2014.

German as a language of science and scholarship as well as the decreasing importance of German-language acquisition are inextricably linked. But disciplines that work in and through language cannot simply throw off the respective language. Humanities scholars have to write in the language in which they think, and at the same time they must learn several languages so as to be able to participate in the scholarly debates in other countries.[8] In a certain way, the following comment by Jutta Limbach, former president of the Goethe Institute, also holds true for the humanities: '*Englisch ist ein Muss, Deutsch ist ein Plus*' [English is a must, German a plus] (Limbach 2005, p. 3). If research quality of the humanities can be measured among other things via international attractiveness, then this does not mean that this attractiveness can be increased by the number of courses of study taught in English offered in the humanities. Instead, it is the bilingual or trilingual courses of study that are conducted in cooperation with universities abroad that can increase the international visibility and attractiveness of the German humanities. Exchange programs and the presence of up-and-coming young scholars and established professors at international conferences promote the networking of the international academic community in all humanities disciplines and make possible the exchange of research findings and methods and, with this, at the same time also make the high quality of humanities research in German-language countries visible in international circles.

Measurement of quality in the humanities along the same lines as in the natural sciences and engineering does not work. The fact that quality in the humanities is more difficult to quantify does of course not mean that quality does not exist. Even though the international attractiveness of the humanities disciplines in German-speaking countries has declined, its transmission has not faded.[9] Humanities scholars trained here, if they also possess the needed language competency, have good chances on the international research labor market. However, the high qualifications of the up-and-coming researchers say only so much about the quality of a discipline in research and teaching. Only a small percentage of university students enrolled in humanities programs seek an academic career or even have any chance at all to have a successful research career, despite the fact that studies at German universities, especially in the humanities, are still frequently mainly geared to qualifying students for research careers. In Germany, a large part of the humanities disciplines belong to the massively attended study programs with high numbers of students, unfavourable teacher-student ratios and in part dramatic drop-out rates.[10]

[8] A conference on the topic *Deutsch in der Wissenschaft* [German in science] was held at the Akademie für Politische Bildung in Tutzing from January 10–12, 2011. The papers were published in a conference volume (Oberreuter et al. 2012).

[9] See Behrens et al. (2010). Die internationale Positionierung der Geisteswissenschaften in Deutschland. Eine empirische Untersuchung. Hannover, Germany: HIS-Projektbericht.

[10] For a current analysis of the situation of the humanities in Germany, see the recommendations of the *Wissenschaftsrat* on development and promotion of the humanities in Germany (Wissenschaftsrat 2006).

6 The Critical Self-image of the Humanities

Summing up the discussion in and about the humanities in Germany, the following picture emerges: Long disregarded by government, poorly equipped, underfunded and standing practically no chance in the competition for the big third-party public funds, the humanities seem to eke out a pitiful existence.[11] The critical self-image of the humanities, which was being clearly expressed already in the 1980s, can be illustrated by the following three quotations:

Joachim Dyck, a Germanist at the University of Oldenburg, lamented in an article in the periodical *Die ZEIT* as early as 1985:

> Wo noch vor 15 Jahren die Rede- und Ideenschlacht tobte, gibt es heute als Geräusch nur noch die leise Klage der Hochschullehrer über die dürftigen Schreib- und Leseversuche einer sprachlos gewordenen Generation und den beflissenen Wortschwall von Studenten, deren abgeleiertes Referat vom meditativen Klappern der Stricknadeln begleitet wird, in der Hoffnung, dem geistigen Leben durch handwerkliche Nebentätigkeit noch einen Hauch von Sinn abzuringen.
> [Where 15 years ago there was a wild war of words and ideas, today there is only the sound of the university teachers' soft complaint about the meager attempts by a generation gone speechless to read and write and the assiduous torrent of words of students whose reeling off of their presentations is accompanied by the meditative rattle of knitting needles, in the hope of wresting some small sense out of the intellectual life by engaging in handicraft.] (Dyck 1985, p. 2)

In 1989, philosopher Jürgen Mittelstraß wrote on the splendor and misery of the humanities as follows:

> Über den Geisteswissenschaften liegt nämlich ein wissenschaftsideologischer Fluch, den 1959 Charles Percy Snow, Physiker, Romancier und hoher britischer Staatsbeamter mit seiner Rede von den zwei Kulturen, der naturwissenschaftlichen und der geisteswissenschaftlichen ('literarischen') Kultur in die Welt gesetzt hat. Er tat dies eher nebenbei, in einer Art Sonntagsrede und doch mit ungeheurer Wirkung, vor allem bei den Geisteswissenschaftlern. Diese Wirkung besagt denn auch vielleicht nicht so sehr etwas über den Wahrheitsgehalt der Snowschen Vorstellungen, als vielmehr etwas über die Nervosität und den Selbstzweifel, die die Geisteswissenschaften ergriffen haben.
> [There is a curse on the humanities, a science ideology curse that was introduced into the world in 1959 by British physicist, novelist and high government official C. P. Snow in his lecture on 'The Two Cultures', namely, the sciences and the humanities (or literary culture). Snow did this rather incidentally, in a kind of crowd-pleasing speech, but it had enormous impact, especially among humanities scholars. The impact possibly says not so much about the truth of Snow's ideas and very much more about the nervousness and self-doubt that had seized the humanities.] (Mittelstraß 1989, p. 7)

And finally, Hans-Joachim Gehrke, former president of the German Archaeological Institute in Berlin, wrote the following in the DFG journal *Forschung* in 2008: 'In vielen geisteswissenschaftlichen Fächern steht man bereits mit dem Rücken zur Wand. Weitere Kürzungen werden in manchen Bereichen unmittelbar zum Exitus führen' [Many humanities disciplines are already standing with their backs to the wall. In some fields any further cuts will lead directly to exitus] (Gehrke 2008, p. 3).

[11] On the self-image of the humanities, see also Koschorke (2007).

Instead of joining in the chorus of complaints, in the following we will attempt, going beyond the *Gekränktheitsrhetorik* [offended rhetoric] (a term by Peter Strohschneider),[12] to point out not only risks but also and especially development opportunities of the humanities, looking at four areas that all begin with the letter 'I', namely, infrastructure, innovation, interdisciplinarity and internationality. At the same time, we will indicate in what areas quality can be found and possibly also measured in the humanities.

7 Quality Indicators: The Four 'I's

The first 'I' stands for *infrastructure*—the foundation of humanities research. Infrastructure is what the humanities disciplines absolutely should have and should strengthen: Libraries, archives and museums are of fundamental importance for cultural memory and for the study of the cultural foundations of societies. However, these institutions are currently undergoing rapid change and are finding themselves caught between the increasing fast pace in the times of the Internet and the central concern of libraries, archives and museums, namely, the long-term availability of their holdings. By promoting simultaneity, interactivity and open access, the new media also open up quite new possibilities for research. But we need to be concerned about the neglect of the permanence of the documentations—short-term life as a consequence of fast availability! Here the task is to assure and protect quality.

The second I stands for *innovation*. This word has so many facets, all of them associated with renewal, novelty and change, that it is difficult to define the term precisely. For many humanities scholars, who see themselves as custodians of their own and others' traditions (Gehrke 2008, p. 3), the concept of innovation and also nearly any future orientation is the opposite of their central concern. They view as their very own and only task the examination of the past—interpretative learning, understanding and imparting traditions. With this attitude, they are in danger of confirming the popular prejudice, often expressed on the part of natural scientists, that says that the humanities deal too much with the ashes of the past as opposed to what is really important, namely, promoting the fire of the future and driving forward scientific and technical research with quickly measurable results. However, this is a false contrast, because a 'fire of knowledge' fed by the here and now alone is all too frequently likely to turn out to be a rapidly extinguishing flash in the pan. However, we can counteract a just as memory-less and unrestrained belief in progress successfully only if we are willing to always create new perspectives and to learn beyond times and borders, in the conviction that the past must always be present in the present day, if we aim to design the future in a responsible way (Krull 2003, p. 32).

In addition to their classical function of cultural memory—namely, mining, saving and conveying the cultural heritage—perhaps the most important function of the

[12] Strohschneider, cited in Hinrichs (2007).

humanities is preventive thinking. The latter is designed to advance our potential to reflect on relevant issues and, with this, to contribute towards working out future options more clearly. Particularly in times of great uncertainty, preventive thinking is more than ever an indispensable task of the humanities. Here lies the innovation potential of humanities research; the full utilization of that potential is without question a quality criterion for humanities research. This, of course, is a criterion that materializes only in idea-rich communication and interaction both within research and also at the interface of research and the public.

The third I stands for *interdisciplinarity*. In academia itself, the disciplinary orientation dominates: Individual disciplines' reference systems with regard to quality assurance (standards), certification through the awarding of academic degrees, reputation, stability of the field and not least career prospects stand in the foreground. They make up, so to speak, the university's organizational form of knowledge.

But government, economy and society expect researchers to provide solutions for the 'big' questions and not just small and fragmented answers from the perspective of one discipline. In the attempt to establish a balance between the necessary raising of the specialist field profile of the individual discipline and the also necessary bundling of research and teaching capacity, what is practiced for the most part is a kind of contact-free, added-on interdisciplinarity. Due to cost-benefit considerations, usually no effort at all is made to produce common methodological procedures or joint publications. This is even often considered to be extremely career-damaging.

In the age of measurements of science that are oriented towards the leading journals in the different fields, this discipline-specific publication strategy may have an understandable rationality, especially for up-and-coming scholars, particularly as the time cycles of research funding (with still predominantly two- to three-year funding periods) practically promote a narrow focus. However, this should be counteracted against and long-term perspectives should be opened up, so as at the same time to encourage researchers to be willing to take risks and to step outside disciplinary boundaries. If the humanities make their contribution towards answering the big questions and make that contribution visible to the outside world, then they will also be demonstrating their high standards of research quality and importance to society.

The fourth I focuses on *internationality*, which was already mentioned above. Research is inconceivable without international cooperation. At the same time, European integration and the process of globalization are presenting a particular challenge to education, science, research and technology. If the university is to remain attractive and alive as a place for teaching, research and innovation, then it will be essential to develop a culture of intercultural openness and internationality. The humanities in particular can contribute towards the creation of new perspectives and learning options that transcend borders and times.

Particularly with regard to the risks and opportunities of globalization processes there are still a lot of open questions. For this reason, what is needed is stronger research collaboration across disciplinary, institutional and national borders; only on the basis of new knowledge can the future global challenges be tackled effectively. For future research projects, this means that they must make the process of globalization a constitutive aspect of the respective project architecture. This requires, for one, the

integration of researchers from different disciplines and cultures and, for another, steady networking with a circle of researchers worldwide, who can all make their contribution in the horizon of the research question. The other way around, effective utilization of globalization opportunities also makes necessary increasing acquisition of culture-specific knowledge. The humanities should increase their commitment also in this area and should make international exchange, international networking and international cooperation an important criterion for quality assessments.

8 Closing Remarks

Today's almost simultaneous production, processing and communication of new knowledge also makes necessary a new self-understanding of science, scholarship and research: a shift from a homogeneously structured process firmly anchored in institutions and characterized by disciplinary discourses to a more open process that is often kicked off by questions from outside the discipline and characterized by a firm connection to society as well as problem-oriented methods.

There is a reason why the humanities in Continental Europe hid from these changes for too long: The model being followed—the research university and its disciplinary top-level research—made Germany a world leader in science and scholarship in the nineteenth century. But already beginning in the 1890s, scientific developments mainly in the natural sciences and engineering began to break up Humboldt's unity of research and teaching, which had been raised practically to an ideology. In an essay on the creation of the German research university, Brocke (2001, p. 386) wrote that the increasing inability of the institution of the university to do equal justice to the tasks confronting it—classical education, professional training and scientific research—caused a constantly growing discrepancy between the neohumanist conception of the university and the universities' actual structure.

Thus, the problems of the Continental European university system virulent today were already marked out at the start of the twentieth century: the insufficient consideration of new disciplines in the traditional university structure, the increasing specialization in all fields, the impossibility of interdisciplinary research within the given structures (which were mostly vehemently defended by the professors) and not least the resulting explosion of costs in the natural sciences and engineering, which through the necessity for savings had a negative impact on the humanities.

The undoubtedly justified sense of pride in an exemplary and productive university system in the past became a counterproductive mentality of protection of vested interests and blindness to scientific, scholarly and societal reality. For this reason it seems all the more urgent now—despite the many difficulties in everyday university operations—to look forward to new possibilities and options. Particularly considering the globalization processes mentioned above, the humanities can definitely profit from the institutional context of increasingly internationalizing universities. To benefit, however, the humanities must be willing to participate more than before in present-day debates and training needs.

There is no reason for the humanities to remain 'with their backs to the wall' or to give up all hope in the face of the supremacy of the natural sciences and engineering. It would also be wrong to overeagerly adopt the research and evaluation modalities of the natural sciences and to artificially create indices for the humanities. The European Reference Index for the Humanities (ERIH) promoted by the European Science Foundation and the controversies over its methodology and meaningfulness will suffice here as an example to point out that the appropriateness of such measurement methods should be called into question.[13]

One thing is clear: Humanities research requires a different kind of 'measurement' and promotion instruments than the instruments used in the natural sciences. If the quality assessment instruments customary in the natural sciences and engineering were applied 1:1 to the humanities, it would only be to the humanities' disadvantage and would lead to a false snapshot showing only a distorted picture far from reality. Nevertheless, the humanities must make stronger efforts to develop criteria and measurement instruments that go beyond the usual activity measures for assessing good housekeeping. They should make quality in the humanities visible, understandable and recognizable not only within the community in specific disciplines but also to the outside world and to the public. Naturally, it can make sense for the humanities to utilize the usual publication and third-party funding indicators as comparison measures. However, they should be embedded in a clearly structured benchmarking concept that can be used to evaluate comparable institutions—such as, for example, German universities with rich traditions and equipped with a high capacity in humanities teaching and research, such as the universities of Bonn, Göttingen, Heidelberg, Tübingen and Freiburg. A concept of this kind might possibly be realizable also across national borders in the European university and research area and could lead to actual 'learning by comparing', if it combined quantitative and qualitative elements of evaluation.

The humanities are very important for the investigation of past problems, the analysis of present-day changes and for coping with future challenges. The humanities can also serve as a reliable compass in times of rapid change if they themselves are clear about their specific quality and significance and demonstrate this to the outside world. The humanities should not respond to the omnipresent call for quality measurement by inappropriately adopting the practices of other disciplines or by fighting a futile defensive battle. Instead, the response should be a committed, interdisciplinary debate, conducted in international dialogue, on suitable methods of transparent quality assessment in the humanities, which know how to utilize quantitative indicators and at the same time combine them with qualitative evaluation methods.

[13] See, for example, the opinion of philosopher Stekeler-Weithofer (2009).

References

Behrens, J., Fischer, L., Minks, K.-H., & Rösler, L. (2010). *Die internationale Positionierung der Geisteswissenschaften in Deutschland. Eine empirische Untersuchung*. Hannover: HIS-Projektbericht.

Berghoff, S., Federkeil, G., Giebisch, P., Hachmeister, C.-D., Hennings, M., Roessler, I., et al. (2009). *Das CHE-Forschungsranking deutscher Universitäten 2009*. Gütersloh: Centrum für Hochschulentwicklung.

Dyck, J. (1985). Stumm und ohne Hoffnung. Die totale Paralyse der Germanistik in den 80er Jahren. *Die Zeit, 25*. Retrieved from http://www.zeit.de/1985/25/stumm-und-ohne-hoffnung.

Gehrke, H.-J. (2008). Erfolg auf wackeligen Beinen. Was die Geisteswissenschaften fördert – und was sie bedroht. *Forschung*, *1*, 3.

Hinrichs, P. (2007). Platons Erben. *UniSPIEGEL, 3*, 8–12. Retrieved from http://www.spiegel.de/spiegel/unispiegel/d-51609364.html.

Historikerverband. (2009). Pressemitteilung: Ausschuss des Historikerverbandes beriet über Forschungsratings in den Geisteswissenschaften und über den Historikertag 2010. Retrieved from http://www.historikerverband.de/fileadmin/_vhd/bilder/Pressemitteilung_WR_Rating.pdf.

Koschorke, A. (2007). Über die angebliche Krise der Geisteswissenschaften. *APuZ*, *46*, 21–25.

Krull, W. (2003). Zwischen Tradition und Innovation. Die Geisteswissenschaften im erwei-terten Europa. In Germanistentreffen Deutschland – Italien. 8.-12.10. (2003). *Dokumentation der Tagungsbeiträge* (pp. 29–44). Köln: Deutscher Akademischer Austauschdienst.

Limbach, J. (2005). Ich liebe unsere Sprache. Rede im Rahmen der Berliner Lektionen anlässlich der Berliner Festspiele 2005. Retrieved from http://www.goethe.de/mmo/priv/367926-STANDARD.pdf

Mittelstraß, J. (1989). *Glanz und Elend der Geisteswissenschaften*. Oldenburg: Bibliotheks- und Informationssystem der Universität.

Oberreuter, H., Krull, W., Meyer, H. J., & Ehlich, K. (2012). *Deutsch in der Wissenschaft. Ein politischer und wissenschaftlicher Diskurs*. München: Olzog.

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/.

Schneider, C. (2009). Zauberlehrlinge im Rate- und Ränkespiel. *FAZ, 228*.

Stekeler-Weithofer, P. (2009). Das Problem der Evaluation von Beiträgen zur Philosophie. Ein streitbarer Zwischenruf. *Deutsche Zeitschrift für Philosophie*, *57*(1), 149–158.

VolkswagenStiftung. (2013). Guide to peer review. Retrieved from http://www.volkswagenstiftung.de/uploads/media/peer_review.pdf.

vom Brocke, B. (2001). Die Entstehung der deutschen Forschungsuniversität, ihre Blüte und Krise um 1900. In R. C. Schwinges (Ed.), Humboldt International: Der Export des deutschen Universitätsmodells im 19. und 20. Jahrhundert (pp. 367–401). Basel: Schwabe.

Welzer, H. (2006). Nur nicht über Sinn reden! *Die Zeit*. Retrieved from http://www.zeit.de/2006/18/B-Interdisziplinaritt_xml.

Wissenschaftsrat. (2006). Empfehlungen zur Entwicklung und Förderung der Geisteswissenschaften in Deutschland. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/download/archiv/geisteswissenschaften.pdf.

Bottom Up from the Bottom: A New Outlook on Research Evaluation for the SSH in France

Geoffrey Williams and Ioana Galleron

Abstract This paper will start with a presentation of the legal French framework for research evaluation, concentrating on the individual level; this first part will also summarize the main oppositions to the idea of evaluation, as they are expressed mainly by unions and other researcher associations. In a second move, we will review the main French actors and practices of evaluation, separating the 'traditional' forms of assessment still in use in the CNU, and the recent evolutions caused by the introduction of a dual financing system (through ANR), of an external evaluation of research units by an independent agency (AERES/HCERES) and by the building of a database in the CNRS. In the light of criticisms that can be formulated about all these practices, we will introduce the projects DisValHum and IMPRESHS, dedicated, respectively, to a study of dissemination strategies in the SSH and to case studies of the impact of the research in the SSH. The third part of the paper will therefore be occupied by a description of our methodology and of a few results.

1 Introduction

The French legal framework for research evaluation underwent major changes following the 'loi relative aux libertés et responsabilités des universités' (loi LRU). This reform left former evaluative practices in place, whilst bringing in a new evaluation agency, AERES, itself recently replaced itself with a 'High Council of the Evaluation' (HCERES). After a presentation of the French research evaluation landscape, as reshaped by the loi LRU, the paper will concentrate on the criticisms that have been formulated about the actors, tools and methods, as well as the place given to the social sciences and humanities (SSH) in this process. In the last section, we will focus on two projects, DisValHum and IMPRESHS, dedicated, respectively, to a study of dissemination strategies in the SSH research and to case studies of the impact of

G. Williams (✉) · I. Galleron
LIDILEM, University Grenoble Alpes, BP 25,
38040 Grenoble Cedex 9, France
e-mail: williams@evalhum.eu

I. Galleron
e-mail: galleron@evalhum.eu

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_14

the research in the SSH. Because both projects are still under development, we will describe our methodology and will present only a few preliminary results.

2 The Need for Evaluation in the Post-'loi LRU' Period

During the last decade, the need for evaluation increased in all higher-education systems. This movement did not spare France, in spite of this country's tendency to stay away from the general trends in culture-related matters[1] and, more specifically, in education issues, as shown, for example, by France's non-participation in the European University Data Collection (EUMIDA) surveys (European Commission 2010). Nevertheless, the claims and methods of the so-called new public management did find a favourable echo in France among some politicians and members of the administrative apparatus. In the meantime, the Shanghai rankings came as a shock to the system, and still create a huge discussion about the low ranking of French universities in the top 50 and top 100 league tables (AEF 2013b, 'Dépêche no. 186447'). A considerable shift in public policy on the higher-education system was, therefore, made under Nicolas Sarkozy's presidency (2007–2012). The most conspicuous and explicitly stated goal of this change was to create 10 highly performing higher-education and research institutions. These were meant to better represent France in international competitions in research and education, as well as to boost academic standards. The latest law on higher education and research ('loi ESR', as it is commonly called in France) brought in by the current government did not renounce this objective, nor did it go against the major changes brought in by the 2008 law (loi LRU)—to the disappointment of many left wing supporters from academia who were pushing for a return to the status quo ante.

Following the changes brought about by this new policy the need for a better organized and a more thorough research evaluation became acute in three key sectors.

2.1 Human Resources

Under the loi LRU, the universities were allotted new duties and competencies regarding the management of their staff. The novelty is that the institutions are now not only allowed, but also invited, to define human resources strategies and policies covering the three major issues of recruitment, promotion and continuous training. Even if this newly acquired freedom is far from complete—as proved by the autonomy dashboard of universities in Europe, in which France scores low (Estermann et al. 2011)—it opened a whole series of possibilities, which in return prompted a new series of questions to be solved.

Under the previous legal framework, recruitment of research and education staff was performed by 'commissions des spécialistes' (recruitment panels). Elected for four years, these panels recruited academic staff, sometimes without any assessment of applications for a position by real specialists in the recruitment field. Now, institu-

[1] This is an accepted political doctrine, well known in France as 'l'exception culturelle'.

tions must put together profile-oriented committees whenever the need arises. These new committees must also justify the ranking of candidates. Thus, both aspects of the hiring process (selection of specialists and candidates), now require a reflection as to quality criteria, even if the rationale is, in most cases, quite flimsy or biased by hidden assumptions.[2]).' The change towards position-specific recruitment panels was also designed to address the issue of endo-recruitment, an issue closely followed by the Ministry of Education, which actively seeks to limit this practice. Panels now include a significant number of members from outside the recruiting university, whose external point of view is supposed to prevent favouritism and to ensure the homogeneity of standards throughout the French Higher Education (HE) system. By making the selection process less opaque, the loi LRU has opened new vistas for research evaluation in France.

The loi LRU not only brought changes in recruitment, but also in promotion practices. The possibility to promote staff members is not a new issue for the French Higher Education Institutions (HEI),[3] but the novelty is that institutions must now publish their criteria for any decision. Such a requirement was nonexistent prior to the accession to 'responsabilités et compétences élargies' (widened responsibilities and competencies) guaranteed by the loi LRU of 2008. Thus, this can be seen as a first step toward a more thorough evaluation of individual careers at the national level, even if numerous voices are to be heard opposing any form of individual evaluation of researchers (CP-CNU 2012; Sauvons l'université 2012). Certain sections of the Conseil National des Universités (CNU), the body that oversees recruitment and promotion procedures,[4] proved, in such a context, more sensitive to the weaknesses in the methodology applied for assessing files (Garçon 2012) and opened internal

[2]In SSH disciplines, particularly in literary and language fields, it is not unusual for members of the selection committees to filter applications by considering if the candidate is an 'agrégé', for holders of the 'agrégration', or 'certifié', for holders of the 'CAPES'. This practice is illegal, as neither agrégation nor certification is among the requirements for recruitment defined by the ministry or fixed by the committees.

'Agrégation' and 'CAPES' are not academic degrees, but are national procedures, based on a set of competitive examinations through which holders of a master's degree can become teachers in the state secondary schools ('professeurs des lycées et des collèges.

[3]Every year, the Ministry of the Higher Education and Research defines a number of promotions for every category of staff, whether they be 'enseignant-chercheur' (EC, i.e. staff for research and education), teaching staff, or administrative staff. There are three types of promotion for the former: 'maître de conférences hors classe' (exceptional senior lecturer), 'professeur première classe' (first-class professor) and 'professeur classe exceptionnelle' (exceptional professor). Candidates eligible for these promotions establish a file that is assessed by the Conseil des National Universités (CNU), as well as by their institution. Half the promotions are decided by the CNU, while the remaining promotions are awarded by the EC members of the administrative council of the institution. In evaluating both teaching and research activity, the statutory obligations of an EC and engagement in administrative affairs are taken into account, although the accent is supposed to fall more heavily on research. Although the CNU promotion criteria are not clear, promotion at the national level is considered more prestigious because of the danger of cronyism, particularly in smaller institutions.

[4]The CNU took its present form in 1992. It is organized according to groups of disciplines and broad disciplinary sections. Each section has a number, which is why a lecturer may say that he or

discussions about criteria. The thorny question of individual evaluation has recently come up again, even if those doing a pilot study on individual evaluation are very careful to avoid pronouncing the word 'evaluation', and talk only about a 'suivi de carrière [monitoring of careers]' (AEF 2013a, Dépêche no. 187254). This 'suivi de carrière [monitoring of careers]' is also the term used by the most recent law on EC (Décret 2014-997, published on the 2nd of September 2014, see Article 21).[5]

2.2 Funding

Following the 2008 law, the Ministry of Higher Education started to implement a dual financing scheme. Eighty percent of state funding to universities—except salaries—is allocated on an 'activity basis', calculated by adding a 'teaching allocation' to a 'research allocation'. These are obtained by multiplying the number of students and tenured academic staff by blocked sums, defined by broad sectors of activity: life sciences, hard sciences and the SSH. The other 20 % rewards the relative efficiency in research and education, compared to that of the rest of the system. But not all the academic staff count in calculating the research allocation, either as activity or as performance; only the 'EC produisants', which roughly translates as active researchers, are taken into account. Thus, the assessment of the research activity became of paramount importance following the implementation of this scheme, and more so as an increase in the number of 'EC produisants' translates more easily into financial gains than any increase in the number of graduating students.[6] At the same time, universities received pressing invitations to increase their 'ressources propres' (own funding), especially by tapping into the competitive research funding resource. This reinforced the need, for the leading teams, to identify the most active and innovative researchers as well as the less-performing areas, either for allocating seed money and administrative support or for designing incentives.

(Footnote 4 continued)
she belongs to, for example, the 7th section (broadly, linguistics) or the 9th section (French language and literature). CNU membership consists of nominated members (one-third) and elected members (two-thirds). The latter is based on a list system, i.e., a dominance of trade union elected members. The CNU is in charge of the 'qualification', a certification system that allows certain doctoral degree holders to become candidates for senior lecturer positions, or senior lecturers to become candidates for professor positions. The problem is that the qualification process is very much a national barrier to the recruitment of foreign researchers in French academia (Sire 2012), and its maintenance is at odds with the ERA process, endorsed by French parliamentary representatives.

[5]This law is accessible under http://www.legifrance.gouv.fr/eli/decret/2014/9/2/MENH1418384D/jo/texte.

[6]In 2010, four more 'EC produisants' in a university brought in the equivalent of a medium salary, while teaching activity required 100 more students to obtain the allocation of the same sum. Calculations were made on the basis of the allocated budget of Université de Bretagne-Sud. Personal data of the authors.

2.3 The National Grant System

The creation of the Agence National de la Recherche (ANR) in 2005 radically modified the research units' access to funds and introduced a new actor to the evaluation sphere. For decades, in spite of an increasing concentration of researchers in the universities, 23.5 % of the budget for civil research was directed towards the Centre National de la Recherche Scientifique (CNRS[7]), while universities received less than 5.82 % (Giacobino 2005).

With the new funding scheme, discussed previously, and the allocation of substantial funding possibilities on a project basis through ANR programs, this unbalanced situation changed significantly. In terms of evaluation, mixed teams[8] (UMR) were no longer automatically recognized as top performers in research, even if, in practice, UMR benefitted from historical prestige when evaluated; at the same time, topics and teams not aligned to the CNRS priorities gained visibility and funding. New forms of evaluation were put into practice, closer to the peer review system used in highly reputable academic journals.

The biggest consequence of the new project-based funding procedure in the ANR grant system is the considerable change in outlook brought about by a radical change from a system in which teams had to work with the more or less generous amount allocated on a quadrennial basis, to a new system in which supplementary resources could be obtained through competitively funded projects. Unfortunately, this revolution only affects the SSH in a limited way, partly because of the long-lived reflexes of managing penury, partly because the available funds are much more limited than the investments in other scientific domains or in technological development. ANR priorities clearly favour scientific domains, which are considered as better contributing to industrial leadership and in responding to societal concerns. The situation is much the same at the regional level, where science policy priorities tend to mimic those established at the national level, which copy, in turn, the European ones, as proved by a recent Ministry discourse and by the subsequent policy document, entitled significantly, 'France-Europe 2020'.

Consequently, a new need for evaluation has arisen, in particular, one stemming from the SSH researchers themselves. The chronic underfunding of the SSH, and, more specifically, of the humanities, can be linked to an insufficient understanding and assessment of their impact outside academia. Impact does figure among criteria taken into account by AERES[9] and by ANR, both for the evaluation of the research

[7] Created in 1939 to bring together various research groups under a government-controlled institution, the CNRS is now the biggest research unit in France. Researchers are employed directly by the CNRS, which is divided into numerous disciplinary fields with associated institutes. There are also mixed teams that include university researchers, who also have a statutory teaching mission. Until the advent of the ANR funding agency, the CNRS had large block grants. It now must compete for project-based funding, and their research is evaluated by the AERES, something to which they have always objected.

[8] 'Mixed teams' gather personnel from the CNRS and from the universities.

[9] AERES was the national evaluation agency created at the time of the LRU reforms. It is now being replaced with an agency under the name of HCERES. See section II for greater detail.

units and for that of projects. However the ANR has no published guidelines for assessing impact, while those of AERES start from a very restricted understanding of the phenomenon. Impact tends to be considered exclusively in the form of patents or spin-offs, two types of results notoriously difficult to obtain when researching SSH topics. In this way, the major contribution of SSH research to the cultural industry is entirely neglected, while the role of SSH research in society is reduced to popularization conferences during specific manifestations ('Fête de la science' is explicitly mentioned), or to contributions to European laws and regulations. The list of impact types published by AERES is not a closed one, but its contours clearly manifest a lack of thorough examination of the matter. The time is, however, not far off when the question of impact will be in the spotlight, as proved by a recent report released by the 'Cour des comptes', the higher administrative court that oversees spending by public bodies and major French NGOs. The report pointed out the considerable budgetary effort made for the research since 2005 and questions whether the nation is getting a sufficient return on its money.

Whether for allocating funds, designing research strategies, supporting teams in their development, or demonstrating value for money, a more objective approach to research evaluation has become a major necessity in France over the last decade.

3 Current Practices and Levels of Evaluation

Unfortunately, in spite of the law and the need for modernized evaluation procedures, many institutions involved in research evaluation remain very vague about their criteria, in general, and about research excellence, in particular. At the same time, the process through which a percentage of the staff of an institution and/or individual persons are labelled as 'produisant' has been constantly questioned but still remains opaque. Finally, a great deal of confusion reigns about the peer review process.

The CNU has been repeatedly criticized over years for its opacity as well as for the weakness of its methodology (Garçon 2012). Because of the large number of applications to be assessed during the qualification or promotion processes, the review process in many sections cannot exceed 10 min/candidate. Furthermore, the relative weight given to the different elements of a CV varies widely from one section to another, and from one evaluator to another. It is to be noted that the way in which CNU members are selected does not require any competency in, or knowledge of, research evaluation, and is indifferent to the scientific merit of the candidates. At the same time, the CNU has no links with entities studying research evaluation, whether these be research laboratories or ministry-related agencies.

The AERES agency, created in 2007 to evaluate French Higher Education and Research Institutes at four levels,[10] never managed to fully implement individual evaluation of EC in spite of the importance of this level in the process of evaluating teams and institutions. The notion of 'EC produisant' does not appear in the official

[10]The teaching courses, the research groups, the doctoral schools and the institutions themselves.

document presenting the evaluation principles of a research unit (see AERES 2012a), but it does exist in a separate document which affirms that '[l]'un des indicateurs est une estimation de la proportion des chercheurs et enseignants-chercheurs "produisant en recherche et valorisation"' [one of the indicators [of the quality and influence of the research unit] is the estimation of the percentage of researchers and EC active in research and development] (AERES 2012b, p. 1).[11] Depending on his or her status, two to four 'first-class publications' ('productions de rang A') by period of four years are supposed to earn a researcher the 'produisant' label; patents, databases and other similar products are accepted as an equivalent. The problem is that there is no clear reason for the number of publications requested (why not one or six, for instance?), while the rigid classification of the outputs is inappropriate in many disciplinary fields (see infra).

Besides, the thorough characterization of journals and books, recommended initially by the AERES to define the channels of first-class publications, proved to be highly complicated. Even a simple glance at the produced lists, displayed on the AERES site, reveals tremendous problems. On one hand, these lists have evolved, following major criticisms from the academic community, from being graded league tables (A, B and C or international, national and limited reputation) to a collection of titles whose very inclusiveness[12] is at odds with the 'first-class publications' claim. On the other hand, such lists do not exist for many SSH domains, including French language and literature research, which is maybe the most striking example. What constitutes a 'first-class publication' depends, therefore, in many domains, on the expert's opinion. This opinion is formed without any reading of the submitted publications—as none were submitted during the assessment process, whether at the individual or the institutional level. To give but one example, the AERES guidelines claim that only collected works presenting a unified critical apparatus and a scientific deepening of the understanding of an original subject can be considered as 'first-class publications'. Unfortunately, the question as to how the experts are supposed to verify these requirements on the basis of a simple inclusion of a title (with its references) in the activity report generated by the research unit is not elucidated.

Conscious of these methodological problems, many visiting committees of the AERES do not release 'produisants' lists; nevertheless, the Ministry for Higher Education and Research, through its directorate for higher education, DGES-IP,[13] still applies very precise numbers per domain when it allocates funds to the universities—a somewhat magical operation if individual evaluation does not yet exist. Universities can propose corrections for these figures by signalling forgotten names. Thus, to a

[11] The notion of 'valorisation' covers, in France, all activities of development and technological transfer, but also social and organizational impact, etc.

[12] The former A, B and C journals were merged in the new lists, which are supposed to designate an 'academic perimeter'. At the same time, researchers can suggest new publication channels to be added to the list. It is not very clear if a further selection is operated among these suggestions (by whom?), or if any suggestion is automatically placed on the list.

[13] DGES-IP (Direction gènèrale pour l'enseignement supèrieur et l'insertion professionnelle) is the directorate of the Higher Education and Research Ministry directly responsible for contractual relations and the budget of French universities.

certain point, higher institutions operate as experts in evaluation, conducting their own analysis by applying, or not applying, AERES-based criteria to evaluate their academic staff.

4 DisValHum and IMPRESHS Projects

However unclear the future of the institutional research evaluation in France may be,[14] far too many questions occur in the day-to-day life of researchers and institutions that require clear answers for the problem to be ignored. Such questions include elucidating who is 'produisant' and who is not, what is to be considered as performance in research and what is not. Whether in France or throughout Europe, the need for clear responses to key evaluation questions is reflected by the growing popularity of Snowball Metrics[15] in the UK with its emphasis on informed decision-making. It is then significant that some major French research universities are also looking closely at this methodology so as to carry out foresight analysis. However, such indicators cannot work until there is critical research into dissemination practices, and this is particularly true in France. The evolution of the French higher-education system during the last years, as well as the external and the internal pressure, has opened the field for initiatives like the DisValHum and the IMPRESHS projects.

The starting point for the DisValHum and IMPRESHS projects is the realization that many of the problems observed in research evaluation in France stem from an insufficient—and, in certain cases, nonexistent—observation of the domain to be assessed and a lack of engagement with the stakeholders, principally the researchers themselves. The situation is even more acute for the SSH, where the preliminary analyses rarely go further than a few platitudes ('SSH publish more books than articles', 'SSH journals are not included in international databases', 'workshops and conferences are important in the SSH'), clumsily taken into account in the various evaluation activities. Both projects seek to contribute to filling this gap. Their intended benefits concern both SSH research, which suffers from its deficit of evaluation, and policymakers by proposing ambitious research policies at the national or institutional level. In general, and despite declarations to the contrary, French evaluation tends to be of a summative type, and is used primarily to allocate funds. Thus, to be effective, it requires a high degree of transparency, and hence faces the challenge of obtaining support from the academic community (Guthrie et al. 2013). Both transparency and support can only be obtained by improving current methodologies, and by listening to researchers at the ground-floor level, who often neither understand the means or the need for an evaluation process, and, generally find the process ill-adapted to their everyday existence.

[14] Under the new law (loi ESR, juillet 2013), the AERES has been replaced by a Haut Commission pour l'Évaluation de la Recherche et de l'Enseignement Supérieur (HCERES), whose organization and methods to date differ little from AERES, despite the recent nomination of a new director.

[15] http://www.snowballmetrics.com/.

Our specific aim is to provide the various evaluation performers (experts of the national agencies, or panels in the universities or research funding institutions, etc.) with objective information about dissemination practices in the SSH, as well as insight into how SSH scholars perceive this dissemination process. We also intend to contribute to the international effort of solving the numerous conundrums implicit in the research assessment of the SSH. This includes issues as the recognition of the specificities of the field, a position that can be seen as somewhat at odds with the claim that they must be taken as an integral part of the whole of scientific effort.

Both projects are supported by the Human Sciences Institute in Brittany (Maison des Sciences de l'Homme en Bretagne), and must be seen as two sides of the same research effort. For administrative reasons, the two projects were submitted for assessment under two separate calls, hence the different acronyms. They concentrate on the dissemination of the research results produced by SSH academics from the four Breton universities. Of the four universities, two, Brest and Bretagne-Sud, are multidisciplinary institutions. Of the two in Rennes, Rennes 1 is predominantly science based, but with a law and economics school, and Rennes 2 is exclusively arts, humanities and social sciences. The four belong to a cluster known as the Université Européenne de Bretagne and share common doctoral schools and joint research groups. Each university retains a degree of specialization in each of the fields studied.[16] For this study, we look only at the output of researchers from the three bigger institutions in Brest and Rennes. The initial results described in this paper refer to a language and literature research group in Brest, a history research group in Rennes 2 and two research teams in the law research group in Rennes 1. The reason for the last one is that this is a large research group with very different research themes. We shall be looking at the output from historical lawyers and specialists in civil law.

Our aims are:

First: to analyse the forms of dissemination, starting from what researchers do (as reflected in their CVs), and not from various preconceptions, based, in most cases, on practices in other fields or on the personal experience of the category designer. The idea is to avoid Procrustean solutions like those imposed by the official reporting, which asks all academics, irrespective of their field, to classify their production in fixed categories. Such categories are not necessarily clear, as there is, for example, no precise definition about what constitutes an international or a national conference. They are also incomplete. Among the most visible gaps are the lack of a category for critical editions or translations, frequent in the SSH, and also the nonexistence of categories such as databases or websites for scientific information. Reporting on forms of engagement with the wider public is also not taken into account, somewhat

[16]Since the first conception of this article, new developments have occurred that are changing relations between universities. The universities Rennes 1 and Rennes 2 were to set to become a single university, the University of Rennes, in January 2015. This project has now been abandoned. However, these two universities along with the two other Breton universities, and with three others from the neighbouring Pays-de-Loire region, will now become members of a new institution labelled 'communauté d'universités' (COMUE: community of universities). This will bring in a number of changes the consequences of which on both research and teaching are as yet unclear.

surprising in that this type of impact is supposedly to be evaluated. Categories can also be redundant, in that an invited conference paper can also be declared as an article in proceedings, or disparate when participation to PhD evaluation panels appears alongside authoring of books, without distinction as to the different nature of the exercise).

Second: to observe productivity curves and averages. As shown previously, an EC is considered to be 'produisant' if he or she has generated two pieces of work over a period of four years, but the reason for establishing such a threshold is not made explicit. At the same time, one of the most frequent criticisms of this requirement from French researchers is that a single-authored monograph should not be accorded the same weight as an article of a few pages in a journal, even if it is a highly reputed international publication.

Another aim in analysing productivity curves is to help render more objective value judgments conveyed in terms of 'average researcher' or 'impressively productive', etc. The CNU reports on individual applications frequently resort to such qualifications, whilst there is no clear definition of the benchmarks taken into account.

Third: to analyse collaborative research practices, as reflected by the disseminated products. The objective is principally to study frequency and forms of co-authorship in the SSH disciplines. We are particularly interested in the identification of transdisciplinary and international cooperation of Breton researchers.

Fourth: to observe channels of dissemination, mainly publishing houses and types of journals favoured by SSH scholars in Brittany, but also channels for oral dissemination. The channels will be further characterized by using objective descriptors, such as presence in international databases or not (for journals), and international distribution or not (for publishing houses), etc. Once again, the aim is to start from the bottom and not from top-down defined lists.

Fifth: to understand the reasons motivating the choice of these channels, as well as of the publication formats adopted. On one hand, we try to understand if maximising the scientific impact constitutes a preoccupation of Breton researchers when they publish; on the other hand, the requirement is to track their ideas about how and why they interact with the wider public.

To fulfill these aims, our first concern has been to build a research products database. A preliminary study was conducted on a small number of CVs published online by researchers in French literature, linguistics, history and law, since these are the domains covered as a priority by the projects. The study was meant to identify the types of research products created by SSH researchers, whether as written material or not. This pilot study was completed by a study of categories selected by various information systems, such as CRISTIN in Norway, VABB-SSH in Flanders (Belgium), or RIN in the United Kingdom. These categories were then tested on a larger scale with the help of the students from the Master of Digital Humanities in Université de Bretagne-Sud. These gathered as many CVs of Breton researchers as possible in the considered domains, helped refine the categories and the structure of the database, and provided the first statistical calculations. For all these reasons, the number of categories finally selected is much larger than that of any of the considered CVs; the differences have proved interesting in themselves as both the focus groups and inter-

views have demonstrated that the non-inclusion of an item in a CV does not translate necessarily into the nonexistence of such a product in the activity of the considered researcher. Its absence is merely a form of self-censorship, sometimes related to the perceived expectations of the external evaluation bodies.[17] In such situations, top-down criteria imposed without a preliminary study of the ground clearly result in a loss in information and, moreover, of potential arguments for demonstrating the social impact of the SSH.

The database, which is currently under development, is organized into four main sections: books, articles (whether in journals or collected works), other written material and non-written material. A comparative list in the appendix of this article shows the types of products it covers, compared to those taken into account by the UK RIN analyses. Authors are characterized by their affiliation (institution and research unit) and by domain (CNU section); a CNU section is conventionally attributed to foreign researchers who cooperate with Breton academics. This has the disadvantage that CNU sections are extremely broad, but does mean that precision can be reached a posteriori using a study of dissemination types and focus group output rather than imposing further subdivision. Co-authorship characterization allows for social network analysis, which will be confronted with a similar analysis conducted on institutional contacts of research units. Moreover, geographical information is available (city and country of authors, and country of publication), making it possible to map visualizations of research contacts.

The basic information as to who, what, where and when is entered in the database. In each section, broad classes of channel and type are used. These remain sufficiently broad to handle all the data included in an individual CV. Only when the database has reached a reasonable size will work start on trying to classify the input in more detail. This is particularly the case with the 'other' section, which contains a rich variety of outputs that probably have a wider social impact than those in a standard CV. As the aim is to get an overall picture of different research groups and different disciplines, we are not concerned with individual researchers, but will look at individual cases when necessary.

The highly time-consuming operation of establishing a database was necessary because information about the SSH production of the researchers in our perimeter is incomplete, unusable, or inaccessible. The institution in charge of producing indicators for research and innovation in France, namely, Observatoire des Sciences et des Techniques, covers the SSH production only on an exceptional basis (Filiatreau 2010) and in doing so relies on the Thomson-Reuters database. If this choice is justified by the benchmarking purposes of the report, it proves clearly inadequate to answer the practical questions listed previously.

As a responsible scientific organization, the CNRS is fully aware of the need for quality checks. Consequently, it has put into place its own internal survey, called RIBAC (Dassa and Sidéra 2011). Unfortunately, this information system concerns

[17] Interview with two historians: 'No, I would not put this on a CV, it is not important enough, and in any case not evaluated by AERES.'

only the CNRS, and despite talk of imposing it on universities, it is more than probable that the current government will abandon the idea. This is not altogether a bad thing, as it is far from certain that RIBAC categories are adapted to the EC. The typology of research products also tends to be very restrictive. A full comparison with other databases has not been possible as yet as the CNRS has not made access to the structure publicly available. It is however clear that the non-written material, as well as research reports of all types and forms, are underestimated, which does handicap impacts studies as that envisaged here.

A national database of research output, HAL[18]—*Hyper articles en ligne*—that collects research outcomes from French researchers, has existed since 2006 ('HAL: Accueil', 2013) as an open repository. HAL SHS, a specific site for the SSH managed by the CNRS, is used by researchers wishing to put data online. This is not compulsory and, given the extreme lack of user-friendliness, many researchers do not submit; thus, its coverage is only partial. Data can be exported in csv format, but an attempt to nourish our database showed that a great deal of what was necessary, coupled with the non-compulsory nature of the repositories, meant that such an operation is not feasible in the immediate future. The imposed categorization also introduces a further difficulty, as researchers either leave out aspects of their work or misinterpret the categories. Technological changes, as well as policies of major research groups, are rapidly rendering the HAL database redundant.

Lastly, research group activity reports, established for the quadrennial evaluation performed by AERES, have appeared unsatisfactory as evaluation research tools. Not only do many laboratories not publish these reports, but when the reports do exist, the laboratories list only the productions of the previous four years. Inside each report, bibliographical references are far from unified, rendering impossible an automatic translation of the information into our database.

Parallel to the building of the database, which is still in the long phase of manual data entry, a series of group interviews with SSH scholars from various research units in Brittany are being conducted. Appendix 2 lists the questions asked. Recorded interviews are supplemented with notes taken in parallel, which are also transcribed and coded using Atlas.ti.[19] These interviews are intended to help refine the types of products included in the database, and, above all, to retrieve 'natural' hierarchies made between forms and channels of dissemination, to understand who Breton scholars consider when they disseminate their research (the 'ideal reader') and to identify their partners from outside academia. A further aim is to build a typology of publishing outlets and to discover what their purpose may be from the scholar's point of view.

[18] http://hal.archives-ouvertes.fr/.

[19] http://www.atlasti.com/.

5 Initial Outcomes

Following initial focus groups and observations of the database, one thing is very clear: there is an enormous mismatch between what goes into CVs, what is accepted by AERES and how researchers see the dissemination of their research. The interpretations of the AERES classification codes vary widely, between those researchers who put in all their activities, no matter how trivial, and those who leave out activities such as speaking to the general public—considering that the CV deals only with 'research'. This is summed up neatly by an English language specialist who asked whether pedagogical dissemination (course material) could be treated as research dissemination: 'Est-ce que la dissémination pédagogique compte, est-ce que les cours comptent?' [Does pedagogical dissemination count, does teaching courses count?] This is a delicate question to ask in that many SSH scholars write material for the French competitive exams governing entry into the secondary school system as teachers. This is output, but not necessarily considered research, as it is, essentially, a compilation of material to be absorbed by candidates. Textbooks in law do, however, carry a certain prestige.

Preliminary conclusions show that impact concerns vary greatly among the SSH scholars. The representatives of socioeconomic and psychology disciplines are more attentive to selecting publication channels and forms according a career plan, or have a genuine expectation to attract the attention of best international partners in their disciplines; these representatives also are very attentive to the requirements of AERES. Scholars in literature and languages, however, generally lack a clear dissemination policy. This observation is also supported by the fact that the latter clearly find difficulty in defining what can be considered an international publishing house or an internationally reputed journal. Two English-language specialists were very clear about the necessity of publishing in English, while recognizing a certain confusion about the value of certain publishing houses. As one said:

> une tendance chez les anglicistes français de publier chez Cambridge Scholars Publishing, la nouvelle maison d'édition à Newcastle, donc on voit bien qu'il y a pas mal de colloques anglicistes qui sont publiés là bas, et autres d'ailleurs, j'ai publié deux là bas donc je trouvais ça très bien, et dernièrement j'ai appris que des chercheurs anglais, eux, considèrent que c'était leur Harmattan, c'est leur Harmattan.
> [A tendency among French English researchers is to publish with Cambridge Scholars Publishing, the new publishing house in Newcastle, so we see clearly that quite a few conference (proceedings) of English specialists are published there, and others elsewhere, I published two there, so I found it quite good, but lately I learnt from English researchers that they consider it their Harmattan, it is their Harmattan.][20]

The interesting fact is that the researcher in question has published books only in the two outlets, but is now doubting whether this is a good thing or not. Whereas in evaluations, the status of publishers is not currently a discriminatory factor, the scholars are clearly sceptical about the pay-to-publish sector.

[20]Harmattan is not greatly considered by 'serious' French researchers as its reputation is of a pay-to-publish outlet with no real quality control.

There was also a tendency to see the English-speaking journals as having higher standards and better review practices, with one scholar very impressed by the facilities offered when asked to review for a major American journal. This researcher insisted on journals being demanding and using the double-blind review, something found in few journals in France in English studies. Her colleague, however, insisted that more local journals should not be written off as 'un cahier local n'est pas forcément de mauvaise qualité, de qualité inférieure, alors qu'on peut avoir des articles de qualité excellente dans une revue locale.' [A local journal is not necessarily bad quality, inferior quality, you can have very good articles in a local journal.] He also pointed out that such journals more readily publish the work of junior researchers, allowing them to get recognition.

Best practices are mainly identified, in the humanities group, as being those recommended by the ministry, less because these are genuinely considered more efficient in developing research, but clearly because 'it is what is expected' (interviews with historians and with language specialists). The influence of evaluation, however, is present in the socioeconomic and psychology group, too. One economics researcher, who professed to having no clear dissemination strategy, found herself classed as non-produisant because of the restrictive list imposed in her field.

Another problem identified by focus groups as weighing on the research and dissemination practices in the SSH is the themes a research group in the humanities imposes on itself to meet national evaluation requirements. These last only for the four years of a contract, and create a straitjacket for any researcher who is thematically or discipline based. This thematic issue is a particularity of certain humanities groups and is imposed to provide a semblance of homogeneity where heterogeneity dominates. Research groups in languages often bring together researchers from different languages and different periods of interest. They are also broadly divided into researchers in literature, cultural studies and linguistics. The third one is largely grammar, because linguists themselves are in a different CNU section and mostly in different research groups. Thus, whereas a scientific research group may be specialized in, for instance, polymers, a language group will give itself a theme, such as 'great men', that is supposed to be a focus point for the four-year contract with the state. This, obviously, requires a fair bit of non-productive acrobatics from the higher-level researchers who have carefully developed a particular area of expertise. As one researcher said:

> la place des SHS est telle qu'on est la 5ème roue de la carrosse donc on nous demande de nous agréger à des champs de recherche et des thèmes de recherche qu'on a pas choisis, à [name of university] c'est ça, si on veut être un peu visible, et c'est un problème de [name of research group] par rapport aux autres labos, même si c'est un peu pareil, si on veut être *visible*, il faut, localement, qu'on *réponde à des appels qui ne sont pas naturellement dans notre champ*. Donc, ce qu'on fait quelquefois avec des déceptions parce qu'il n'y a pas de publication par derrière parce que justement c'est trop large...
> [The SSH are excess to requirements, so they ask us to group our areas and themes of research that we have chosen, in [name of university] it is just that. If you want a minimum of visibility, and it's a problem for [name of research group] in relation to other research groups, even if it's a bit the same. If you want to be visible, you must, locally, answer calls for tender which are not naturally in your field. Thus, it is what we do, but sometimes with regret as there are no publications forthcoming as the theme is too wide...]

Table 1 Output types across four disciplines in percentages

	Civil law	Law history	History	Literature
Journal	66	32	22	37
Book chapter	18	12	23	30
Encyclopaedia	0	7	3	4
Proceedings	6	21	17	4
Press	0	11	3	0
Miscellaneous	0	6	14	0
Books	10	11	19	26
Total	100	100	100	100

Note Some columns might not sum to 100 % due to rounding

Another interesting observation can be made about the contrast between the practices and perceptions of engagement with non-academic representatives. The discourses present this activity as a one-way process, in which the Researcher transmits Knowledge to a passive Receiver; the idea of a possible influence of stakeholders on one's own research triggered vivid reactions in some cases. But examples cited during the discussion proved that outside academia, stakeholders are, at least in certain cases, valuable collaborators as much as passive receivers. We try to collect precise identifications of these partners to conduct cross-interviews in the manner of those recommended by the ERiC method.

In quantitative terms, the image about SSH publication coming from the database is, for the moment, as in Table 1.

The dominance of books and book chapters is clear in history and literary studies, but these figures must be treated with care. Published chapters may be, in fact, published proceedings, something that is rarely declared in English, but is always noted in the sectors of law and history. The AERES classification lumps together books and book chapters and groups papers in proceedings with either national or international conferences. It is possible that the book section is considered more prestigious by English specialists, hence the preference to declare a chapter to a proceedings article. The absence of certain items may simply show that these disciplines do not deem such outputs as worthy of mention in a CV. The very high percentage of journal publications in civil law also requires caution, because many of these may be short legal commentaries. While we are attempting to track the length of papers, not all CVs give full references. Obviously, miscellaneous publications and books will require close attention. However, what these statistics do show is that simplistic evaluations based on declared data do not give a genuine picture of the complex dissemination patterns across disciplines.

Some factors are becoming clear. Each discipline has its own publication patterns and its own channels, with no similarity across even legal history and history. To date, there is little sign of interdisciplinarity or internationalization. The rule is single authorship for papers and books, except for proceedings and collected works that tend to be co-edited. The exception to this rule was a specific case in law, relating to

scientific and medical fields, but the co-author was another lawyer and not someone from outside the discipline. Most publications are in French, and in France, although there are also major legal publications in francophone Belgium.

The regional university press, the Presses Universitaires de Rennes, is the main publisher for books in history and, to an extent, in literary studies. This publisher has built a strong reputation in regional history and is an obvious publisher for collected works and proceedings. Civil law tends to have its own highly specialized publishers.

As research groups can be fairly homogeneous, it is interesting to look at the 'anomalies'. To date, three examples stand out: a researcher in languages publishing in high-impact journals in a research group that tends to remain at local or national levels; a researcher in history whose subject area, piracy, has strong popular appeal and, therefore, gives numerous radio broadcasts; another is a researcher who has a particular interest in one legal field that links him to a particular form of local court. Other broad cross-disciplinary tendencies also are beginning to appear, as language researchers closer to the visual arts, notably those studying cinematographic productions, have dissemination patterns different from those more concerned with producing scholarly editions. As one researcher said:

> je suis un peu partagé en fait puisque je fais de l'édition de textes, l'édition de textes se prête assez mal à la communication; l'édition de textes a plutot tendance à la publication directe. [I am of two minds about this in fact as I have worked on critical editions. Critical edition work is not adapted to popularization; critical editions tend more toward direct publication.]

6 Conclusion

The Loi LRU caused a sea change in French research by bringing in internationally certified evaluation procedures. The modification of that law by the Loi ESR watered these procedures down, at the demand of trade unions and a vocal section of the research community. As a result, evaluation procedures that might allow for informed decision-making and foresight activities are now far off. The situation has become more, rather than less, confused, leaving opaque recruitment and promotion practices in place, and not really providing, the tools for a better-informed monitoring of research. Existing systems may work more or less well in some disciplines, where internationalization and, therefore, international benchmarking of research are strong, but this is not the case in the SSH.

Despite resistance in some quarters, greater attention to quality criteria is inevitable as France remains a major player in international research in all fields, including those of the SSH. Current research is leading to better bibliometrics and an understanding of research practices and dissemination. However, although common terminology is developing, the interpretation of that terminology will inevitably remain anchored in national practice, needs and research traditions. Thus, any attempt at benchmarking must be based on an analysis of the situation in each large field and in each country. An overall picture is needed before indicators are imposed. This global picture is

what IMPRESHS is setting out to achieve, starting from one region of France with the aim of launching a larger study across university research in the SSH across France.

There are numerous threads to be followed before a clear picture of French SSH research can be obtained. What is already clear is a very complex situation dominated by national parameters. What this means in practice is that a neutral study based on bottom-up procedures will encourage greater understanding of output types and the motivations of researchers behind their choice of those output channels. Only then will it be possible to equate research outcomes with possible societal impact. SSH research covers a broad spectrum of activities, outcomes and impacts. Understanding this is the key to better quality research evaluation criteria and, therefore, better research. The wealth is in the variety; IMPRESHS aims to help bring about a better understanding of this variety.

References

AEF. (2013a). Comment les sections 34 et 63 du CNU experiment l'évaluation individuelle des enseignants-chercheurs (Dépêche no. 187254). Paris. Retrieved from http://www.aef.info/enseignement-recherche.

AEF. (2013b). Shanghai 2013: 'La France seule est responsable du résultat mitigé qu'elle obtient', estime Edourad Husson (Dépêche no. 186447). Paris. Retrieved from http://www.aef.info/enseignement-recherche.

AERES. (2012a). Critères d'évaluation des entités de recherche. Le referential d'AERES. Retrieved from http://www.aeres-evaluation.fr/content/download/17661/271795/file/R.

AERES. (2012b). Critères d'identification des chercheurs et enseignantschercheurs 'produisant en recherche et valorisation'. Retrieved from http://rbdd.cnrs.fr/spip.php?article141.

CP-CNU. (2012). Motion contre l'évaluation individuelle des enseignants-chercheurs. Retrieved from http://evaluation.hypotheses.org/1683.

Dassa, M., & Sidéra, I. (2011). Présentation du système d'information RIBAC du CNRS. Un outil de valorisation de l'activité des acteurs de la recherche en SHS. In Evaluation des productions scientifique. Des innovations en SHS? Actes du colloque de Paris, Juin 2011 (pp. 19–30). Retrieved from http://journalbase.sciencesconf.org/conference/journalbase/eda_fr.pdf.

Estermann, T., Nokkala, T., & Steinel, M. (2011). *University autonomy in Europe II — the scorecard*. Brussels: European University Association.

European Commission. (2010). Feasability study for creating a European university data collection. Final study report (Contract No. RTD/C/C4/2009/0233402). Research-Directorate General of the

European Commission. Retrieved from http://ec.europa.eu/research/era/docs/en/eumida-final-report.pdf.

Filiatreau, G. (Ed.). (2010). *Indicateurs de sciences et de technologies. Edition 2010. Rapport de l'Observatoire des sciences et des technologies*. Paris: Economica.

Garçon, F. (2012). *Le dernier verrou*. Paris: The Media Faculty.

Giacobino, E. (2005). Le nouveau paysage de la recherche institutionelle en France. Retrieved from https://www.dgdr.cnrs.fr/daj/archiv-actus/2006/decembre%202006/Docs-dec06/Giacobino06.pdf.

Guthrie, S., Wamae, W., Diepeveen, S., Wooding, J., & Grant, J. (2013). Measuring research: a guide to research evaluation frameworks and tools. Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/pubs/monographs/MG1217.html.

Sauvons l'université. (2012). Contre l'évaluation quadriennale des enseignants-chercheurs: signez la pétition! Retrieved from http://www.sauvonsluniversite.com/spip.php?article5602.

Sire, B. (2012). Nous voulons partout les meilleurs. *La Dépêche*. Retrieved from http://www.ladepeche.fr/article/2012/11/12/1486615-brunosire-nous-voulons-les-meilleurs-partout.html

Part V
The 'Forschungsrating' of the German Council of Science and Humanities. Risks and Opportunities for the Humanities: The Case of the *Anglistik/Amerikanistik* Pilot Study

Rating Research Performance in the Humanities: An Interim Report on an Initiative of the German *Wissenschaftsrat*

Christian Mair

Abstract The author, a professor of English linguistics at Freiburg University, was a member of the German Council of Science and Humanities (*Wissenschaftsrat*) from 2006 to 2012 and, in this capacity, was involved in this advisory body's rating and assessment activities. The present contribution focusses on issues arising in the rating of research output in the humanities and is informed by his dual perspective, as planner and organizer of the ratings undertaken by the *Wissenschaftsrat* and as a rated scholar in his own discipline, English and American Studies.

Over the past decade, rankings—whether home-grown or international—have had a profound impact on higher education in Germany, although the way in which they are being used tends to reveal a degree of tactical short-termism if not downright cynicism. Institutions which come out on top rarely question the procedures by which the welcome result has come about, but are happy to make the most of the free advertising provided. Those not placing so well do not take the result as a motivation for systematic self-study, but rather look to convenient quick fixes which, they hope, will enable them to move ahead in the league tables the next time around.

Within the academic community, rankings have become an informal mechanism of reputation assignment which is not entirely unproblematical but which—at least so far—has had few tangible consequences in terms of structural reform or strategic planning. In wider society, rankings may have some influence on students' and parents' choices of institutions and programmes, though there is as yet no evidence that they are a crucial factor in such decisions, which is probably not a bad thing, either, as the criteria which rankings are based on usually have no very direct bearing on the needs of first-year undergraduates.

In this situation, the German Council for Science and Humanities (*Wissenschaftsrat*), decided to carry out an analysis of the extant rankings in 2004. Its main finding was that the systematic, comparative and often quantitative assessment of research performance had come to stay, but that the methods and criteria employed by the

C. Mair (✉)
English Department, University of Freiburg,
Kollegiengebäude IV, Rempartstr. 15, 79085 Freiburg, Germany
e-mail: christian.mair@anglistik.uni-freiburg.de

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_15

various rankings were usually not fully transparent and that, moreover, the relevant academic communities had little say in how they were framed (Wissenschaftsrat 2004). The *Wissenschaftsrat's* suggestion for improvement was to develop a rating system in which research output in a particular field would be evaluated comparatively on the basis of criteria developed in consultation with the relevant research community.

As such a rating exercise involved substantial preparation and considerable investment of labour from all parties concerned, pilot studies were deemed essential. The concept was first put to the test on a nationwide scale in the fields of chemistry and sociology—and proved generally workable in both fields, despite their very different objects and methods of investigation (Wissenschaftsrat 2008). Encouraged by this, in 2008 the *Wissenschaftsrat* decided to carry out two further pilot studies, which were supposed to conclude the test phase, and then make the new instrument available on a large scale. The disciplines selected for this second phase of pilot studies were electrical engineering and informatics, on the one hand, and history, on the other. While the engineering pilot was successfully completed in June 2011 (Wissenschaftsrat 2011), the history pilot ended in a deadlock between the *Wissenschaftsrat*, representing the advocates of measuring research output in the humanities, and the Verband deutscher Historiker (Association of German Historians), representing the research community to be rated. As some of the debate was conducted in the culture pages of major national broadsheets, it generated an amount of publicity which, at least for the *Wissenschaftsrat*, was not entirely desirable in such an early phase of testing the new instrument.

On the other hand, it is the high profile that this episode gained which makes it instructive and interesting beyond its immediate academic-political context. In the remarks which follow I shall therefore take it as a starting point for a discussion of the particular difficulties—objective and subjective—surrounding the comparative measurement and evaluation of research output in the humanities and to present the *Wissenschaftsrat's* line of argumentation on this important issue.

In principle, there is no reason why a rating exercise as envisaged by the *Wissenschaftsrat* should be offensive to scholars' sensibilities in the humanities. After all, in its critique of the current situation, the *Wissenschaftsrat* points out the superficiality and lack of transparency of most existing *rankings* and makes the point that any instrument used to measure research performance needs to fit the discipline it is applied to. The *ratings* which the *Wissenschaftsrat* (Wissenschaftsrat 2004, pp. 33–43) suggests as the appropriate alternative are supposed to:

- be conducted by peers who understand the discipline they are evaluating,
- apply criteria specific to the field being evaluated,
- evaluate research output in a multi-dimensional matrix rather than a simple rank list,
- differentiate between achievements of individual 'research units' representing the field at a particular institution.

The last-mentioned criterion in particular should be welcome to scholars in the humanities, who define their research agenda very much as individuals and would resent their achievement to be levelled into departmental averages in a rating exercise.

While the preparation for a rating may involve a certain degree of nuisance and the rewards may be uncertain, the overall design features should find a sympathetic audience among humanities scholars. As a principle, informed peer review is accepted in the humanities as in other academic fields. It determines what gets published or who gets selected for positions, and at conferences or similar forums humanities scholars certainly enjoy the opportunity of showcasing their work and benefit from constructive criticism and advice extended by peers as much as anyone in academia.

What then is the cause of the hostility towards the rating exercise articulated by German historians (or at least their spokespeople in the association)? At least in part, I would contend, the conflict was due to a communication problem. Rankings and ratings, including the *Wissenschaftsrat's*, tend to be presented in a discourse of administrative control and neoliberal new public management which makes many scholars in the humanities suspicious from the very start. Their main experience with this discourse has so far been gained in the defensive rather than the offensive mode. Strategic planning of research has been experienced as increasing regimentation, increasing pressure to produce largely bureaucratic documentation and—in the extreme case—withdrawal of personnel and resources. That the humanities stand to gain from strategic planning—for example through improving career prospects for young scholars or claiming their due place in expensive digital infrastructure projects—has been less obvious by comparison. In this situation, any type of ranking or rating is thus likely to be considered as part of an unhealthy trend towards the bureaucratization, commercialization and commodification of higher education.

Let me briefly illustrate the type of miscommunication I have in mind with one of the *Wissenschaftsrat's* own formulations. Both internally and in several external presentations it has defined the purpose of the rating exercise as 'Unterstützung der Leitungen bei strategischer Steuerung durch vergleichende Informationen über Stärken und Schwächen einer Einrichtung' [supporting administration in its strategic planning by providing comparative information on strengths and weaknesses of a unit] (see Wissenschaftsrat 2004, p. 35, for a published version). Putting things in this way is certainly not wrong, but—in view of what has been said above—clearly not the best way of enlisting the support of the scholars whose participation is required to make the exercise a success. While the formulation allows us to infer the threats that may accrue from under-performance, it is not very explicit on the rewards to be derived from co-operation, both in terms of a particular field and the individual researcher. Researchers in the humanities are generally individualists and therefore sceptical about higher-level strategies of promoting or regimenting their scholarly creativity. They are competitive but not necessarily in the corporate sense of championing their institution. Successful teams are more likely to be composed of scholars working in different places than of colleagues belonging to the same department.

In his public debate with the *Wissenschaftsrat*, Werner Plumpe, the renowned historian and president of the German Historians' Association at the time, emphasizes exactly these points in his critique of the proposed rating (Plumpe 2009). Quan-

tification and standardization, he claims, may suggest the simplicity that political decision makers in university administration and higher-education bureaucracies crave, but this simplicity is a spurious illusion [in his own words (Plumpe 2009, p. 123): 'teilweise quantifizierte, immer aber parametrisierte Informationen für politische Diskussions- und Entscheidungsprozesse, die gemessen an der Realität des Faches unterkomplex [sind]']. An even bigger illusion is the assumption that success in research is the result of stimuli set in the system or advance planning of other kinds ['Illusion, Wissenschaft lasse sich parametrisch durch das Setzen bestimmter Anreize steuern'] (Plumpe 2009, p. 123). According to Plumpe, a standardized rating is not merely useless but counter-productive, because it encourages scholars to focus on meeting the targets of the system rather than the often different standards of professional integrity and scholarly excellence ['Herausbildung und Verfestigung strategischer Verhaltensweisen, die zumindest in den Geisteswissenschaften die akademische Kultur zerstör[en]'] (Plumpe 2009, p. 123). In short, the field of history does not owe it to itself or anyone else to take part in such a problematical project:

> Das Fach habe es aber weder nötig noch sei es im eigenen Interesse verpflichtet, die gefährlichen Illusionen der derzeit politisch hegemonialen Strömungen zu bedienen.
>
> [Neither self-interest nor external necessity forces the community to pander to the current hegemony's dangerous illusions.] (Plumpe 2009, p. 123)

As we see, the opposition is comprehensive and formulated with considerable rhetorical investment. A compromise between the Historians' Association and the *Wissenschaftsrat* was not possible. While the opponents of rating could claim a victory and were in fact heralded as champions of academic freedom in some of the press reportage, the *Wissenschaftsrat* found itself in a bit of a fix. In an atmosphere thus charged, it would have been futile to just move on and approach another field in the humanities to enlist its co-operation. The way out of the impasse was the creation of a working group bringing together a wide range of scholars in the humanities—from philosophy through literature and linguistics all the way to area studies, including the *kleine Fächer*, highly specialized areas of enquiry such as cuneiform studies or Albanology, which in the German system are frequently incorporated as micro-departments consisting of one professor and one or two lecturers or assistants. This interdisciplinary working group was expected to assess the suitability of the *Wissenschaftsrat's* proposed rating to the humanities and suggest modifications where it held them to be necessary.

The present author was privileged to be part of this working group and can testify to the open atmosphere of discussion which made all participants aware of the wide range of research methods and theoretical frameworks found in the contemporary humanities. Most members of the group eventually (though not initially) accepted that rating research output according to the *Wissenschaftsrat's* model was possible in the humanities, might even have beneficial side effects for maintaining and developing quality in the individual fields, and be a means of securing the humanities' general standing in the concert of the other disciplines. Intense disputes, however, arose every time concrete and specific standards of evaluation had to be formulated. Early drafts

of the recommendations contained fairly contorted passages on the relative merits of the traditional scholarly monograph as against the co-authored paper in a peer-reviewed journal, on the need to encourage publication in English while safeguarding the continuing role of national languages as languages of scholarly publication, and so on. About half way through the proceedings, participants realized that the best way to solve these issues for the time being was to defer them, i.e. to state the problem but to expect the solution to emerge from subsequent discussions in the individual research communities concerned. The recommendations thus grew slimmer, but improved from meeting to meeting as discussants realized that they had to aim for a mid-level of abstraction and leave the concrete fleshing out of standards to the discipline-specific experts. In a slight departure from existing *Wissenschaftsrat* rating conventions, the following three dimensions of evaluation were proposed (Wissenschaftsrat 2010, p. 20):

- Forschungsqualität [quality of research]
- Forschungsermöglichung [activities to enable research]
- Transfer von Forschungsleistungen an außerwissenschaftliche Adressaten [transfer of research achievement into non-academic domains].

To accommodate possible slower rates of maturation of research results and slower dissemination and reception, the standard five-year cycle of assessment was extended to seven years. It will be a major challenge to rating exercises based on these recommendations that qualitative measures were prioritized over quantitative ones. Thus, for the assessment of research quality, each 'research unit' will be asked to submit the five publications from a relevant seven-year period which are considered most important. The technical designation 'research unit' is intended to make possible reporting at a contextually appropriate level intermediate between the individual researcher and an institutionalized administrative unit such as a 'department' or an 'institute'. In a traditional German humanities context, this level would typically be understood to be the '*Professur*', i.e. the professorial '*Lehrstuhl*' or chair comprising the professor and his or her assistant(s). Discussions in the working group suggested that some academics would be quite happy to dispense with this intermediate layer in practice and submit five publications per professor, thus defining the relevant unit of documentation as the individual advanced researcher. Clearly, those responsible for the next pilot study will take the opportunity to clarify this contested issue against the background of their discipline.

The most salient feature of the proposed procedure when compared to rating in the natural sciences is that quantitative information, such as number of publications, will play an ancillary role only. This is justified, though, in view of the fact that standard quantitative indicators such as impact factors or citation indices are only marginally relevant in the humanities. One additional dimension of evaluation which it was judged necessary to include in rating research quality similarly defies quantification, namely a researcher's scholarly reputation. In view of reputation's auratic and intangible nature, those members of the working group who would rather not have included it as a criterion will probably take consolation from the fact that it will not have the same importance for all disciplines and certainly not for all individuals.

One of the more convincing ways of measuring reputation was considered to be taking note of the award of prestigious research prizes, such as the German Research Foundation's (DFG) Leibniz Award. Those who advocated considering reputation emphasized that it was not something which lapsed in the seven-year time-window relevant for measuring performance.

The term *Forschungsermöglichung*, not conventionally established, was used as a cover for activities which did not necessarily result in research publications by the principal investigator, but promoted research activities in a wider sense. Typical examples would include contributions to the development and maintenance of important research infrastructures, such as digital text archives or linguistic corpora, acquisition of external funding for research teams providing career opportunities for young researchers, etc. The distinction between the two dimensions of quality and enabling was felt necessary as (a) the mere fact that research in the humanities was funded by external grants did not mean that it was necessarily of high(er) quality and (b) across virtually all humanities disciplines the individual researcher was considered to be in a position to produce first-rate research unaided by teams or expensive infrastructure.

Transfer was expected to take forms appropriate to the individual disciplines, ranging from involvement in exhibitions and museums (art history) via in-service teacher training (foreign languages) to consulting activities (philosophical ethics).

As I briefly hinted at above, it is also very interesting to note the points on which the general recommendations are silent. They do not pronounce on the relative merit of different formats of publication, such as the article in a refereed journal, the article in a volume of conference proceedings, or the monograph. What constitutes an effective or prestigious place of publication is a question for individual disciplines to decide, and linguists' answers will certainly be different from historians'. Personally, I found this attitude of tolerance a little too generous as I am convinced that publishing cultures in all humanities subjects are in a state of transformation. The bad news is that too much is published, and too little is read, but the good news is that in many disciplines informal hierarchies of publishing outlets are emerging which may not be as rigorously enforced as the impact-factor-based reputation hierarchies in the natural sciences, but nevertheless provide orientation to scholars as to where they should strive to publish in order to ensure a maximum audience for their findings.

Another important point the recommendations are silent on is language(s) of publication. Research in the humanities is informed by culture- and language-specific traditions of academic writing, and most scholars in the humanities consider multilingualism an asset in their practice. Arguably, however, our current practices and the academic language policies currently advocated do not promote the most intelligent kind of academic multilingualism in the humanities. Knee-jerk reactions to combat the spread of English and promote academic publication in the respective national languages will usually find favour with the public but are potentially harmful. Consider the following example. A German specialist on the Portuguese language with interesting results on the specificities of Brazilian as against European Portuguese has three theoretical options: (a) publish the findings in German and guarantee dissemination in the peer group most relevant to his or her career, (b) publish in Portuguese

and thus reach the speakers of the language itself, and (c) publish in English to reach the global community of experts on Portuguese. Each of the strategies will potentially lose some readers: people interested in the Portuguese language not reading German (a), general linguists with no particular fluency in Portuguese (b), and people interested in the Portuguese language unable to read English (c). To compound the issue further, the strategy adopted will partly determine the use made of the findings. Publication in German or English will attract additional readers with no specific interest in Brazilian Portuguese as such, but with an interest in the standardization of pluricentric languages in general (e. g. Canadian English vs. United States English, or convergence and divergence between Standard German as used in Austria, Switzerland and Germany). Publication in German may lead to more intensive popularization of the findings among the small group of German-based teachers of Portuguese as a foreign language. These are merely some of the legitimate motivations which guide writers in the choice of languages for publication.

Conceivably, publication in German or Portuguese might also be employed for less than honest purposes, for example as a convenient method to get away with the unreflected use of traditional philological methods by insulating one's work from potential criticism articulated by a now largely English-speaking international community of 'modern' general linguists. But then again, this very Anglophone global linguistic establishment could be accused of cultural imperialism, which for example indeed manifests itself often in refusing to recognize important innovations until they are made available in English. Given the complexity of the politico-linguistic terrain in the humanities, researchers need more support than they are getting now. For example it is much better to fund the translation of excellent work published in languages other than English than to force researchers who are not entirely confident in their language skills to write in English themselves.

The labours of the working group have had one immediate positive result. The group's recommendations have made it possible for the relevant professional associations in the field of English and American Studies to participate in a pilot study. The panel started work in March 2011. Its findings were published in November of the following year (Wissenschaftsrat 2012). The results of the research rating *Anglistik/Amerikanistik* will eventually help determine whether the *Wissenschaftsrat's* approach to measuring research output in consultation with the relevant communities will have a future as a routine tool in the German system of higher education.

If the pilot study turns out to be successful, English and American Studies in Germany will take the rating exercise as the external stimulus to undertake the necessary critical stock-taking that every department needs at intervals. Owing to the safeguards described above, researchers can rest assured that their output is measured against criteria developed by their peers. In the full concert of disciplines in the university, scholars in English and American studies will not have to plead that their subject represents a special case—a strategy which may bring short-term rewards but which is sure to marginalize a field in the long run.

In marketing the rating exercise to the community, both the *Wissenschaftsrat* and the professional associations will be well advised to rephrase the definition

quoted above ('Unterstützung der Leitungen bei strategischer Steuerung durch vergleichende Informationen über Stärken und Schwächen einer Einrichtung') as:

> Unterstützung der Einrichtung bei Standortbestimmung und Weiterentwicklung durch vergleichende Informationen über Stärken und Schwächen der Leistungen der Forscherinnen und Forscher am Ort.
>
> [Supporting the unit in its efforts to assess its position and develop its potential by providing comparative information on strengths and weaknesses of research carried out locally.]

Understood in this way, the rating exercise can become part of a dialogue between scholars and the other stakeholders in the academic system: administrations, funding authorities, other (and sometimes competing) disciplines and, not least, the educated public whose support the humanities need more than other subjects in order to survive and prosper.

If this sounds too good to be true, consider the following three alternative scenarios which might result from a successful pilot study. It is the year 2027, and we are going through the preparations for the second routine rating for English and American Studies in German higher education (after two seven-year cycles: 2014–2020, 2021–2027).

The first scenario is the dystopian one. Status hierarchies and the peculiarly strong German fixation on the professorial chair[1] will still reign supreme, and we will witness a replay of a heated debate which took place in the 2010 meetings of the working group: 'Is my colleague allowed to report a publication by his assistant, just so he can boost his standing in the rating?' Assuming that there are two 'chairs' in English linguistics in a department, the chief motivation of each chairholder to take part in the rating will still be the hope that each one will turn out the better one of the two (rather than both putting on a good show jointly, in the interest of their department and university, and—not least—for current and prospective students). Among the publications reported we will find a 500-page tome titled *Morphologische Kreativität im nigerianischen Englisch: Neologismen aus der Presse*, published in German, by a German academic vanity press, with a subsidy, and a print run of 150, only five of which are sold outside Germany. This notwithstanding, it is cited as a 'magisterial treatment of its topic, well written and with many interesting case studies'.

This, on the other hand, is the utopian scenario. While the pilot rating (2012) stirred up a lot of furore at the time, the first routine exercise in 2020 added modifications to reduce the burden on evaluators and evaluees, thus increasing acceptance in the community. By 2027, ratings have become socially embedded practice in the academic community, including the humanities, and apart from mild irritation caused by the inevitable bureaucratic requirements, the general response is positive – along the lines of 'good thing somebody is taking note of the research we're doing here', 'well, they've politely pointed out the weaknesses that, to be honest, we have been aware of ourselves—in fact, they've given us free expert advice' and 'good thing we know where we stand this time, and good thing we've improved since the last one'.

[1] Consult the web for the collocation 'member(s) of my chair' and observe how much of the material emanates from the .de top-level national domain.

Neither of the extreme scenarios is likely. As an optimist, I hope for a moderately positive reception of ratings in the humanities. Colleagues will actively embrace ratings as an opportunity to showcase their achievement, but, as in the pilot study, researchers will groan at the tedium of compiling the self-report, and this will be echoed by assessors' groans at the tedium of some of the writing they will have to read.

Acknowledgments I would like to thank Dr. Elke Lütkemeier and Dr. Veronika Khlavna, of the *Wissenschaftsrat's* head office, for reading and commenting on an earlier version of this paper. I have profited greatly from their long involvement in the *Wissenschaftsrat's* rating projects.

References

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/.

Wissenschaftsrat. (2004). Empfehlungen zu Rankings im Wissenschaftssystem. Teil 1: Forschung. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/download/archiv/6285-04.pdf.

Wissenschaftsrat. (2008). Pilotstudie Forschungsrating. Empfehlungen und Dokumentation. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/download/Forschungsrating/Dokumente/FAQ/Pilotstudie_Forschungsrating_2008.pdf.

Wissenschaftsrat. (2010). Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/download/archiv/10039-10.pdf.

Wissenschaftsrat. (2011). Forschungsrating Elektrotechnik und Informationstechnik. Einzelergebnisse der Universitäten und ausseruniversitären Forschungseinrichtungen. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/download/archiv/1328-11_Ergebnisdarstellungen.pdf.

Wissenschaftsrat. (2012). Ergebnisse des Forschungsratings Anglistik und Amerikanistik. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/download/archiv/2756-12.pdf.

‘21 Grams’: Interdisciplinarity and the Assessment of Quality in the Humanities

Klaus Stierstorfer and Peter Schneck

Abstract In their joint contribution, the president of the German Association for English Studies (Deutscher Anglistenverband), Klaus Stierstorfer, and the president of the German Association for American Studies (Deutsche Gesellschaft für Amerikastudien), Peter Schneck, describe the central motivations behind the decision to actively support the pilot study for the research rating of the German Council of Science and Humanities (*Wissenschaftsrat*) despite some fundamental skepticism among the associations’s members. On the basis of five basic propositions—different in each argument—they both insist that the assessment of research quality in the humanities inevitably requires the central involvement of the disciplines assessed in order to reflect on and formulate the central categories, standards and procedures best suited for such assessments. Such a process must take into account the complexity of research processes and results in the humanities whose qualitative dimensions cannot be fully measured by quantitative methods.

1 Rating Research: Who Needs It, and What Is It Good For? (by Klaus Stiersdorfer)

Research rating and ranking is happening now, at least in German academia in my experience, and it has been growing in the anglophone countries, with which I deal professionally, at an alarming pace and as a kind of *menetekel* for whatever other countries may be planning to do in the future. This is why, and here is my *first thesis*, research rating and ranking cannot be avoided at present. If my first thesis is accepted, then it is worth exploring what it looks like at present in the humanities.

K. Stierstorfer (✉)
English Department, Westfälische Wilhelms-Universität Münster,
Johannisstr. 12–20, 48143 Münster, Germany
e-mail: stiersto@wwu.de

P. Schneck
Institute for English and American Studies, Universität Osnabrück,
Neuer Graben 40, 49069 Osnabrück, Germany
e-mail: peter.schneck@uni-osnabrueck.de

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_16

Most rating and ranking systems I have come across involve any one of the following procedures: peer reviewing of research publications; measuring of quantities of publications; opinion polls on the research reputations of individual institutions and agencies, or any combination of the three. I will not dwell on the latter two as they seem the most obviously inadequate for rating in the humanities, but do want to broach briefly the topic of peer reviewing which is widely seen as the fairest and most reliable tool of the three. The problems I see with it in its current form have, however, to do with fairness and transparency. With most reviewing procedures, the image of the administration of justice attributed to the so-called dark middle ages seems appropriate. There is little transparency in the application of pre-specified criteria; the actual judges (peer-reviewers) are still shielded from the person under review (the defendant) by the inquisitorial screen of anonymity; and the defendant has hardly any means of recourse to plead his or her case when the verdict is negative. This leads to a situation when most researchers in my field, at least where they have the choice, avoid such reviewing processes as the impression (true or not) arising from this black-box juridical system is imputations of favouritism, nepotism and the pursuit of non-scholarly, strategic or political ends under cover of this anonymity. The much-propounded 'blind' or even 'double blind' peer-review really does not mean that justice is iconically blind (as she should be) as to the addressee of her ministrations (projects under review are all too easily attributable in small research communities), but that reviewees are blinded (as they should not be) as to who is their judge and on what grounds their verdict is really passed. Hence, on this ground and many others, my *second thesis* is, current research rating needs improvement if we want to stick to this practice.

How such improvement can be brought about is, of course, the philosopher's stone here, but before its quest is started, the issue of the necessity of rating research in the humanities in the first place must be dealt with. As this is a short statement, the answer suggested here—which is also the prevalent opinion in the Deutscher Anglistenverband and the official position of its presidency and council—is essentially twofold. First, and this is my *thesis number three*, we need research rating because it is there or, more precisely, scholars in the humanities and their societies and associations should get involved in research rating because they are being practiced at the moment; trying to make oneself heard and get involved in establishing the fairest and best practice possible seems reasonable if not logical and unavoidable. Experience has shown that outright refusal to join the discussion does not help to avoid rating and ranking but produces bad, because inexpertly designed procedures.

Why then has research rating been established in the first place? The simple answer is: money. In the progressive commercialization and economization (if that is a word) of our academia, the political focus on money invested in research has been immense, and hence a mechanism for its distribution was sorely needed. On a simple, outcome-oriented economic model, the logical system is to put money where the best outcome is. Hence the idea to measure research outcomes and put most money where the best outcomes can be registered or at least expected. Thus,

research rating is primarily an administrative tool that has to do with investing and distributing limited funds for research. The crux of defining and comparing precisely these outcomes has long been overlooked or neglected. In the most negative reading, the whole process only shifts the problem to another scenario.

Does rating have any benefits for the scholar or researcher in the humanities? My answer is: No, surely not primarily. In a slightly more personal explanation I would stress that I am not interested in knowing whether my colleague X’s new monograph is better than mine, and if so how much on a scale from 1 to 10, neither do I need to know whether colleague Y’s article in a field I am interested in is rated high or low before I read it as the specific questions I bring to it in my specific research context may differ from quality criteria, nor do I have any desire to be informed whether my publications of the last 5 years are to be graded as 5, 6 or 7 on a scale of 1–10. For purposes of orientation which books and articles to look at in the first place, I have sufficient bibliographic and reviewing tools at hand which are well-established and efficient, even if not easily translatable onto scales from 1 to 10. Thus, my *thesis number four* says research rating is next to useless for the purposes of research itself and time spent on it would be immeasurably better spent on such research.

But, if we cannot reasonably avoid research rating at present, and even if it seems pointless for research, can we gather some lateral benefits from it, although it remains primarily superfluous in the eyes of the researcher? Here my *fifth thesis* is yes, research rating could be devised in such ways that a number of collateral benefits might accrue. Again, a lot of creative thinking could and must go into this question, but I only want to focus on one possible aspect here, that is disciplinary self-reflection. By thinking about criteria how quality of research can be measured and understood, scholars in the humanities will be forced to reflect on their current standards and aims of research and how to define them. This process can help individual disciplines to identify where they stand as a discipline and where they might want to be going in the future, as the steering function of rating procedures can hardly be underestimated. While rating may thus be a good thing for initiating and furthering discussions in disciplines and professional associations such as our Anglistenverband, this does not mean that these guidelines agreed on for the entire discipline are really a good yardstick for individual instances of research. Especially in the humanities we know too well that innovative research is, as Thomas Kuhn, Paul Feyerabend and others have argued, all too often not the kind that is immediately recognizable as such by current disciplinary standards.

Conclusion: Although the benefits seem lateral at best, rating of research is nothing that the humanities can easily avoid at the moment, so it seems better to embrace the discussion leading to its implementation with full commitment in the service of the colleagues for whom we speak in our various associations. The search for a fair, transparent and equitable rating system in the humanities may be a quest for the philosopher’s stone, but that does not mean that, under current circumstances, we should not try as best we can.

Thesis 1: Research rating and ranking cannot be avoided at present.
Thesis 2: Research rating and ranking needs improvement if it is to be continued.
Thesis 3: Research rating and ranking is needed because it is there.
Thesis 4: Research rating and ranking is useless for research itself.
Thesis 5: Research rating and ranking can produce collateral benefits.

2 'Weighing the Soul' of the Humanities (by Peter Schneck)

Let me begin with a little historical anecdote: On April 10th 1901, Dr. Duncan MacDougall, a medical researcher from Dorchester, Massachusetts conducted an experiment to determine the physical existence of the soul. Placing six moribund patients on specially designed scales, the doctor tried to quantify the soul by measuring the weight of the patient's bodies shortly before and shortly after their death. Comparing the difference between the two assessments, MacDougall found that each of the patient's bodies lost precisely the same amount of weight, which was around three-fourth of an ounce, or about 21 g. Since he could think of no other explanation for the difference in weight, the doctor concluded that in the moment of death the soul had left the patient's body; thus the soul not only existed, it's weight could also be pinned down rather precisely at 21 g—which is probably less than one would have expected for such a 'weighty' phenomena as the soul given its metaphysical significance throughout our cultural and spiritual history.

While MacDougall's weighing of the soul may be regarded as one of the countless, equally eccentric and futile attempts to measure the immeasurable—an attempt which is symptomatic for a climate of extreme scientific optimism and positivism around the turn of the 19th to the 20th century—it may nevertheless be instructive for understanding the current struggle between those who propose to assess, rate or quantify the quality of research in the humanities with objective methods of weighing and measurement, and those who think that this attempt would amount to a futile 'weighing of the soul'—that is, an absurd, useless and basically misguided exercise.

The anecdote may be instructive in the context of our discussion for more than one reason, but before I turn to the problem of measuring the immeasurable in the main part of my short remarks, let me clarify a few things from the start.

On the one hand, I am talking to you as a humanities scholar whose teaching and research has been subjected to various forms of quality assessment by an extended number of parties: by other scholars, both from my own field and from other neighbouring fields, by various university administrations and committees, by the review boards of various national and international research funding agencies and institutions, as well as by various assessment boards of the federal state and on the national level. Last, but not least, I have also been asked numerous times to assess myself not by mere introspection, but in a more regulated and prescribed form.

Ever since my performance as a scholar became the subject of a standardized questionnaire for the first time in 1984 at a leading American university, quality

assessment in all its different forms has remained an inescapable part of my scholarly and professional existence.

From this perspective of personal experience as an individual scholar, my feelings towards the continuous increase of assessment processes, the growing repertoire of procedures and protocols, as well as in face of the various institutional and public ratings and rankings in which they result—my sentiments in regard to all this excessive monitoring and controlling could best be described by quoting Elvis Costello: ‘I used to be disgusted, now I’m trying to be amused.’

To put it a bit more precisely; even though over the last decades I have come to experience and somewhat grudgingly accept an astounding number of forms of quality assessment and rating processes in the humanities as inescapable, that does not in any way mean I deem them indispensable. On the contrary, as an individual scholar in the humanities, I have increasingly come to doubt and, in fact, severely question both the essential necessity and the positive effect of quantifying ratings and rankings in and for the specific form of research that is being done in the humanities. To put it bluntly: I find it rather hard, if not impossible, to conceive of any process of calculating and expressing in numbers the difference in quality in regard to research in my field that would actually have any impact other than to regulate it (mainstreaming it, prescribing it) by rather artificial measures of comparison.

Thus, the only thing I learned so far from the ongoing and increasing assessment and quantification of research quality in the humanities is this: Whatever can be quantified, will be quantified—and if it hasn’t been quantified yet, it will be quantified eventually. So I agree with my colleague Klaus Stierstorfer that if ratings and rankings are here to stay there is hardly a way to avoid them—but that doesn’t make them more useful or attractive.

As Werner Plumpe, the president of the Association of German Historians has recently argued with considerable gloom, the sheer pressure of and rush towards ratings and rankings may eventually even reach the unquantifiable soul of the humanities: enforcing quantifying methods on central dimensions of research that cannot and should not be measured and expressed by numerical values only.

There are good reasons to accept some of the more convincing arguments that Plumpe brings forth against rating and ranking procedures in the humanities based on quantification, and I easily agree with most of his criticism and scepticism in regard to the uselessness of quantification for the acknowledgement and assessment of research quality in the humanities. There may also be good reason to subscribe to Plumpe’s skepticism that there is a great danger of misinterpretation, or even misuse by third parties, resulting from the suggestive comparability of mere numerical values—something that must be seen as a central concern given the fact that all these numerical values are (increasingly) used as evidence and arguments for the distribution of resources by universities, by the state (both on the federal and the national level) and by third party sponsors like research foundations (both national and international).

And yet there is something slightly uncomfortable and counterintuitive in this well-stated arguments, and even though I share both the reasoning and the sentiment to a certain degree, eventually the conclusions I draw from the current situation are rather different.

In fact, while Plumpe (and the majority of his colleagues in the association of German historians) have emphatically decided not to take part in the preparatory study initiated by the *Deutscher Wissenschaftsrat* (German Science Council), the *Deutscher Anglistenverband* (German Association for English Studies) and the *Deutsche Gesellschaft für Amerikastudien* (German Association for American Studies) have decided to do just that—despite the fact that we share the fundamental scepticism of our colleagues from the history departments about essential aspects of rating and ranking in the humanities per se.

But there are several reasons for this decision, and some of them have already been presented in summarized form by Klaus Stierstorfer. My task in the following parts of these short remarks will be to describe the specific perspective of the association which I represent in respect to the projected study but also in general. This perspective is particularly characterized by the strong interdisciplinary traits of the research that is being done in German American Studies (or more precisely *Amerikaforschung*).

I said there is something counterintuitive or uncomfortable about the complete rejection of the quantification of research quality in the humanities. While there are, as I readily acknowledged, good arguments against quantification as such, these arguments should not (and probably cannot) obscure our perception of the high degree of assessment by quantification that is already in practice in the humanities—in fact, one could argue that it is quantification which dominates the assessment of individual research in the humanities from the very start until the moment when one has successfully become installed by a committee—on the basis of other assessments—as a university professor. In other words, the professional success in the academic field of the humanities is essentially based on ratings and rankings and other accepted assessment procedures within the field. While these procedures are of course not completely based on or expressed in numbers, one cannot overlook or deny the existence and significance of quantification within these assessment practices in the humanities.

This is not meant to be a rhetorical move—I don't think that my colleagues from the history departments would deny the existence of quantification and ranking procedures within their field and as part of their own daily academic practice. Yet while they would readily attest this, they would probably also insist that all this rating and ranking is only done by peers, and based on meticulous and highly reflected methods of reviewing and critical acknowledgment.

However, if there are procedures of assessment involving quantification established in the field as such, it is obvious that the argument against quantification in the humanities is either a universal one—then it either works or it doesn't; and if it does not work because it can never capture the 'soul' that is the real quality of research done in the humanities, then one should drop it altogether: no more grading of research papers, no more graded forms of assessment for doctoral theses on a standard scale (even when using the Latin terms this is still a quantification of quality), no more ranking lists in committees etc.

On the other hand, if the argument is *not* a universal one (and I don't think it is or can be) then the debate should not be about quantification at all, but, rather about consensual standards of comparison and accepted and/or acceptable conditions of

assessment which make the quantified expression of quality not only possible but even desirable for pragmatic reasons (and a number of factors have been named already during our discussions: the sheer increase of scholarship and its ever growing diversity, international competition and funding schemes within the common European research area etc.).

Another aspect that also tends to be neglected in the debate (and I am only talking about the debate about the pros and cons of assessment and quantification of research quality) is the increasing development of new transnational research and study programs, especially on the young researchers level, i.e. joint doctoral programs within the humanities offered and designed by institutions from different countries across Europe. One of the most challenging tasks is to find a common denominator for the assessment and control of the quality of the study programme and the research of the individual researcher. The same is true for international research consortia: there has to be a shared understanding of the quality standards that would guide and make possible the assessment of the research to be conducted. This is an aspect that is of special significance for American Studies as a discipline and a field of research, since in contrast to English Studies (*Anglistik*), American studies has been conceived from the start as a fundamentally interdisciplinary enterprise. In fact, one could argue that American Studies is the name for research done across the boundaries of various disciplines and since its inception this understanding has always led to intense struggles about the proper methodologies, the common concepts, the shared terminology and, last but not least, the commonly accepted standards of quality in research between all participating disciplines.

Therefore, from the perspective of the scientific community involved in research in American Studies in Germany, the participation in the proposed pilot study by the Science Council has both professional, strategic and pragmatic reasons. On the one hand, it presents a calculated step to maintain a central role in the debate and definition of standard criteria and procedures to assess the quality of research done within the discipline. At the same time, it acknowledges the increasing dynamics of collaborative research agendas across disciplines and across national research areas, which are at the heart of the current struggles for standards, criteria and indicators that may be transferable and commonly acceptable at the same time.

In conclusion, one could summarize the motivational aspects that has guided the decision of the DGfA as follows:

- To assure the active participation and indispensable involvement of the field/scientific community in the process of defining standards and criteria of assessment for the quality of research within the field
- To allow for an open and ongoing debate about standards and criteria within the field and across the disciplines ⇒ interdisciplinary research community
- To actively take on responsibility for the development of common standards and criteria
- To make transparent and critically debate existing standards
- To develop common consensual standards across disciplines that meet the requirements and the dynamics of today's interdisciplinary research in the humanities

Let me end with a caveat: The process certainly is not an easy one, and we do not think that we should drop our guard by replacing our healthy scepticism with a naïve trust in the evidence of numbers and graphs. As has been emphasized, the process of arriving at the shared and commonly accepted standards and criteria I talked about can only be a mixture of top-down and bottom-up approaches and perspectives. To return to my initial historical anecdote: Weighing the 'soul' of the humanities should not simply be translated into a question of grams and ounces, nor should the wealth and diversity of humanities research be assessed as a *quantité negligeable*.

Research Rating *Anglistik/Amerikanistik* of the German Council of Science and Humanities

Alfred Hornung, Veronika Khlavna and Barbara Korte

Abstract The pilot study *Forschungsrating Anglistik/Amerikanistik* is the first implementation of the *Forschungsrating* in the humanities. This chapter presents the findings and conclusions of the rating. It consists of three parts: First, the results of the rating, first published in December 2012, are presented, as well as the conclusions drawn by the German Council of Science and Humanities. Second, Alfred Hornung who chaired the review board reflects on the *Forschungsrating* from the point of view of the chair of the review board as well as an *Amerikanistik* scholar. Third, Barbara Korte writes about the *Forschungsrating* from her perspective as a member of the review board and *Anglistik* scholar.

1 Research Rating in English and American Studies (by Veronika Khlavna and Alfred Hornung)

1.1 Introduction

In May 2008, the German Council for Science and Humanities, which provides advice to the German Federal Government and the State (Länder) Governments on the structure and development of higher education and research, decided to extend its pilot studies of research rating in the fields of Chemistry and Sociology to the

A. Hornung (✉)
Department of English and Linguistics, American Studies, Johannes Gutenberg University Mainz, Jakob-Welder-Weg 18 (Philosophicum), Raum 01-597, 55099 Mainz, Germany
e-mail: hornung@uni-mainz.de

V. Khlavna
Research Policy Department, German Council of Science and Humanities, 50968 Cologne, Germany
e-mail: khlavna@wissenschaftsrat.de

B. Korte
University of Freiburg, English Seminar, Rempartstr. 15, 79098 Freiburg, Germany
e-mail: barbara.korte@anglistik.uni-freiburg.de

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_17

fields of Technical Sciences and the Humanities (Wissenschaftsrat 2008, pp. 11–17). The overall goal was to test the applicability of research rating methods also in the Humanities. The disciplines selected were *Anglistik/Amerikanistik*, which comprises the subfields of English linguistics, English-language literatures and cultures, American Studies, and English didactics.[1] The results of this research rating of *Anglistik/Amerikanistik* were published in December 2012 (Wissenschaftsrat 2013, pp. 271–333).[2]

The pilot study of the research rating in the discipline of English and American Studies builds on the methodologies and criteria of procedure developed in conjunction with the pilot studies in Chemistry, Sociology, and Electrical and Computer Engineering.[3] One of the most important and essential features of the research rating is that its procedure is *explicitly designed by academic standards*. Academic standards for the research rating are guaranteed by male and female evaluators in review boards as well as by the respective academic associations. The responsibility for the first pilot study of the research rating and its further development were in the hands of a steering group consisting of the members of the scientific commission of the Wissenschaftsrat, individual and institutional members of the major science organizations as well as guests from state ministries and the Federal Ministry for Education and Research. As in the previous pilot studies, the steering group entrusted a review board with the implementation of the research rating for English and American Studies. The scientific organizations and professional associations were asked to nominate potential reviewers with an international reputation who could cover the most important subfields. The review board on English and American Studies, chaired by Prof. Dr. Alfred Hornung, consisted of 19 members. The main objectives of the review board were the definition of the field *Anglistik/Amerikanistik* and its subfields, the determination of criteria for application in the review process, the creation of appropriate questionnaires and the eventual assessments.

Based on the assumption that universities and other academic institutions pursue research in their respective fields and beyond, the assessment of research performance in English and American Studies followed the convention established in the other pilot studies and applied multiple criteria of evaluation, each of them specified by several aspects and operationalized by different quantitative and qualitative data.

---

[1] All institutions active in the research of at least one of the defined subfields were able to participate in the research rating of *Anglistik/Amerikanistik*. The time period chosen for the assessment was 7 years (1 January 2004–31 December 2010). To participate institutions had to have existed for at least half of the survey period. No other criteria, such as minimum number of personnel, were determined. As in the previous pilot studies, the response to the research rating was also very high in English and American Studies. 358 participating professors at the reporting date in 2010 represent 94 % of the 379 professors registered by the Federal Statistical Office for Teaching and Research in 'English and American Studies' (see Statistisches Bundesamt 2010, p. 94).

[2] The results of the participating institutions can be found at: http://www.wissenschaftsrat.de/nc/arbeitsbereiche-arbeitsprogramm/forschungsrating/anglistikamerikanistik.html.

[3] See Wissenschaftsrat (2008, 2013).

As in the previous pilot studies, the assessment of the research performance was based on an *informed peer-review process* by expert reviewers. For each evaluated institution, the reviewers received extensive data with quantitative and qualitative information.

In the following, the levels of the research ratings in English and American Studies and the experiences made in the review process will be outlined and explained. Subsequently, the criteria will be described. The last part will give an outlook on further procedures.

1.2 Procedural Steps

As in other disciplines, the implementation of the research rating in English and American Studies can be subdivided into four phases: 1. subject-specific operationalization, 2. collection of data from the institutions, 3. assessment of the data reviewed by the review board, 4. publication of the results and recommendations for the procedure.

1.2.1 Subject-Specific Operationalization

The subject-specific adaptation of the research rating to English and American Studies included the definition of the field and the subfields, the definition of the criteria and the data, the terms for the participation as well as the preparation of the data collection. The definition of the discipline and its subfields in English and American Studies agreed upon by the review board proved to be adequate and manageable. For comparison purposes the established definitions of the subfields (English linguistics, English Studies: Literature and Cultural Studies, American Studies, Didactics of English) should be reused in future research ratings of English and American Studies. At present the adequate assessment of interdisciplinary research is an area of concern. In order to reflect the different roles and profiles of institutions and to identify their strengths and weaknesses, the research achievements in English and American Studies were also evaluated according to multiple criteria (research quality, reputation, facilitating research and transfer to non-university recipients), each of them with differentiating aspects of assessment. These were mostly operationalized by qualitative information. The background information provided by the institutions on human resources and teaching workloads permitted the contextualization of the data with regard to research activities.

1.2.2 Collection of Data from Institution

The collection of publication lists and data in the institutions were based on the *current-potential* principle (the status of performance of actively employed scholars

at a respective institution on the reporting date of 31 December 2010 over the past 7 year period). The *work-done-at* principle was applied in cases where not all relevant data was available at the reporting date (performance of all scholars employed at the given institution in the 7 year period from 01 January 2004 to 31 December 2010). Thus, the data collection was based on the 'hybrid' approach of *current-potential* and *work-done-at*.

The data collection followed three steps: 1. personnel data, 2. publication data and 3. main data collection. In a first step, the institutions classified scholars actively engaged in English and American Studies according to professional positions, and assigned them to the four subfields. Subsequently, the institutions were asked to submit for each professor three exemplary publications from the survey period. In the course of the subsequent main data collection all other data relevant to the assessment were collected.

Except for the exemplary publications, the data of the institutions were collected in online questionnaires.

1.2.3 Assessment of the Data by the Review Board

As in previous pilot studies, the methods and the *informed peer-review* approach proved to be successful. The assessment was carried out in three steps: First, the two reviewers assigned to respective institutions reviewed the publications and data individually and independent of each other for a preliminary assessment prior to the meetings of the review board. At the meetings the review board formed two separate panels to discuss the preliminary results in subfield-specific groups. Thus English Studies: Literature and Cultural Studies joined up with American Studies, English linguistics with Didactics of English. In a final step, all reviews were put to vote in the general meetings of the plenum.

All criteria were evaluated on the level of the subfields to adequately account for the constitution of the field. After a first review of the data and in preparation for the assessment phase, the reviewers of the respective subfields met with the staff from the Office of the German Council of Science and Humanities to develop criteria for a subfield-specific assessment. This procedure allowed an early analysis of the data material and provided an appropriate access for the assessment of the individual subfields. This approach proved to be successful and should be applied in the future with particular attention to the consolidation of the results gained in subfield-specific meetings with the collectively defined criteria in the review board.

The data assembled for the assessment proved to be of different relevance. While the data collected for the assessment of the criteria 'research quality' and 'facilitating research' provided a solid and reliable basis, the assessment of the criteria of 'reputation' and 'transfer to non-university recipients' was less reliable, also due to some incomplete data. In general, the assessment model however worked out and should be retained with respect to the adjustments recommended in the Final Report of the Review Board (Wissenschaftsrat 2013, pp. 219–271). Efficiency measures were not calculated. The background information provided turned out to be helpful for the

qualification and contextualization of the other data. The high degree of agreement between the reviewers in their rating is a strong support for the reliability of the *informed* peer-review process.

1.2.4 Publication of Results

As in the previous pilot studies, the publication of the results consisted of two parts, the result report (Wissenschaftsrat 2013, pp. 271–333) and the institution-based presentation of results. The results are also available online[4] and allow a direct comparison of the institutions on the level of the different criteria for the four defined subfields.

1.3 Criteria

In line with the rating procedure the following criteria were used for the assessment of English and American Studies: 'research quality', 'reputation', 'facilitating research' and 'transfer to non-university recipients'.[5]

1.3.1 Research Quality

Quality of research is of particular importance in the assessment of research performance. Contrary to previous pilot studies, the assessment of the criterion 'quality of research' was primarily based on the assessment of the quality of the publication output. In addition, information on the quantity of the publication output was used. The focus on a qualitative assessment of the publications in English and American Studies was necessary because a citation-based performance assessment of publications does not exist, which is the case in many disciplines of the humanities.[6]

The qualitative assessment of publication performance was primarily based on the reading of the submitted exemplary publications. For this purpose, each professor

[4]The general results are published at www.forschungsrating.de. The results of the participating institutions can be found at: http://www.wissenschaftsrat.de/nc/arbeitsbereiche-arbeitsprogramm/forschungsrating/anglistikamerikanistik.html.

[5]The complete scoring matrix is available at: http://www.wissenschaftsrat.de/download/Forschungsrating/Dokumente/Bewertungsmatrix_ANAM.pdf.

[6]There are many reasons for the absence of citation indexes: lists of books and monographs in publication and citation databases are often incomplete, publications tended to be in German and hence did not figure in international citation databases, collections of essays and anthologies are not systematically evaluated, the number of citations is no clear information on the quality of a publication, since a citation can indicate both an appreciation and a critique of the respective research positions, and finally there does not seem to exist a unanimous opinion on a quality ranking of journals and other publications.

could submit three publications or publication excerpts of max 50 pages. One of the publications could be that of a young academic affiliated with the professorship. This procedure and, in particular, the possibility of considering a publication of young scholars proved to be advantageous. The assessment of publication excerpts, especially those from monographs, proved to be difficult when the reviewers did not know the complete publication. In the future it should be possible to submit the monograph and to mark the section of about 50 pages to be considered in the assessment. The qualitative assessment of the publication lists and their quantitative information (number of publications according to publication types) enhanced the reading of the submitted exemplary publications. The criteria relevant for the assessment of the publications, namely 'importance', 'degree of innovation', 'originality', 'timeliness', 'impact' (national and international), 'quality of research methods' and the range and influence of the research question for one's own discipline as well as for other fields proved to be adequate.

1.3.2 'Reputation'

The assessment of the criterion of 'reputation' was entirely based on qualitative information given for the assessment aspects of 'recognition' and 'professional activities'. The submitted entries for this criterion were very heterogeneous in terms of quality and quantity which rendered its assessment more difficult. The assessment of data given for 'recognition' proved to be especially difficult. Overall, the assessment of 'reputation' as a separate criterion was justified. To improve data quality, the definition of this criterion and its aspects should be more specified in the future, prior to the collection of data.

1.3.3 'Facilitating Research'

The assessment of 'facilitating research' intended to account for activities immanent in academic fields which enable the performance of research in the first place.[7] The evaluation aspects ('third-party funding', 'young talent', 'infrastructure and networks') and data selected for the assessment of this criterion proved appropriate. Particularly the quantitative data and indicators contributed to the simplification and transparency of the ratings.

The data collected for funding sources and the years of the expenditure of third-party funds was relatively unproblematic for the individual subfields. A possibility to optimize the collection of information on third-party funding activities might be the adaptation of the collection principle for the externally funded projects and the expended third-party funds. Since the records covered externally funded projects granted during the survey period on the one hand and the expenditure of third-party

[7] Refer to Wissenschaftsrat recommendations for comparative research assessment in the humanities (Wissenschaftsrat 2013, pp. 345–367).

funds in each year of the survey period on the other, a connection between the two pieces of information was difficult to assess.

The lists of current doctoral dissertations submitted by the institutions proved to be inconclusive. The assessment of these lists was difficult as the successful graduation can actually not be predicted. Accordingly, this data had lesser importance in the assessment process. At the beginning of the review process, the review board had decided not to assess the achievements in the promotion of young talent on the basis of the number of granted PhDs since this figure just provides information about the quantity but not the quality of the young talent. This approach proved appropriate. To allow a more precise assessment of the success of support for the young talent, this information should still be supplemented by quantitative details of completed PhDs in the future. For the assessment of the achievements of the promotion of the young talent, the collected qualitative information (name of the doctoral candidate, name of the supervisor, title and year) of completed dissertations were more important for the assessment process than information on ongoing dissertations.

An adequate assessment of information on networks and research collaborations, in which the reported scholars were significantly involved, was difficult because of the great heterogeneity of the entries and their varied significance. In some cases, major national and international networks, associations and research centres figured next to less significant and informal networks. In the future, this data should be more distinctively described.

1.3.4 'Transfer to Non-university Recipients'

This criterion assessed the contribution of the institutions with respect to research-based knowledge transfer distinguishing between 'personnel transfer' and 'knowledge transfer'. The institutions attributed different meanings to this criterium, so that the quality of the supplied entries varied accordingly. Moreover, the distinction made by the institutions between scholarly activities and those that are more likely attributable to the domain of transfer was not always comprehensible to the reviewers.

Despite the above difficulties and in view of the increasing importance of the transfer of research results, the record and assessment of transfer activities, especially to the non-university recipients, should figure prominently in the future. The distinction of the assessment aspects 'personnel transfer' and 'transfer of knowledge' was not useful since it was not always reflected in the completion of the questionnaire. In future surveys, this criterion should be defined by more distinctive aspects of assessment and more precise survey instructions.

1.3.5 Background Information

Within the scope of the assessment, the background information was used to qualify all other data. The background information provided about institutions and subfields

turned out to be extremely meaningful and helpful. The possibility to describe the local conditions for the evolution of research projects allowed the reviewers to contextualize the specific research activities, in particular the publications. The information on the teaching and examination workload as well as the personnel situation helped to account for the lack of activities in other areas. For an adequate treatment of this information, self-descriptions should be kept and should not exceed a given space.

The information on vacancies in particular was extremely useful. In order to include this information even more systematically in the assessment process as well as to integrate it into the publication of the results, the collection of data needs to be standardized.

Despite the extremely high value of the background information for the qualification of the other data, it proved nevertheless insufficient. In the interest of a more objective consideration of available resources, a separate calculation and assessment of the efficiency should be included in future reviews.

1.4 Conclusion and Outlook

The successfully conducted pilot study of the research rating in English and American Studies shows that an adequate comparative assessment of research performance in the humanities in general, and in English and American Studies in particular, is possible. The research rating is an apt procedure to account for the particular practices of research in the humanities in the context of research assessment. This is reflected in the development and operationalization of the assessment model and in the specification of the survey period. The mode of representation according to subfields and specific criteria offers addressee-oriented information.

In October 2013, the German Council of Science and Humanities proposed recommendations for the future of the research rating (Wissenschaftsrat 2013) and suggested the extension of the research ratings to more disciplines. The experience gained from the research rating in English and American Studies was incorporated into these recommendations. The financing of the implementation is currently under discussion between federal and state governments.

2 Chairing the Research Rating of *Anglistik/Amerikanistik* (by Alfred Hornung)

The research rating *Anglistik/Amerikanistik* (English and American Studies) carried out under the auspices of the *Wissenschaftsrat* formed part of the pilot studies to assess and establish quality standards in the natural sciences and the humanities. Starting out with chemistry and sociology in 2007–2008, electrical engineering and information technology as well as English and American Studies followed in

2011–2012. Recommended by professional associations and based on my record as member of the review board of the German Research Foundation on European and American Literatures I was asked to chair the review board. Acting on the proposals of the Steering Committee of the German Council of Science and Humanities and a subcommittee, which had developed criteria for the assessment of disciplines in the humanities, a group of eventually 19 members from England, Germany and Switzerland was selected from a list of national and international candidates, provided by their professional associations, the German Research Foundation and the Steering Committee of the *Wissenschaftsrat*. The Steering Committee appointed this group of reviewers and entrusted them with the research rating, supported by administrators of the Head Office (Dr. Rainer Lange, Dr. Elke Lütkemeier, Dr. Veronika Khlavna). In the first session the review board decided over the subfields of the discipline of English and American Studies and the procedure and criteria for the evaluation. Eventually four distinct subfields were defined: English linguistics, English literary and cultural studies, American Studies, and English didactics. The separate treatment of English Studies and American Studies as well as the nonrecognition of a subfield of Medieval Studies were the most controversial points in the discussions. The retrenchment of Medieval Studies, which in the past used to be a subject of English linguistics, turned out to be a fact at most universities which had sacrificed both the language and literature of the Middle Ages to new curricula in Bachelor and Masters of English degrees. The argument for the separate evaluation of the American Studies Master advanced by the Americanists was based on the interdisciplinary nature of this field of studies, which in its best representation at the John F. Kennedy Institute in Berlin, comprises the cooperation of literature, linguistics, culture, history, politics, geography and economics of North America. Indeed, the strengths of American Studies in a number of universities are based on the cooperation of these different disciplines, mostly of literature, culture, politics and history. The creation of these four subfields also necessitated an increase of the number of evaluators in American Studies and didactics of English, eventually making for a parity of respectively five colleagues in linguistics, English and American Studies, and four in didactics.

Guided by the previous pilot studies and considering the special features of disciplines in the humanities, the group eventually settled on four main criteria for the evaluation: research quality, reputation, facilitating research, transfer of research to non-university recipients. The report of the *Wissenschaftsrat* specifies the differentiation of aspects and problems in the evaluation of each of these categories. While the assessment of the research quality and facilitating research proved to be reliable categories, reputation and transfer were difficult to assess. This difficulty might also reflect a difference between national standards. North American and British universities are much more interested in communicating their work to their students and the public. Part of this community service is an adequate and comprehensible representation of a discipline and the profile of a department and its personnel. Such promotional activities also serve to attract students in a strongly competitive system of tertiary education. German academics, especially in the humanities, still seem to be hesitant about the promotion of their work and could learn from their English-language

colleagues. An explanation for this hesitancy could also be the often minimal attention and the low status accorded to disciplines in the humanities in the universities as well as in the public perception. The criterion of facilitating research might contribute to a change in this respect. Facilitating research comprises all measures taken to promote the careers of young researchers in the field. Next to the often long and time-consuming processes of directing individual dissertations, the establishment of structured PhD programs for cohorts proved to be very advantageous. This is also reflected in the successful applications for third-party funds, especially in the constitution of research training groups funded by the German Research Foundation or other sponsors. Our review of these very positive achievements also showed that the major research universities profit most from these joint research programs. At the same time the promotion of many PhDs also necessitates the creation of new avenues for jobs outside of academic careers. In this respect, more attention needs to be directed toward transfer activities and to a more pragmatic orientation of doctoral training programs.

This diversification of research and research training also pertains to the self-conception of the four subfields of the discipline *Anglistik/Amerikanistik*. German linguists of the English language have successfully adapted to international standards, which also includes a trend toward publications of articles in journals instead of lengthy monographs. While the monograph still represents the major piece of original scholarship in the humanities and allows scholars also in smaller departments to document their special expertise, the publication of articles gains increasing importance. This move from monographs to articles also reflects the time available for research in most disciplines of the humanities. Next to German Studies, *Anglistik/Amerikanistik* has the highest number of students who pursue academic degrees or want to enter a teaching career in secondary education. Much time is spent in teaching crowded lectures and seminars and grading papers. Many colleagues of the participating universities used the sections of the questionnaire provided for background information, comments about local conditions, to point to the disparity between teaching and research and to the disregard of teaching in the evaluation process.

The coexistence of academic and teacher training curricula also makes for the hybrid nature of the discipline of *Anglistik/Amerikanistik*. On the one hand the subject of 'English' for future teachers unites all four subfields and combines the tasks of linguists, Anglicists, Americanists and didacticians in teaching courses with a focus on teacher training. In most instances only colleagues in the didactics of English do research in this particular area and hence often score highly in transfer to schools and the public. On the other hand each of the four subfields pursues their research interests geared primarily to academic careers and less to teacher training. Historically the common denominator used to exist in the definition of the comprehensively defined discipline of *'Anglistik'* as philology. The study of etymological features of the English language and close readings of great literature basically stressed the competence of the language as a system, and courses as well as research were conducted in German. Starting in the 1980s this situation has changed with an emphasis on the practical knowledge of English and the performance of the language both in the classroom and

in publications. This change was a response to the powerful influence of English and American popular cultures on young people as well as the increasing importance of ethnic minorities, which challenged the mainstream cultures in the English language countries of immigration: Australia, America, Canada and Great Britain, including the former Commonwealth. Consequently the common bond of philology moved into the background and the four subfields further specialized with an emphasis on cultural studies. The formation of new cooperations and exchange programs with international colleagues and institutions intensified these specializations. The call for inter- and transdisciplinary research programs in the universities corresponded with the new application programs of academic sponsors and favoured adequate research activities. Initially the interdisciplinary nature of research and training in American Studies favoured this field, a fact which also figured prominently in the number of successful applications for third-party funds.

An important part of the research rating carried out by the review board under the auspices of the *Wissenschaftsrat* was its acceptance by institutions, colleagues and professional associations. Early on the *Wissenschaftsrat* organized two meetings in Berlin and Mainz for academic and administrative coordinators from each institution to communicate the process of evaluation and assist in the collection of data about personnel, students and research activities. Representatives of the Wissenschaftsrat, Dr. Veronika Khlavna and Dr. Elke Lütkemeier, and I attended the 2011 and 2012 annual conventions of the *Deutscher Anglistenverband* (German Association for English Studies) and the *Deutsche Gesellschaft für Amerikastudien* (German Association for American Studies) as well as the meeting of the *Deutsche Gesellschaft für Fremdsprachenforschung* (German Association for Foreign Language Research) to inform their members about the evaluation process, to gain their support and to listen to their concerns. Apart from questions about the constitution of the review board, the subdivision of the discipline into four subfields or missing ones, such as Postcolonial Studies or Medieval Studies, the strict time-period of 7 years (2004–2010) for the assessment proved to be the most important points. Even the hybrid approach to the evaluation of *current-potential* and *work-done-at* seemed inadequate and colleagues felt that the work of emeriti and the rupture caused by vacancies were not accounted for. Also, the absence of teaching from the criteria of evaluation was criticized. The differences in department structures in terms of personnel and budget, the comprehensive conception of English as one discipline as opposed to separate subfields and their number of representatives were felt to effect the comparative analysis of ratings. A serious concern was the potential usage of the evaluation results by the authorities in the universities and ministries and pursuant repercussions. In spite of these initial reservations, our reports on first results in the 2012 conventions found more acceptable audiences and many of the concerns raised initially proved to be less relevant in the review process. Maybe the knowledge about such evaluations at American universities made for the more ready acceptance of the research rating among the Americanists.

Reservations about the evaluation of a discipline in the humanities were initially also raised by some members in the Steering Committee of the *Wissenschaftsrat*. The presentation of the results, however, reconciled most members with the evaluation

process, especially since it revealed a number of analogies with the previous pilot studies, not least among them the overall average rating in research quality. At the press conference in Berlin in December 2012 journalists addressed results connected with their local universities and the relevance of the results for the discipline and their fields. My work as chair of the review board ended with a report in the general session of the Scientific Commission of the *Wissenschaftsrat* in January 2013. The high number of participants in the *Anglistik/Amerikanistik* research rating, ca. 90 % of all institutions, and the reliable results convinced the members of the Commission that the research rating developed by the *Wissenschaftsrat* could be applied to a discipline in the humanities. The successful completion of the fourth pilot study also led to the installment of, and my participation in a committee charged to prepare the basis for the extension of the research rating to all disciplines in German universities. In October 2013 the *Wissenschaftsrat* discussed the recommendations of this committee and suggested the extension of the evaluation to other disciplines on a regular basis.

The work in the review board over a 2 year period was carried out in a very cooperative and communal spirit and proved to be rewarding. The feedback between the representatives of the four subfields in separate sessions as well as their cooperation in plenary sessions contributed to the speedy conclusion of the research rating and the successful rendition of the report and its communication to our colleagues at the participating institutions. It was a professional pleasure to chair these sessions and share the insights gained from the informed-peer-review of submitted data with reviewers and the participators from the *Wissenschaftsrat*. The basically good national and international status of the discipline *Anglistik/Amerikanistik*, which emerged from the evaluations and which is documented in the report, is a very satisfying compensation for our work. Feedback from the institutions and subfields as well as positive reactions from ministerial and university authorities to the research rating further substantiate its successful application in the humanities.

3 Quo Vadis *Anglistik*? On Rating a Disintegrating Academic Field (by Barbara Korte)

The German Council of Science and Humanities' 2012 review for *Anglistik und Amerikanistik* gave rise to controversial debate in one branch of the field in particular, namely *Anglistik*. This was once the denomination for English Studies, understood as the study of the English language as well as the literatures and cultures expressed in it from the middle ages to the present, as practiced within departments of English. The results of the rating process document how one traditional area in which German scholars used to occupy a leading position has been practically eliminated from English Studies at German universities: Medieval Studies has survived at only a handful of universities, and it seems to be more strongly connected with other disciplines concerned with the period than with English Studies. Conversely, the field of English Studies now comprises many new interests and specializations, and it has

therefore split up in ways that contributed to dissent over the rating process and its categories.

The decision to run the review under the designation *Anglistik* and *Amerikanistik* was discussed in the raters' preliminary sessions and was determined to be the least controversial appellation for the field as a whole. It pays tribute to the fact that *American Studies* has emerged as a strong and highly visible branch within the study of English literatures and cultures, with a distinct profile defined by its region of scholarly interest (the United States, or North America if Canada is included), with specific inter- and transdisciplinary connections, an internationally renowned beacon (the Kennedy Institute in Berlin) and, last but not least, a very active association that promotes the distinct nature of American Studies (although most professorships for American Studies are still situated within departments of English). From the perspective of *Amerikanistik*, a separate rating category was understandably favoured over the alternative, namely to be rated in a joint group with researchers engaged in the study of all other literatures and cultures in the English language, which the assessment lumped together as *Anglistik*: *Literatur- und Kulturwissenschaft* (English literary and cultural studies). It is scholars from the latter group, or *Anglisten* in the narrow sense, who most frequently voiced objections to the separate rating category for *Amerikanistik*. The two other groups in the pilot study, namely English linguistics and English didactics, remained uncontroversial since their profiles are sufficiently distinct from literary and cultural studies in terms of research interests, methodologies and links with other disciplines.

Arguments for the joint rating of *Anglistik* and *Amerikanistik* asserted, firstly, that they still share major interests in and approaches to the study of literature, film and other areas of cultural production, and, secondly, that the separate treatment of American Studies might further promote a profiling of *Amerikanistik* against—and possibly even at the cost of—*Anglistik: Literatur- und Kulturwissenschaft*. This umbrella term also invited critique since it covers a great diversity of interests and subfields that have emerged over the years in non-Americanist English Studies: *Anglistik* (in the narrow sense) has re-invented itself significantly (not without impulses from American Studies), retaining its historical depth (if diminished as regards the Middle Ages) and some of its traditional philological orientations, but significantly expanding and complementing them under the influence of the various 'turns' of the past two or three decades.

The most prominent and consequential changes within *Anglistik* have been effected through the advance (and institutionalization) of Cultural Studies and Postcolonial Studies, for which we have now also established professorships and, in a few instances, institutes. What the *Wissenschaftsrat*'s review understood as 'English' literary and cultural studies was therefore a much bigger and far more heterogeneous bag of scholarship than that of American Studies. It is unsurprising that there were demands to split this bundle up. It was suggested, in particular, that Postcolonial Studies has become so established in the German academic system that it should have been rated on its own, as in the case of American Studies. But how, then, could one name the rest? Could 'British' Studies contain 'Irish' Studies? And where should one stop? Should specializations in Gender Studies also be rated separately? Or

Shakespeare Studies? The research landscape that the rating exercise was expected to chart would have then become too splintered for the results to be significant. In any case, it is undeniable that, if British, Postcolonial Studies *and* American Studies had been treated as one unit, the results for some universities might have been different.

However, the *Wissenschaftsrat's* pilot study did not only point to rifts within literary and cultural studies: The separate rating categories for linguistics and didactics, though less contested, indicate how it is taken for granted that these two areas have drifted apart from literary and cultural studies. Their umbilical connections to English Studies have not been cut, but some of the linguistic research conducted by members of English departments now seems just as closely affiliated with other linguistics or with cognitive studies, while English didactics is strongly connected to that of other foreign languages or with general didactics and pedagogy. Once more, this emphasizes that *Anglistik und Amerikanistik* is a vexed denomination for an academic field that has become increasingly difficult to define because of internal diversification and crossovers with other disciplines. In this respect, the 2012 study with its four groups reflects a state of disintegration that is not of purely academic interest but implies questions of an eminently political nature that affect individual scholars, individual departments and the profile of the entire field. Departments with strong overall ratings will, arguably, have a better standing within their institutions than those with weaker overall results; they might be in greater demand for collaborative projects within their institution, and hence have better chances of acquiring the third-party funding and number of doctoral students that were important criteria in the 2012 pilot study. Within departments, strongly rated subfields might desire to see their symbolic capital matched by a greater share of the budget. Weakly rated professorships might be abandoned in a department in order to strengthen more strongly rated areas, and so on.

Apart from such political consequences, the discipline might also take the rating exercise as an occasion to reflect upon where it is heading: Are we content to see the field of English Studies become increasingly split up? Do we gain or lose by progressive specializations? To what extent can our universities and departments afford or support such specialization? And how should we advise young scholars in terms of career paths? For instance, should and can English Medieval Studies be revived within the German system? It would be unrealistic to assume that the major divisions within English Studies as it currently stands are reversible. American Studies will remain strong, and Postcolonial Studies will not permit itself to be once more reduced to an appendix of 'British' (?) Studies. Yet English Studies as a whole might profit if its internal *connections* became more visible once again. It is not that these connections were not already there: they exist in the form of organizational units (departments of English), in the cooperation of individual scholars, and they are still implemented in courses of study, notably those that focus on English as a school subject. It is no coincidence that, of the rating's four groups, didactics was the only one with a truly integrative approach to 'English' in all its subfields: language, literature and culture, and significantly also across the *Anglistik/Amerikanistik* divide. Current research interests such as Transatlantic Studies, Migration Studies, Transnational

and Globalization Studies also help to bring the branches of English Studies closer together again and to generate new research areas.

The carving up of an academic field into units suitable for rating creates a publicly visible 'image', but it also gives scholars in the field an occasion to reflect upon whether they see themselves—or their subfields—as adequately represented by that image. The image of English Studies created by the 2012 pilot study seems to have aroused more thought about divisions than about the connecting lines and common research interests that prevent the field from falling apart. A reprisal of the exercise should be sensitive to the criticism voiced against the categories used in the 2012 review. And it should introduce criteria that acknowledge not only transdisciplinary research, but also *intra*disciplinary activities and their importance for the future of English Studies.

References

Statistisches Bundesamt. (2010). Bildung und Kultur. In *Personal an Hochschulen* (Fachserie 11 Reihe 4.4). Wiesbaden: Statistisches Bundesamt.

Wissenschaftsrat. (2008). Pilotstudie Forschungsrating. In *Empfehlungen und Dokumentation*. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/download/Forschungsrating/Dokumente/FAQ/Pilotstudie_Forschungsrating_2008.pdf.

Wissenschaftsrat. (2013). Pilotstudie zur Weiterentwicklung des Forschungsratings. In *Ergebnisse und Dokumentation*. Köln: Wissenschaftsrat.

Research Assessment in a Philological Discipline: Criteria and Rater Reliability

Ingo Plag

Abstract This article reports on a large-scale peer-review assessment of the research done in English departments at German universities, organized by the German *Wissenschaftsrat*. The main aim of the paper is to take a critical look at the methodology of this research assessment project based on a detailed statistical analysis of the 4,110 ratings provided by the 19 reviewers. The focus lies on the reliability of the ratings and on the nature of the criteria that were used to assess the quality of research. The analysis shows that there is little variation across raters, which is an indication of the general reliability of the results. Most criteria highly correlate with each other. Only the criterion of 'Transfer to non-academic addressees' does not correlate very strongly with other indicators of research quality. The amount of external funding turns out not to be a good indicator of research quality.

1 Introduction

There are some general concerns with regard to attempts to assess the quality of research carried out in public institutions. At the political level, it is, for example, unclear, what the aims of such assessments might be, and who might use them for which kind of decision-making. Furthermore, scholars complain that such assessments involve a great amount of effort, but it is more than doubtful that assessing research leads to higher quality of research. Another big issue is methodological in nature. Different kinds of methodologies are being employed without any clear evidence about their usefulness or reliability.

In spite of these concerns the English departments at German universities decided to participate in a large research assessment organized by the *Wissenschaftsrat*. The assessment was carried out by peers and explicitly aimed at testing the possibilities and problems of assessing research quality in the humanities, and in a philological discipline in particular. The idea that such an assessment might be especially problematic in the philologies arises from the fact that these disciplines are internally extremely

I. Plag (✉)
Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany
e-mail: ingo.plag@uni-duesseldorf.de

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_18

heterogeneous, with subdisciplines ranging from historical-hermeneutically oriented research to experimental-quantitative approaches, from highly theoretical to thoroughly applied. For this reason, the peers were explicitly asked to critically assess not only the research they had to review, but also the assessment process itself, over the two years of the project.

At the beginning the peers were highly skeptical concerning the assessment criteria and their operationalization. The assessment was supposed to be based chiefly on qualitative instead of quantitative data, and especially the reliability of these qualitative data was called into question.

The aim of the present paper is to address these concerns from an empirical perspective, answering the following research questions:

- How reliable are the judgments made by individual reviewers? How far do different raters agree, especially on criteria that cannot be quantified? Can one trust these ratings?
- What is the relationship between different quality criteria? For example, is it true that the amount of external funding attracted by a researcher is a good indicator of the quality of the research done by this researcher, as is often assumed?

These are empirical questions that can be answered through a quantitative analysis of the judgment data. The group of peers asked the present author to carry out such an analysis and publish the results in pertinent publications. Previous versions of this paper have appeared in German as Plag (2013a, b). The present version also contains some additional analyses.

In the next section I will give some background information about the procedure, which is followed by an analysis of the rater reliability in Sect. 3. Section 4 investigates the relationship between different assessment criteria.

2 Assessing Research Quality in English Departments: Methods and Procedures

This section presents a short summary of the methods and procedures developed and applied in the research rating. A more detailed discussion can be found in the pertinent report by the *Wissenschaftsrat* (Wissenschaftsrat 2012a, b).

As a first step, the peers discussed the division of English studies into pertinent subdisciplines and the categories for the rating. The group agreed to supply ratings according to four subdisciplines or 'sections': English Literature and Culture (ELC), American Studies (AS), Linguistics (LX), and Teaching English as a Foreign Language (EFL). Each section had a similar number of reviewers (19 overall).

With regard to the categories to be rated the peers agreed on four different so-called 'dimensions': *Research Quality*, *Reputation*, *Enablement*, *Transfer*. For each of the four dimensions a number of more detailed criteria were developed. Institutions were then asked to provide certain types of information for each of the criteria.

Table 1 lists the dimensions and the criteria. Table 2 illustrates the kind of information elicited from the institutions (see Wissenschaftsrat (2012a, b) for a complete list and more detailed discussion).

The information provided by the institutions was then rated according to the nine-point scale shown in Table 3.

Each section of each institution was rated by two peers (referred to as 'raters' in the following). Each rater provided their rating independent of the other rater's

Table 1 Rating dimensions and criteria

Dimension	Criterion
Quality	Quality of output
	Quantity of output
Reputation	Recognition
	Professional activities
Enablement	Junior researcher development
	External funding
	Infrastructure and networking
Transfer	Transfer of staff
	Transfer of knowledge

Table 2 Kinds of information

Criterion	Kind of information (selection)
Quality of output	Three self-selected publications per professorship, lists of publications
Quantity of output	Lists of publications
Recognition	Prizes, research fellows
Professional activities	Journal editorship, reviewing, editorial-board-membership
Junior researcher development	Dissertations, habilitations, prizes, job offers
External funding	Projects, money spent
Infrastructure and networking	Networks, research centers, conferences
Transfer of staff	Course offerings, lectures
Transfer of knowledge	Textbooks, other materials

Table 3 Rating scale

Numeric value	Linguistic value
5	Outstanding
5–4	Oustanding/very good
4	Very good
4–3	Very good/good
3	Good
3–2	Good/satisfactory
2	Satisfactory
2–1	Satisfactory/not satisfactory
1	Not satisfactory

rating. The group of peers discussed the ratings in joint meetings of all raters of a pertinent section. Based on this discussion this group decided on the ratings for the four dimensions. The vast majority of these decisions were unanimous. The resulting ratings by the sections were later discussed and approved in a plenary session with all raters from all sections. Occasionally, ratings were revised based on a re-evaluation of some of the arguments that had led to a certain rating. The final report of the group only contained the ratings of the dimensions, not the ratings for the nine criteria.

For the purpose of this paper two data sets were used. The first one (data set A) contains all independent ratings by all raters. This data set allows us to investigate the level of agreement between the two raters and the relationship between the different criteria. The second data set (data set B) contains the ratings for the four dimensions as decided in the plenary session of the group of peers. This data set is used to investigate the four dimensions on the basis of the final ratings.

For the quantitative analysis the above scale was transformed into a 9-point scale with 5 as the highest score and 1 as the lowest with intervals of 0.5. We will use standard statistical procedures, as implemented in the software package R (Core Team 2012).

3 Reliability of the Ratings

3.1 Rater Reliability

The ratings in data set A show a mean of 2.95 (standard deviation: 0.27). An analysis of variance reveals that there are significant differences between raters (*ANOVA*, $F_{(18,348)} = 188$, $p < 0.05$). Such differences are expectable as each rater reviewed a different set of institutions. Figure 1 shows the means by rater (including 95 % confidence intervals), with each rater being represented by a capital letter.

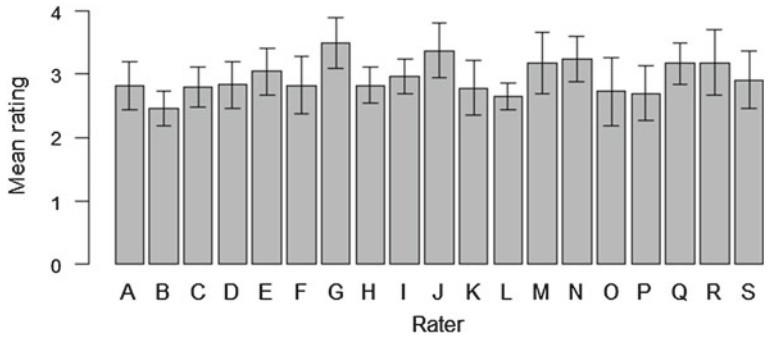
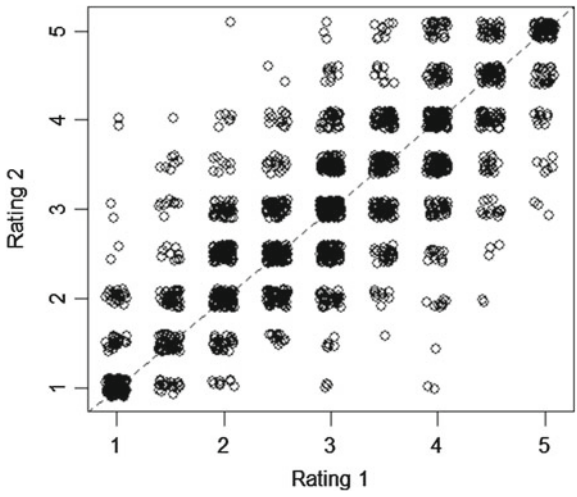


Fig. 1 Mean rating by rater

Let us now turn to the rater pairs and their agreement. 4,110 paired ratings entered our analysis. Figure 2 shows the distribution of the ratings, with some jitter added to each rating for expository purposes. Each of the 2,055 dots in the graph represents one pair of ratings. The scatter is unevenly distributed with most ratings on or close to the diagonal, where the two ratings are identical. Thus we can say that the raters tend to give similar or identical ratings. A look at the differences between ratings corroborates this impression. Figure 3 shows the distribution of the differences between ratings. 40 % of the ratings are identical and another almost 40 % differ only by 0.5. To assess the reliability and consistency of the two raters more formally, we used Cohen's Kappa and Intraclass Correlation (ICC) (see, for example, LeBreton and Senter (2007) for discussion). For our data both measures indicate that there is very strong agreement between two ratings of a given item (Cohen's Kappa: $\kappa = 0.82$, $ICC = 0.802$).

Fig. 2 Ratings by rater



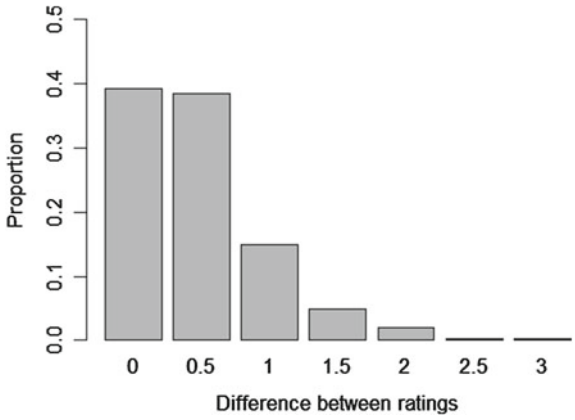


Fig. 3 Distribution of difference between ratings

To summarize, the raters very much agree in their assessment of the criteria, which means that it is obviously possible to reliably assess the quality of research in the disciplines at hand.

It is still an open question, however, whether this reliability differs with regard to the different criteria being rated. This question will be answered in the next subsection.

3.2 Rating Variation Across Different Criteria

An analysis of variance with 'criterion' as independent variable and 'difference in rating' as dependent variable yielded a significant effect of criterion (*ANOVA*, $F_{(12,\ 2012)} = 1.96$, $p < 0.05$). In other words, the difference in the ratings of two raters is dependent on what kind of category was rated. Figure 4 shows the distribution of mean differences by criterion or dimension. Regression analyses show that the six categories with the lowest mean differences do not differ significantly from one another. *Enablement*, however, differs from recognition ($p < 0.05$, $t_{(2012)} = 2.02$) and from all categories to the right of it in Fig. 4.

The dimensions *Research Quality*, *Reputation*, *Enablement*, *Transfer* do not differ significantly from one another concerning the rating differences. With the rating criteria the situation is different. The rating of external funding is least variable, an outcome that is unsurprising given that this criterion is largely dependent on counting sums of money. At the other end of the scale, knowledge transfer seems much harder to reliably evaluate.

It is perhaps striking that the dimension *Research Quality*, which rested primarily on the qualitative assessment of sent-in publications, reached the second best agreement (measured in mean rating difference) in the ratings. This fact can be interpreted in such a way that there are apparently quite clear quality standards in the disciplines

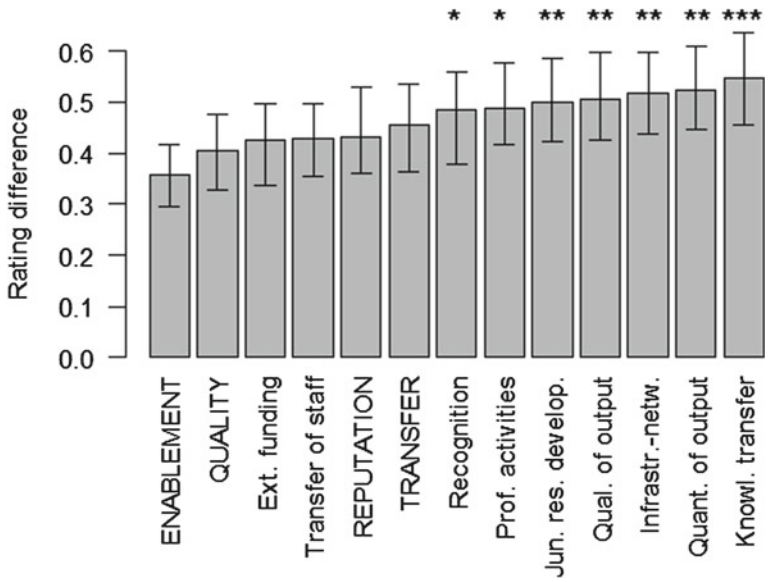


Fig. 4 Mean difference in ratings by category (significance levels for these differences are given by asterisks: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

under discussion, and that these standards were applied by the raters in a consistent fashion.

In sum, there is very good evidence that the peer review procedure as implemented in this project has led to reliable ratings and trustworthy quality assessments.

4 Rating Categories: What Do They Really Tell Us?

In this section we take a closer look at the categories to be rated in order to see in which relation they stand to each other.

4.1 Criteria

If we look at the correlations of the ratings in data set 1 across the nine criteria, we see that all 36 correlations are positive and highly significant (Spearman test). This means that, for a given institution higher scores on one criterion go together with higher scores in any other given criterion. This effect varies, however, quite a bit. Figure 5 illustrates the distribution of the 36 correlation coefficients.

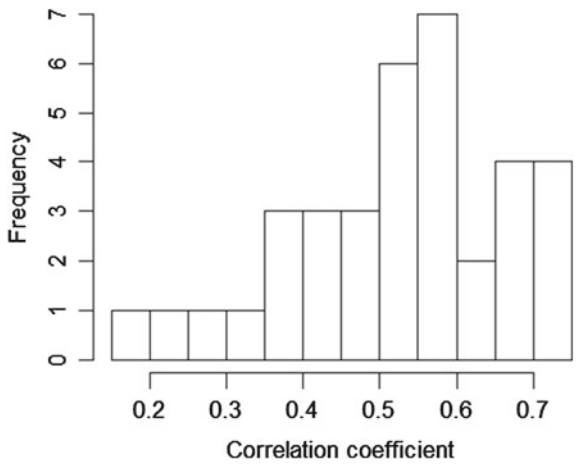


Fig. 5 Distribution of the 36 correlation coefficients for the 9 criteria

Table 4 Highest and lowest correlations between rating criteria

Correlation	Criterion 1	Criterion 2
Strong ($\rho > 0.68$)	Quality of output	Quantity of output
	Professional activities	Recognition
	Professional activities	Infrastructure and networking
	External funding	Infrastructure and networking
	Transfer of staff	Knowledge transfer
Weak ($\rho \leq 0.3$)	Transfer of staff	Quality of output
	Transfer of staff	Quantity of output
	Knowledge transfer	Quality of output

A closer look at these correlations is interesting. Table 4 lists the highest and lowest coefficients.

We can see that some criteria have close relationships to others. A high quality of the publications goes together with a high quantity. This means that people who have very good publications are also the ones that publish a lot. Other very high correlations might be less surprising. That external funds may lead to good infrastructures seems quite predictable, for example.

In the context of today's impoverished universities, external funding has become a prominent issue in political debates inside and outside academia. A common, even if often implicit, assumption in these debates is that attracting external funding is an indication of a researcher's excellence. The present data show that this assumption is not justified. There is a positive correlation between the amount of external funding and the quality and quantity of the research output ($\rho = 0.47$ and $\rho = 0.45$, respectively), but these correlations are not particularly strong. In fact, more than two thirds of the correlations between criteria are stronger.

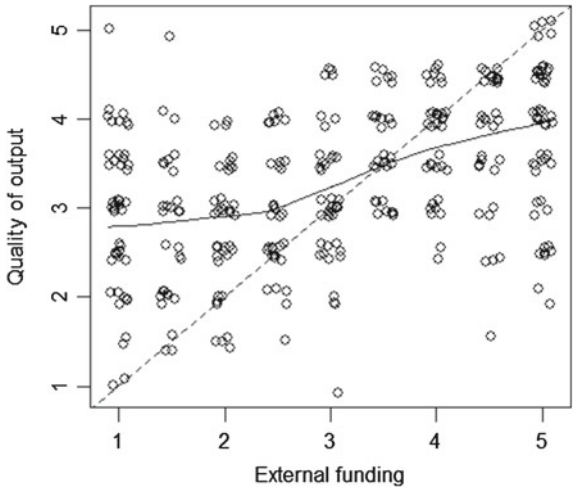


Fig. 6 Quality of output by external funding

Figure 6 shows the relationship between external funding and the quality of the output ($N = 335$, again I have added some jitter). The solid black line gives the trend in the data using a non-parametric scatterplot smoother (Cleveland 1979), the broken line represents a perfect correlation ($\rho = 1$). We can see that the general trend is not particularly strong, at both ends of the x-axis there is a lot of dispersion. What we can say, however, is that high quality research tends to go along with higher amounts of external funding. Conversely, we can state that high amounts of external funding do not necessarily mean high quality research. And there are also two institutions that lack external funding and output top quality research.

These facts suggest that the amount of external funding is not a very reliable way of measuring the quality of research.

4.2 Rating Dimensions

We can apply a similar procedure to data set 2, which contains the final results for the four rating dimensions. Table 5 summarizes the correlation coefficients in a correlations matrix.

All correlations are highly significant ($p < 0.001$, Spearman), but *Transfer* behaves differently from the other three dimensions. Whereas *Research Quality*, *Reputation* and *Enablement* highly correlate with one another ($\rho = 0.73$ or 0.69), *Transfer* does not correlate so well with the other three dimensions (with ρ-values ranging between 0.39 and 0.5). This is also illustrated in the scatterplots in Fig. 7. The left column of panels show the correlations of *Quality*, *Reputation* and *Enablement*, the right column the correlations of *Transfer* with the other three dimensions. The panels on the left show much less dispersion than those on the right, and the trend

Table 5 Highest and lowest correlations between rating criteria

	Quality	Reputation	Enablement
Reputation	0.73		
Enablement	0.69	0.73	
Transfer	0.39	0.49	0.50

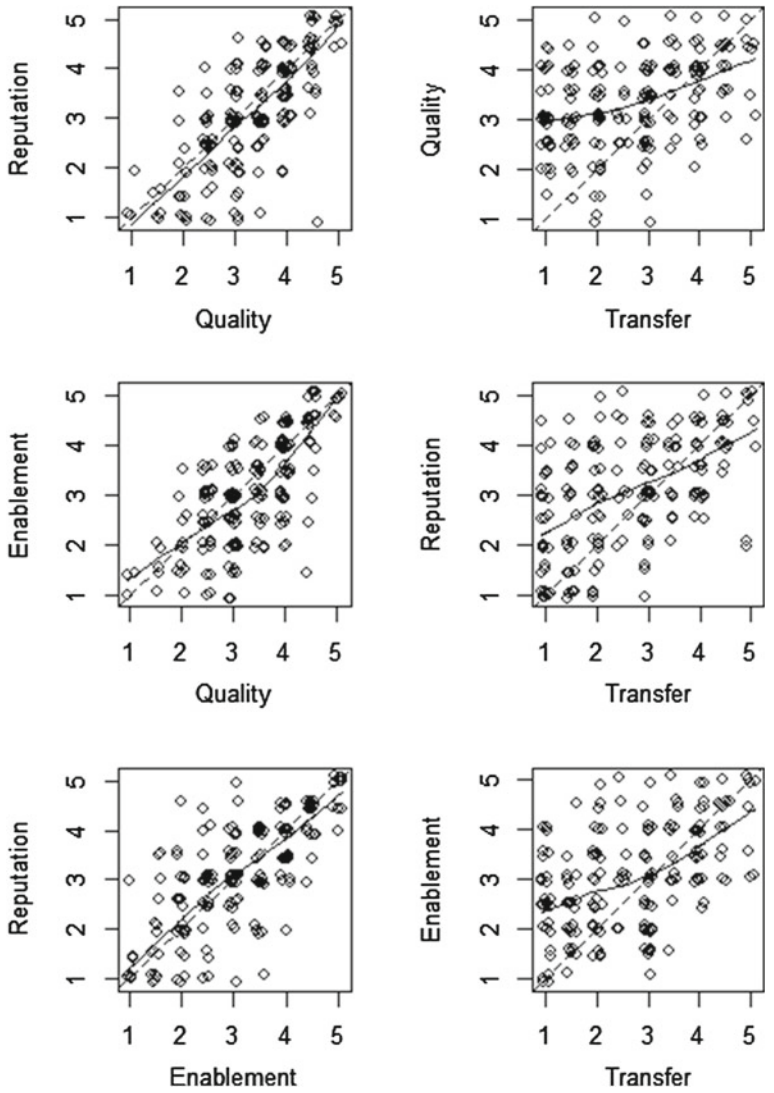


Fig. 7 Relationship between rating dimensions

as shown by the scatterplot smoother in the left panels is also much closer to the diagonal than the one in the right panels.

5 Summary and Discussion

Our analysis revealed that there is strong agreement between raters. This means that the categories to be rated were well operationalized and allowed for a consistent and transparent rating, even if the consistency varied somewhat between categories. It also means that the different subdisciplines represented in English departments in Germany have developed quality standards that are widely shared and that can be used to reach fairly objective assessments of research activities.

With regard to the relationship between the categories three main results emerged. First, there is a significant positive correlation (of varying strength) between all categories. This means that a section of an institution has received similar ratings across the categories to be rated. From a statistical viewpoint this means that the different criteria to a large part reflect the same underlying properties. This was expectable to some extent, but it raises the question of how much effort is actually needed to reach reliable results. The present project involved a considerable investment of time and money, and there is some concern whether such an investment is justified. Politically, the inclusion of many different categories is of course desirable, as it makes the assessment more acceptable for those who are being rated.

Second, not all categories correlate equally strongly, and especially the amount of external funding does not correlate well with measures that directly assess the quality of the research output. This also means that a qualitative evaluation of publications is indispensible for any attempt to assess the quality of research.

Third, we have seen that transfer does not stand in a very strong relationship to other dimensions. This can be interpreted in such a way that transfer to non-academic institutions does not play a prominent role in the research activities of English departments.

Overall we can say that the results of the assessment can be regarded as highly reliable. This result will be to the liking of those that have received good ratings and will be sad news for those who have not reached satisfactory ratings. This brings us to the perhaps decisive question: so what? Or, more concretely, who will use these results and to what end? Who is the addressee of all these assessment efforts?

One might first think of the ratees as primary addressees, as they receive feedback on many aspects of their work. It is highly doubtful, however, whether these scholars need such an assessment in order to learn something about the quality of their research. The scientific community provides constant and ample feedback, either by senior scholars (in the case of dissertations or habilitations, for example) or by peers (in the case of articles, books, jobs, promotion, project funding, prizes etc.), so that all of us seem to get enough feedback to have a fairly good idea about the quality of our own research. Furthermore, for reasons of privacy protection, the present project did not assess research quality at the level of the individual but only at the

level of sections of institutions. The peers were actually sometimes quite unhappy about this restriction since there were sometimes large differences between individuals of one section. These differences then had to be averaged out, which made the assessment less accurate and meaningful than it could have been. For the individual scholar the assessment as done in this project is therefore not really helpful, unless it could be used to improve the situation of an individual section. A reality check of this aspect is sobering, however. While it has happened that universities boasted the achievements of their respective English department as attested in this project on their university websites, I have heard of no tangible increased support (financial or other) accompanying such advertisements.

Let us therefore turn to the other potential addressees of research assessments, i.e. institutions that could use the data for their decision-making (at the departmental, faculty or university level). A discussion of the details of how exactly assessment results may feed into structural or financial decisions taken by university bodies are beyond the scope of this paper, but in general one should be in favour of such decisions being based on trustworthy and reliable data, rather than on the personal biases of decision-makers and their advisors. The present assessment of the research quality of English department certainly provides such a data base.

It should be clear, however, that success in the domain of research is only one criterion for decisions in very complex institutional settings. Apart from information on their research the institutions were also asked to provide information on the institutional settings (e.g. number of students, number of exams, number and structure of staff, number and kinds of study programs etc.). This information clearly indicated that the structural and institutional conditions in many of the departments we assessed are often quite detrimental to the aim of generating excellent research.

References

Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*, 829–836. doi:10.1080/01621459.1979.10481038.

LeBreton, J. M., & Senter, J. L. (2007). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815–852. doi:10.1177/1094428106296642.

Plag, I. (2013a). Forschungsrating der Anglistik/Amerikanistik: Analysen und Reflexionen zur Bewertung von Forschungsleistungen in einer Philologie. *Zeitschrift für Fremdsprachenforschung*, *23*, 177–194.

Plag, I. (2013b). Forschungsrating der Anglistik/Amerikanistik: Analysen und Reflexionen zur Bewertung von Forschungsleistungen in einer Philologie. Anglistik: International Journal. *English Studies*, *24*, 181–194.

R Core Team. (2012). A language and environment for statistical computing. Wien: R Core Team. Retrieved from http://www.R-project.org.

Wissenschaftsrat. (2012a). Ergebnisse des Forschungsratings Anglistik und Amerikanistik. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/download/archiv/2756-12.pdf.

Wissenschaftsrat. (2012b). Hintergrundinformation: Pilotstudie Forschungsrating im Fach Anglistik und Amerikanistik. Berlin: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/download/archiv/hginfo_2612.pdf.