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

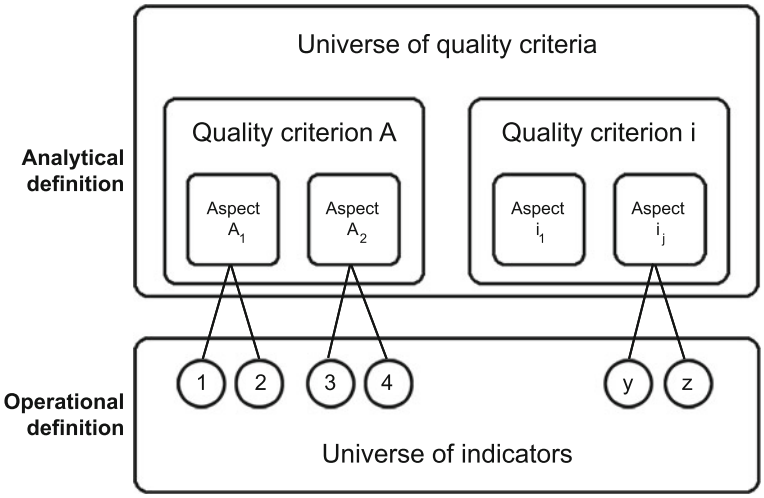

Fig. 1 Measurement model for developing quality criteria and indicators for the humanities. *Source* Hug et al. (2014)

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-

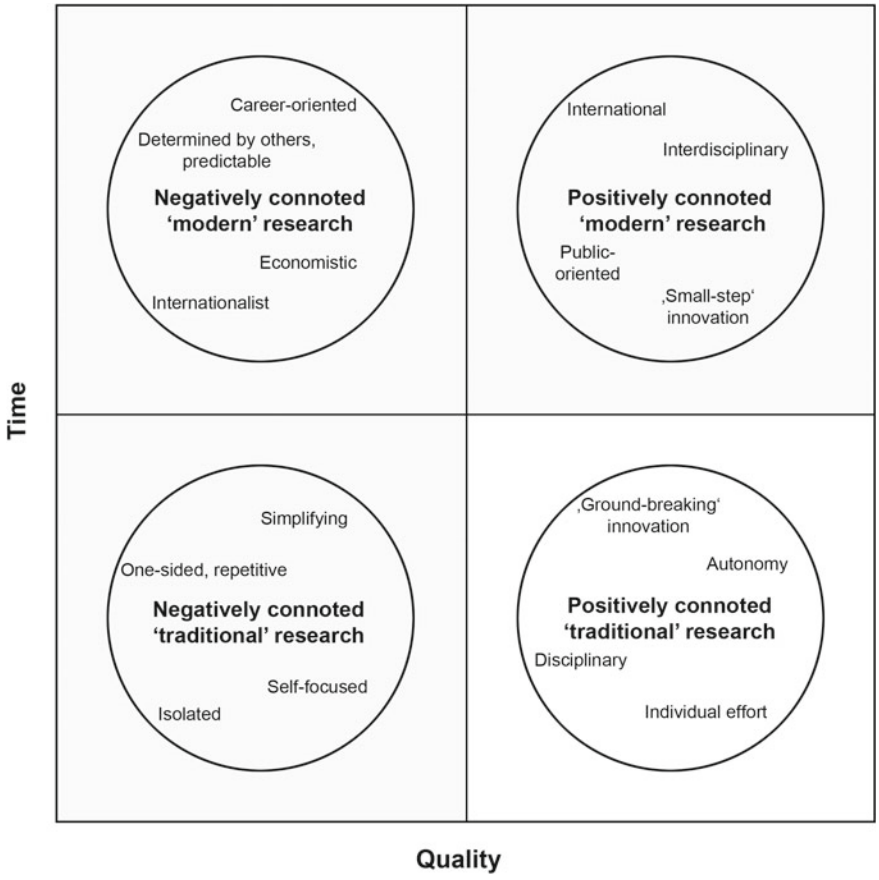

Fig. 2 Four types of research in the humanities. Commonalities across the disciplines. *Source* Ochsner et al. (2013), p. 86

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

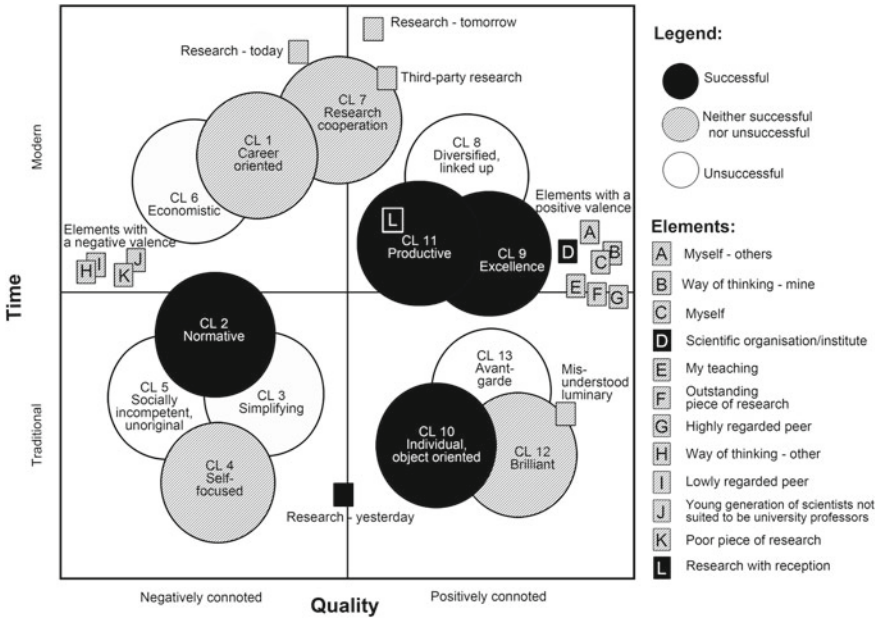

Fig. 3 Schematic representation of the clusters and elements in the discipline *German literature studies*. Slightly modified version of Ochsner et al. (2013), p. 84

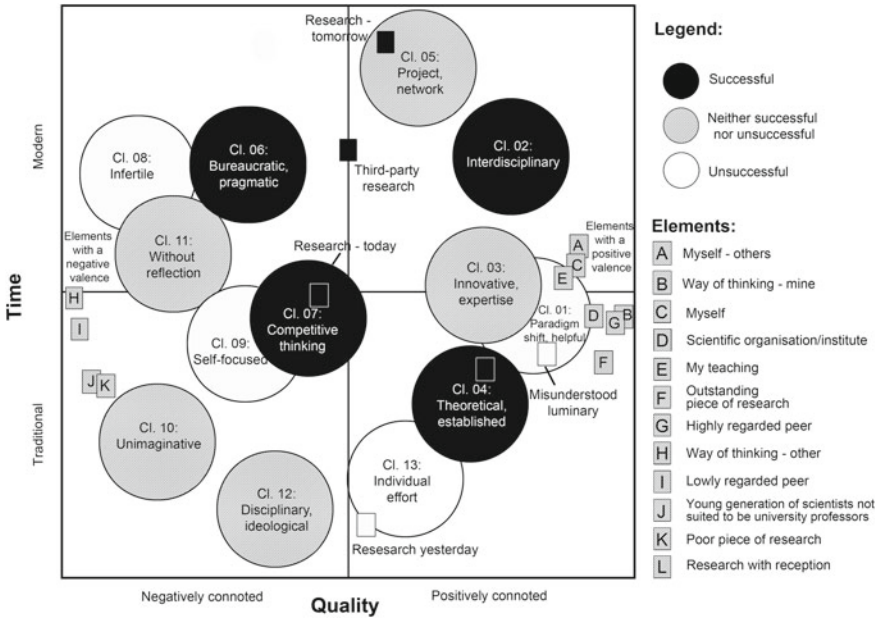

Fig. 4 Schematic representation of the clusters and elements in the discipline *English literature studies*

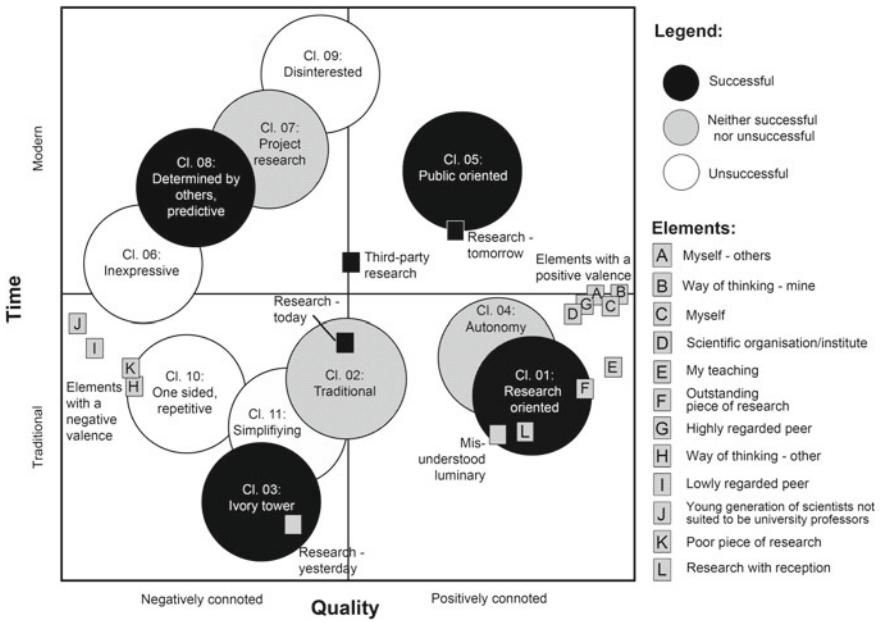

Fig. 5 Schematic representation of the clusters and elements in the discipline *art history*