Quality criteria for Sociology? What sociologists can learn from the project «Developing and Testing Research Quality Criteria in the Humanities»

Michael Ochsner, Tobias Wolbring and Sven E. Hug

Reference:
Ochsner, M., Hug, S. E. & Wolbring, T. (2015). Quality criteria for sociology? What sociologists can learn from the project «developing and testing research quality criteria in the humanities». Sociologia E Politiche Sociali, 18(2), 90–110.
https://doi.org/10.3280/sp2015-002005

Quality criteria for Sociology? What sociologists can learn from the project «Developing and Testing Research Quality Criteria in the Humanities»

Michael Ochsner[*], Tobias Wolbring[**] and Sven E. Hug[***]

Abstract:
Universities take an important role in the knowledge-society. For reasons of accountability to the public or in order to assure or enhance research quality, many universities implemented assessment procedures, often using bibliometric and other performance indicators. These procedures are mostly developed in a data-driven manner and not much is known about what the indicators in these procedures actually measure and how they affect behavior. Furthermore, the methods stem from the natural and life sciences and cannot be readily transferred to the social sciences and humanities. In this article, we present (i) quality criteria for research from the perspective of humanities scholars and how they can be transferred to sociology (ii) summarise the opportunities and limitations of the research rating of the German Council of Science and Humanities, and (iii) suggest that sociology as a discipline should develop a discipline-specific approach to research evaluation that takes into account the sociology scholars' notions of quality and the disciplines' research practices, that is bottom-up in nature, and uses both quantitative as well as qualitative data.

Keywords: Research Evaluation; Quality Criteria; Sociology; Research Policy; Research Assessment

1. Introduction

Universities take an important role in the information-, and especially in the knowledge-society (see, e.g., Välimaa and Hoffman 2008). While first mentions of the term knowledge-society surfaced already in the late 1960ies (Lane 1966) and 1970ies (Bell 1973), the concept reached policy makers and the society in general only in the 1990ies, especially when Stehr (1994) created a social theory based on the knowledge society, where he stated that labour and capital are no longer providing enough insight to understand and explain modern societies (see, e.g., Välimaa and Hoffman 2008). Since then, the generation of information and its transformation to knowledge available to all members of the society is seen crucial for the economic success of a country and also for social goals (UNESCO 2005). Hence, the last decades have been marked by a huge expansion of the higher education sector.

The importance of, and amount of public spending for, higher education asks for legitimacy and accountability of the work done at universities. In other words,

[*] Researcher, D-GESS, ETH Zurich, ochsner@gess.ethz.ch and Senior Researcher, FORS, University of Lausanne.

[**] Assistant Professor for Sociology, Department of Sociology, University of Mannheim, wolbring@uni-mannheim.de.

[***] Researcher, D-GESS, ETH Zurich, sven.hug@gess.ethz.ch and Project Manager, Evaluation Office, University of Zurich.

universities depend on public support of higher education. Consequently, the World Declaration on Higher Education for the Twenty-first Century states: «*Public support for higher education and research remains essential* to ensure a balanced achievement of educational and social missions» (UNESCO 1998, Article 14, Financing of higher education as a public service, subparagraph (a), emphasis from the source). The transformation to knowledge society and the expansion of the higher education sector fell together with the growing importance of neoliberal ideologies that promoted lean states and caused many countries to switch from Keynesianism to austerity politics[1]. Hence, since the 1990ies, in many countries the expansion of the education sector went together with budget cutting or at least budgets grew slower than student enrolments, leading to reform pressures (Geschwind and Larsson 2008). Therefore, efficiency and management grew more and more important in higher education institutions and new public management practices did not stop at the gates of the universities (see, e. g., Alexander 2000; Mora 2001; Readings 1996; Rolfe 2013).

The policy makers' demand for instruments to parametrically steer research grew particularly prominent in the twenty-first century. While during the Cold War this would have caused controversy since Western science policy held dear the scientific autonomy in the tradition of the Humboldtian university in order to contrast the state-controlled and planned science of the socialist countries (see, e.g., Weingart 2008), after the fall of the Iron Curtain universities were subject to strong competition due to globalisation and internationalisation of research. The need of accountability and the internationalisation led to the rise of quantitative assessments of research in many countries (e.g., funding, rankings, professorship appointments) – of course not without the scholars' criticism who feared the loss of their autonomy as well as unintended effects of such assessments. In the meantime, there is some experience with research assessment exercises and research evaluations on the national as well as on the institutional level and there is evidence for at least some unintended effects of quantitative assessment (see, e.g., Butler 2003; Gläser *et al.* 2002; Lawrence 2003).

The social sciences and humanities (SSH) proved especially difficult to assess quantitatively, even bibliometricians caution from readily applying bibliometrics to the SSH (see, e.g., Hicks 2004; Lariviere, Gingras and Archambault 2006; Nederhof 2006). On the one hand, there are technical problems like coverage issues in the databases used for bibliometric analyses caused by different research practices in the SSH than in the natural and life sciences. However, there is also a paradigm shift of publication and research practices due to technological change (e.g. open access, digital humanities, see Dávidházi 2014). On the other hand, there is currently a utilitarian approach to science policy and research assessments focus on direct utility of research (so-called societal impact), which, in its current interpretation, is favoring technical, natural and life sciences by relying on economic outcomes or direct, measurable usages by the public (Donovan 2007). However, a big part of SSH research is not aimed at direct, visible returns but at critically questioning the status quo of a society, preparing knowledge for later use as well as educating critical

[1] It is no surprise that the idea of knowledge economy is said to go back to Hayek (1937), who criticized communism and state planning and envisioned a democracy based on market logic and was the founder of what is today labelled neoliberalism (see Peters 2007).

citizens (see, e.g., Nussbaum 2010). Actually, SSH research is aware of the fact that the needs of the future's society cannot be predicted by today's needs or knowledge[2]. Hence, the conception of research of most SSH disciplines is not that of linear progress of knowlegde but that of a coexistence of different, even contradicting, paradigms, theories, or lines of thought (Lack 2008). Therefore, research assessments in SSH must reflect this other type of knowledge generation in order not to drown the peculiarities of SSH research and their very reasons of existence.

We argue that, up to now, this fact has not been taken into account in the development of research assessment procedures and tools very often. We suggest that assessment procedures should reflect the scholars' notions of quality and relate to the research practices of the disciplines. In a Swiss project entitled «Developing and Testing Research Quality Criteria in the Humanities», two of the authors developed quality criteria that reflect the scholars' own notions of quality for three humanities' disciplines which are especially difficult to assess with existing approaches. This paper aims at reflecting the potential of adapting such an approach to research evaluation in sociology.

We first present the project's framework to explore and develop quality criteria based on the scholars' notions of quality. We then briefly describe key findings of the project and their relevance to sociology. Next, we present an elaborated pilot study for research evaluation in sociology – the research rating of the German Council of Science and Humanities (Wissenschaftsrat 2008a) – that uses both quantitative and qualitative data. Finally, drawing on the findings of the project on quality criteria in the humanities and the experience of the research rating pilot in sociology, we suggest that sociology as a discipline should develop a discipline-specific approach to research evaluation that takes into account the sociology scholars' notions of quality and the disciplines' research practices, that is bottom-up in nature, and uses both quantitative as well as qualitative data.

2. Framework to Develop Quality Criteria for Research

With the exception of psychology, economics, and management[3], research evaluations are highly controversial in the humanities and social sciences. Some initiatives to develop instrument for research assessment in the humanities stirred strong rejections by the scholars (e.g., the European Reference Index for the Humanities, ERIH, see Andersen *et al.* 2009; or the research rating of the German Council of the Science and Humanities, see Plumpe 2009). Therefore, the idea of the project «Developing and Testing Research Quality Criteria in the Humanities» was to learn from these experiences and analyse the reasons for the rejections of evaluation tools and procedures first. We thus reviewed a broad range of documents (cf. the bibliography in Peric *et al.* 2012) and identified four main reasons for reservations against the measurement of research quality (Hug *et al.* 2014). First, the methods used

[2] The difference between the utilitarian value of Islam studies before and after September 11 2001 might illustrate this argument.

[3] All of which are often not part of SSH faculties or try to distinguish themselves from SSH disciplines, psychology by searching proximity to the (life) sciences, economy by building an own faculty.

in research assessments originate from the natural and life sciences and do not fit well to the research practices in the humanities (this is also well documented in the bibliometric community, see, e.g., Hicks 2004; Nederhof, 2006). Second, humanities scholars have reservations against quantification, especially so regarding research quality, as the statement of a humanities scholar in Fisher *et al.*'s study nicely shows: «Some efforts soar and others sink, but it is not the measurable success that matters, rather the effort» (Fisher *et al.* 2000, «The Value of a Liberal Education», para. 18). Third, humanities scholars fear dysfunctional effects of the use of quantitative indicators, e.g. loss of diversity (see Andersen *et al.* 2009). Fourth, there is a lack of consensus on research topics and on the meaningful use of methods. Hence, there is no consensus on quality criteria neither (Herbert and Kaube, 2008).

In order to address these issues, we suggest the following framework for developing quality criteria for research in the humanities of which we think that it is also applicable to the social sciences. It consists of four pillars:

The first pillar demands that the development of criteria and indicators be rooted in the disciplines themselves, thus, *adapting an inside-out approach*. Such a bottom-up procedure ensures that the disciplines' unique conceptions of research quality are reflected in the criteria. The involvement and adequate representation of the research community in the development process is an integral part of this pillar as is an open outcome (i.e., any quality criterion defined by the scholars is accepted, no matter how different it is from existing criteria).

The second pillar of the framework, *relying on a sound measurement approach,* responds to the humanities scholars' reservations against quantification. This means that it must be clear what is being measured. For this end, research quality is defined by quality criteria that are specified more clearly by aspects (i.e. analytical definition) that can be linked to quantitative indicators (i.e. operational definition). Such an approach makes it possible to identify of quantifiable and non-quantifiable criteria. That is, if for an aspect or a criterion no quantitative indicators can be found, the aspect or the criterion can only be judged by peers (see Fig. 1).

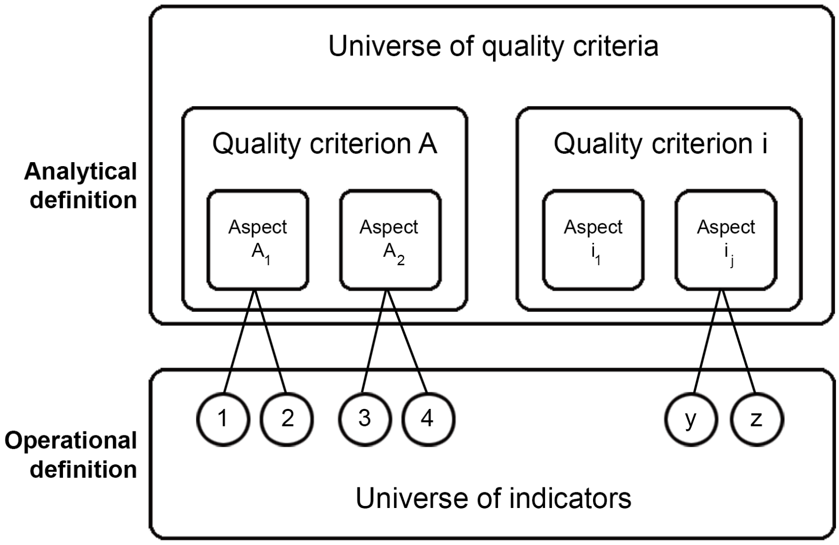


Fig. 1 – Mesurement Approach
Source: Hug *et al.* (2014)

Making the notions of quality explicit is the third pillar of the framework. It consists of two parts: First, the notions of quality that underlie the evaluation procedure or measurement instrument must be made as explicit as possible in order to reduce the uncertainties of what exactly is being measured and to make it clear in which direction research is being steered. Second, the scholars' notions of quality must be taken into account when developing quality criteria and indicators in order to assure that research is steered into the direction of research quality as perceived by the scholars (who are the only ones able to judge quality) thus reducing the likelihood of negative steering effects. However, the scholars' notions of quality might not be known and must be made explicit

The fourth pillar is *striving for consensus*. Methods have to be applied that make it possible to determine which criteria and indicators are accepted by the scholarly community and which are not.

We have implemented this framework using two specific methods: the Repertory Grid technique (Kelly 1955) and the Delphi survey method (Linstone and Turoff 1975). The Repertory Grid technique explores research conceptions and notions of quality of the scholars, thus addressing the inside-out approach of the first pillar and it enables the explication of tacit knowledge thus addressing the third pillar of the framework. The Delphi method addresses three pillars of the framework: First, it contributes to the first pillar by involving a large group of scholars; second, it assures the application of a sound measurement approach by structuring the scholars' communication process; third, the Delphi method is an excellent tool to find consensus, thus addressing the fourth pillar.

3. The case of the Humanities: Notions of Quality and Quality Criteria and Indicators

In our study, we chose three disciplines that elude particularly the commonly used evaluation procedures: German literature studies (GLS), English literature studies (ELS), and art history (AH).

First, we conducted 21 Repertory Grid interviews in order to explore the humanities scholars' notions of quality. Second, we administered a three-round Delphi survey in order to validate the results from the Repertory Grid interviews with a large international group of scholars, i.e. all scholars holding at least a PhD in one of the three disciplines GLS, ELS, and AH at the Swiss universities or at the member universities of the League of European Research Universities (LERU).

3.1 The Repertory Grid interviews

The Repertory Grid method is a versatile instrument that combines quantitative (i.e. numerical) and qualitative (i.e. syntagms) data allowing an idiographic (i.e. the scholars describe their notions of quality in their own words) as well as a nomothetic (i.e. developing of discipline-specific criteria by summarising the individual perceptions for each discipline) approach and it is capable of explicating tacit knowledge (see, e.g., Jankowiecz 2001). Due to restrictions of space, we do not describe the method in detail, for further information about the method, please confer to Ochsner *et al.* (2013).

The Repertory Grid interviews revealed that the scholars differentiate between a more «traditional» and a more «modern» conception of research, each of which can be of higher and of lower quality, thus revealing four types of research: (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 «modern» research that is characterized by pragmatism, career aspirations, economization, and pre-structuring.

Thus, pure «traditional» research can be described as disciplinary, local, individual, and autonomous research, whereas pure «modern» research is characterised by interdisciplinary, internationality, team- or project work, and societal orientation. This reveals a first insight into the relationship of research quality and some indicators that are commonly used in evaluation procedures: interdisciplinarity, cooperation, internationality, and societal impact are not related to the *quality* dimension but to the *time* dimension and are thus not indicators for research quality but for the «modern» conception of research. They are double-edged in nature because they can characterise positively connoted research as well as negatively connoted research.

We also found two different kinds of innovation related to the two concepts of research: *ground-breaking* innovation and *small-step* innovation. *Ground-breaking* innovation on the one hand is related to the «traditional» conception of research, leads to new theories, and can cause a paradigm change but might not yet be crowned by success. *Small-step* innovation on the other hand is related to the «modern» conception of research and describes innovation that finds strong reception and starts from and ties into existing knowledge.

On the quality dimension there is practically no difference between the «modern» and the «traditional» conception of research (the «traditional» conception of research scores just a little bit higher). Hence, it is important to take both of these conceptions into account in evaluations defining either a «traditional» or «modern» conception of science as quality standard in research evaluation might risk sacrificing one of the two types of innovation.

Besides general observations about scholars' conceptions of research, we were able to extract quality criteria from the scholars' notions of quality derived from the Repertory Grid interviews. Some of the criteria are already well known (e.g., *innovation*, *rigour*, *connection to society*) but we identified also some less-known criteria (e.g., *continuity*, *inspiration*, *topicality*, *openness and integration*, *connection between research and teaching*, and *intrinsic motivation*). For a detailed description of the results of the Repertory Grid interviews, see Ochsner *et al.* (2013).

The Repertory Grid method is a time-consuming method and hence the criteria extracted from the Repertory Grid interviews are only based on few cases. Therefore, we validated the results with a three-round Delphi survey in which we asked all scholars holding at least a PhD from one of the three disciplines under study at Swiss universities or at the member universities of the League of European Research Universities (LERU).

3.2. The Delphi survey

A Delphi survey is a «method for the systematic solicitation and collection of judgments on a particular topic through a set of carefully designed sequential questionnaires interspersed with summarized information and feedback of opinions derived from earlier responses» (Delbecq *et al.* 1975, p 10). The Delphi study was designed as follows: In the first round, the scholars completed the quality criteria derived from the Repertory Grid interviews and the literature; in the second round, the scholars rated the aspects of the criteria; in the third round, they rated quantitative indicators attached to the aspects of the criteria. Due to restrictions of space, we describe the method and results only briefly. For detailed information on the two first Delphi rounds see Hug *et al.* (2013), for information on the third Delphi round, see Ochsner *et al.* (2014).

In the first round, the scholars suggested new quality criteria and aspects and reformulated or complemented some criteria derived from the Repertory Grid interviews. We thus created a catalogue of criteria for research quality in the humanities that reflects the notions of quality of the humanities scholars and is formulated in the language of the scholars. The catalogue consists of 19 criteria further specified by 70 aspects. In the second round the scholars rated the quality aspects on a six-point scale, with 1 to 3 showing disapproval and 4 to 6 approval of the aspect. None of the aspects was rejected, thus the catalogue reflects adequately the research quality in the three disciplines. In order to identify those criteria for assessing research quality that find acceptance in the research community, we defined consensual criteria in each discipline. We classified a criterion as consensual when at least one of its aspects was clearly approved by a majority (i.e., at least 50 % of the discipline's respondents rated the aspect at least with a 5) and disapproved only by very few scholars (i.e., not more than 10 % of the discipline's respondents rated the aspect with a 1, 2 or 3). Tab. 1 shows the 19 criteria and an indication of consensuality in the three disciplines.

We found a set of eleven shared criteria that reached consensus in all three disciplines.[4] Moreover, six criteria were consensual in one or two disciplines and can be considered discipline-specific criteria. Finally, two criteria did not reach consensus in any discipline, namely *productivity* and *relation to and impact on society*. For the detailed description of the results, see Hug *et al.* (2013).

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

[4] Note, however, that for criteria that consist of several aspects, it is possible that in one discipline the first aspect is consensual while in the others the second and third is. Thus, not all of these criteria are specified with the same consensual aspects in the three disciplines. For example, the criterion *scholarly exchange* was specified differently in the three disciplines: In GLS, two aspects of this criterion reached consensus: «disciplinary exchange» and «interdisciplinary exchange»; in ELS, the two aspects «disciplinary exchange» and «international exchange» reached consensus; and in AH, all three aspects that build the criterion *scholarly exchange* reached consensus: «disciplinary exchange», «interdisciplinary exchange», and «international exchange».

7. Reflection, criticism $^{\text{GLS, AH}}$	14. Self-management, independence $^{\text{GLS, ELS}}$

Tab. 1 – All criteria with an indication of their consensuality in the three disciplines German Literature Studies, English Literature Studies and Art History.
Source: Ochsner *et al.* (2012)
Note: GLS = German Literature Studies; ELS = English Literature Studies; AH = Art History.

For the third Delphi round, we collected indicators for the consensual quality aspects. We conducted a comprehensive literature research in order to find as many quantitative indicators[5] as possible. We included a broad range of documents spanning from bibliometric and scientometric literature and government or institutional reports on how humanities are evaluated to grey literature on critiques of those procedures by humanities scholars. This resulted in a bibliography of literature on quality criteria and indicators for the humanities (Peric *et al.* 2012). We also collected indicators from the humanities scholars themselves who suggested indicators after the Repertory Grid interviews and in the first Delphi round. We thus found a huge amount of potential indicators, some very specific, some more vague. Because the scholars had to rate them in the third round, we clustered the indicators into 62 indicator groups and assigned them to aspect they can potentially measure. By assigning the indicator groups to the aspects of the quality criteria, it is possible to quantify the amount of aspects that cannot be measured quantitatively. We were only able to assign indicators to 23 of the 42 aspects that reached consensus in one of the three disciplines, which amounts to a share of 55% of aspects that can be measured quantitatively. If we look at the single disciplines, the share is even lower: In GLS, the share amounts to 53%, in ELS to 52%, and in AH to 48%. Thus, according to these results, quantitative indicators can capture only about 50% of the humanities scholars' notions of quality.

The rating of the indicators followed the same procedure as the rating of the quality aspects, again using the same six-point scale. Most indicators were approved by at least 50% of the scholars. However, in order to be used in assessments, the indicators should be accepted by the affected scholars, hence we again identified the consensual indicators using the same classification as for the aspects. Only very few indicators proved to be consensual: In GLS, 10 indicator groups reached consensus (12%); in ELS, only one indicator group was classified as consensual (1%); and in AH, 16 indicator groups were consensual (22%). We also asked a direct question whether the scholars think that it is conceivable that experts (peers) could evaluate the participants' own research performance appropriately based only on the quantitative data that the scholars had just rated. This question was clearly rejected by the respondents of all three disciplines (GLS: 88%; ELS: 66%; AH: 89%).

If we look at the response rates of the three surveys, we see that humanities scholars are willing to talk about quality criteria and discuss their notions of quality: in the first two rounds, the response rate was 30% and 28% respectively. This is a

[5] By indicators we mean clearly quantifiable entities such as number of publications, number of collaborations, or number of different methods used in research. We of course exclude ratings of the aspect by peers because this does not inform the peer but is simply a judgement of the aspect by the peer.

rather good response rate for a survey among professors. However, the third round reached a response rate of 11% during the same time as the second round and with a significant extension of the fieldwork period stopped at the maximum of 20%. This shows that humanities scholars are suspicious of indicator-based assessments of research quality, which was also reflected in their comments to the surveys (see Ochsner *et al.* 2014). Furthermore, the ratings suggest a similar interpretation. In all disciplines the grand mean of all rated aspects was considerably higher than the grand mean of all rated indicators. While 41% to 55% of the aspects reached consensus, only 1% to 22% of the indicators reached consensus. This points to the fact that only a small part of the humanities scholars' notions of quality can be adequately quantified or measured: From the 29 to 36 aspects that reached consensus, only 15 to 19 aspects, or 48% to 53%, can be potentially measured quantitatively. Finally, only 3% to 32% of the consensual aspects can be measured with consensual indicators (see Tab. 2). Because criteria can be specified by several aspects, the fraction of measurable criteria is somewhat higher but still less than two thirds. It has to be kept in mind, however, that when one aspect of a certain criterion is measurable but another aspect of it is not measurable, the criterion seems to be measurable but is missing an aspect entirely. Thus, it is only partially measurable and needs a peer to be evaluated adequately.

Discipline	Consensual		Theoretically measurable				Consensually measurable			
	Criteria	Aspects	Criteria		Aspects		Criteria		Aspects	
GLS	16	36	14	74%	19	53%	4	25%	7	14%
ELS	13	29	11	58%	15	48%	1	8%	1	3%
AH	13	31	11	58%	15	52%	8	62%	10	32%

Tab. 2 – Measurability of consensual criteria and aspects.
Note: GLS = German Literature Studies; ELS = English Literature Studies; AH = Art History.

3.3 Main results of the four studies

Summarising the results from the four studies (see Ochsner *et al.* 2014), we can conclude that an assessment of research performance by means of *indicators* will be met with resistance in the humanities: We have found that (1) only about 50 % of those quality criteria and aspects which are rated as most important can be measured with quantitative indicators. As long as 50% of the most relevant criteria and aspects cannot be measured with indicators, humanities scholars will be very critical of purely quantitative approaches to research assessments; (2) while most indicators are accepted for use in peer review-based assessments, a non-negligible minority does not approve of the use of most indicators to inform peer reviewers; (3) some indicators that are often used in evaluation schemes are not measuring research *quality* but differentiate between more «traditional» and more «modern» research, both of which can be of high or low quality and importance; and (4) *purely* indicator-based research assessments are disapproved of by a vast majority of the humanities scholars.

Concerning an assessment of research performance by means of *quality criteria*, however, we can conclude at the same time that humanities scholars are willing to reflect on research quality. They are ready to take part in the development of quality criteria if a bottom-up approach is chosen. A performance assessment on the basis of relevant criteria is possible if the humanities scholars are involved. We have found that (1) a broad range of quality criteria has to be applied to adequately assess research quality in the humanities; (2) there are shared criteria that are consensual in all disciplines that have been studied; (3) the disciplines should not be lumped together for evaluation purposes as we found discipline-specific criteria; and (4) with a certain amount of care, research indicators linked to the relevant criteria can be used to support the experts in research assessments (informed peer review).

4 Discussion of the project's results with regard to sociology

The above-mentioned results are based on data on German literature studies, English literature studies, and art history. However, we think that many of the findings translate quite well to sociology. When presenting the project we are getting often comments from scholars of such different disciplines as genetics, chemistry, or biology stating that they can identify with most of the criteria and certainly with the framework. While sociology is not a pure humanities discipline as the three subjects of the project, sociology is certainly closer to them than genetics or chemistry. Two authors of this article are sociologists; therefore, an interpretation of the results for sociology is a logical step.

The main finding of the Repertory Grid study, i.e. the four types of research, can easily be transferred to sociology. Certainly, these types of research are present in sociological research: Maybe there is a combination with methodology, the modern concept being more quantitatively oriented, the traditional concept being more theoretically oriented. However, it would be interesting to replicate this study with sociologists in order to find out whether these four types are the most prevalent types of research or whether there are more distinctions.

The results of the Delphi study can also be used for sociology. Of the 19 criteria for research quality, only *Fostering cultural memory* is not a straight-forward criterion for sociology. However, it might just be interpreted differently, changing the historical connotation for a current one and rename it maybe to *Fostering cultural conscience*. Furthermore, the criterion *Relation to and impact on society* may be needed to split into two criteria: *Relation to society* will be one of the most important criteria for sociology, while *impact on society* might be more important than for humanities disciplines but not necessarily be a quality criterion since sociological research can be of excellent quality but at the same time may *not (yet)* have an impact on society. However, in order to find adequate quality criteria for sociological research, we suggest to repeat the Delphi survey with a broad sample of sociologists, in the best case all scholars that are going to be evaluated by the criteria. This is a prerequisite of the framework that suggests a bottom-up procedure. The criteria from the project on quality criteria in the humanities, however, can be used as a starting point and, during the multi-round Delphi survey, the scholars will adapt the criteria and suggest new ones.

The main findings of the project certainly hold true also for sociology: A *purely* quantitative approach to evaluation would be detrimental to sociology as well, not only because there is no data base that can provide useful and valid data for all relevant aspects of such an exercise, but also because important aspects of research quality are hard or impossible to capture solely on the basis of quantitative indicators; in addition, sociological research always has a temporal aspect and what might seem important today might be irrelevant tomorrow and vice versa. Hence, a broad spectrum of research must be supported and mainstreaming must be avoided. Thus, the recommendations can also be used for sociology: (1) a wide range of quality criteria must be taken into account; (2) sub-disciplines might differ regarding certain quality criteria, which should be reflected in the evaluation procedure; (3) research must be evaluated by peers along criteria to reduce subjectivity and enhance transparency of the review process; and (4) with a certain amount of care, indicators can be used to inform peers on certain aspects and criteria. Therefore, we think that informed peer review is the preferred procedure to evaluate sociological research.

The project, however, has a quite substantial drawback: While the Delphi-study and the criteria and indicators have been replicated for the institution-level of a fourth discipline, Romance studies, the criteria were never used in an evaluation, an implementation thus is missing as of yet. However, The German Council of Science and Humanities conducted a pilot study for its research rating using informed peer review. The next section presents some insights from this pilot study.

5. The research rating of the German Council of Science and Humanities in Sociology

In Germany public media (Focus, Spiegel, Zeit) started to conduct teaching and research rankings in the early 1990ies. Debates about them were heated from their very beginning as they were apparently initiated for non-scientific reasons of attention seeking and print run increasing. Scientist were involved in these popular rankings as advisors giving input in how to collect and analyse data or as experts rating the reputation of different universities in surveys. However, the objectivity, reliability, and validity of published indicators and corresponding rankings remained open to discussion and were seriously questioned on methodological and substantial grounds. Nonetheless in practical terms being placed in the top or end section of one of these rankings could have dramatic consequences for universities and departments if government or university administration took the results at face value and used them as a rationale for distribution of monetary resources.

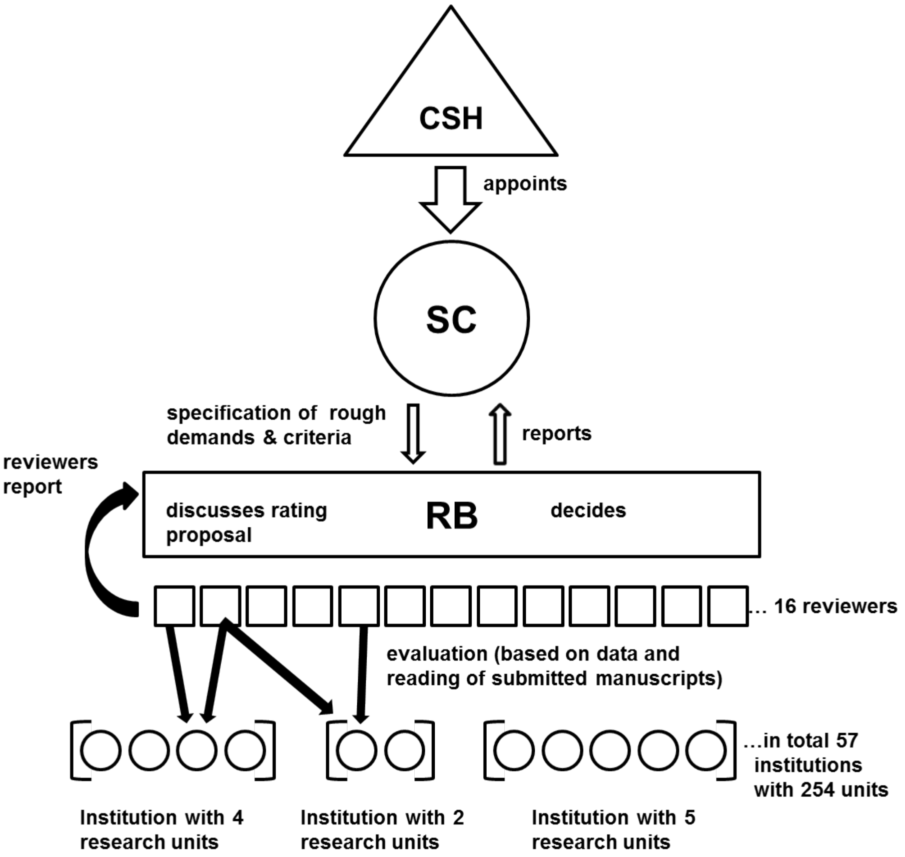
Against this background the German Council of Science and Humanities adopted in November 2004 recommendations for conducting research rankings and decided in July 2005 to develop and test methods for research evaluation in an elaborated pilot study for two selected disciplines of the social and natural sciences: sociology and chemistry. The decision was driven by the pragmatist view (Weingart 2015) that – due to huge practical relevance and demand – science indicators are here to stay but should be as scientific and reliable as possible. Therefore, to improve existing evaluation practices the Council of Science and Humanities (Wissenschaftsrat 2008b, 8; see also 2004) formulated three demands which the approach should meet at minimum: (a) *multi-dimensionality* including the criteria research (subdimensions: research quality; impact/effectiveness; efficiency), promotion of young scientists, and knowledge transfer (subdimensions: transfer in other domains of society; transfer and distribution of knowledge), (b) *informed peer review* combining quantitative and qualitative data with peer evaluations, (c) *rating instead of ranking* of research achievements. Due to these specifications, the exact implementation of these requirements varied between the two disciplines sociology and chemistry. For obvious reasons we will concentrate on the case of sociology in the following.

As can be seen in Fig. 2 the research rating involved several central actors: the Council of Science and Humanities (*CSH*), a steering committee (*SC*), and a review board (*RB*) (see also Wissenschaftsrat 2008a). The CSH initiated the rating and appointed a SC which consisted of members of the CSH, representatives of other scientific institutions, and additional experts on the subject. The SC coordinated the constitution of the RB, formulated rough requirements for the rating, and supervised operationalization of the indicators and review processes by sending a SC representative to the RB meetings. The 16 members of the RB were nominated by influential science organizations aiming to secure both the reviewers' outstanding reputation and a broad coverage of the various sociological sub-disciplines, methodological know-how and international expertise. These conditions were deemed as particularly important as both the implementation of the rating and the later evaluation of the research units almost exclusively depended on the RB.

At the conceptual level the RB had to decide on the unit of analysis and the operationalization of the abstract assessment criteria (see Neidhardt 2008 for details), particularly research quality as the single most important dimension of the rating. As regards the unit of analysis, sociology departments could divide themselves into so-called research units which were usually chairs but could also encompass larger units, in rare cases even whole departments. Operationalization of research quality followed the requirement of informed peer review. Two reviewers were assigned to each research unit and received information on publications, number of publications in peer-reviewed journals, third-party funded research projects, and self-assessed strengths and weaknesses. An additional source of information that is exemplary for this endeavour of research evaluation and distinguishes it from most other approaches is that reviewers actually had to read papers, book chapters, and books published by members of the research unit to assess quality. Although the use of citation data as an additional indicator was discussed extensively and controversially in advance, the RB decided – in contrast to the pilot study in chemistry – against it due to low and strongly varying coverage of sociological literature in the available citation databases.

Fig. 2 – Organization of Research Rating in Sociology
Source: Riordan *et al.* (2011, 152)

Based on the above-mentioned quantitative and qualitative material, the two reviewers then independently rated research quality on a five-point scale ranging from "excellent" to "unsatisfactory". Although no guidelines existed on how to weight the available indicators for an overall evaluation and reviewers thus likely gave different priority to the different data sources, qualitative interviews with members of the RB and quantitative analyses of the rating showed that the number of publications in refereed journals and the reviewers' quality judgements of the submitted manuscripts had by far the strongest influence on quality ratings (Riordan *et al.* 2011).


CSH
appoints
SC
specification of rough demands & criteria
reports
reviewers report
discusses rating proposal
RB
decides
... 16 reviewers
evaluation (based on data and reading of submitted manuscripts)
...in total 57 institutions with 254 units
Institution with 4 research units
Institution with 2 research units
Institution with 5 research units

Finally, reviewers submitted their ratings to the RB where they were discussed more or less extensively and modified according to the evaluation of the research unit by the RB. Extensive discussions were particularly likely if reviews differed in their assessment. Qualitative interviews with the reviewers indicate that such disagreements were not uncommon (Riordan *et al.* 2011), even though reviewers were encouraged to solve substantial disagreements before submitting their reviews leading to an artificial homogenization of ratings. Thus, the necessity to agree upon unambiguous ratings consumed a substantial amount of time of the RB and resulting ratings might be partly influenced by the course of discussions, particularly by the influential actors in the group, and the practical need to finally reach agreement due to time constraints.

Irrespective of this potential impairment of validity, the reception of the research rating by the CSH was mostly very positive. The community widely agreed that the rating constitutes the best available effort to rate research quality so far. Advantages were mainly seen in the bottom-up approach involving representatives of the discipline in the development of the tools and methods, the informed peer review accounting for both quantitative and qualitative data, and the use of a broad set of indicators taking the inherent complexity of the research object "quality" seriously.

Nonetheless critical voices remained. Concluding this section we want to highlight two lines of criticism which are informative as regards the aim of this paper to derive recommendations for research evaluation in sociology (see Münch 2009; Neidhardt 2009 for a discussion of additional aspects).

On the one hand, critics both from members of the RB as well as from sociological commentators centered on the enormous *temporal and monetary costs* of the project (Auspurg *et al.* 2015; CSH 2008a; Riordan *et al.* 2011). Roughly estimated, the ratings costs amounted 1.1 million Euros for administration, preparation, and board meetings, 219 working weeks of reviewers' time, and 432 working weeks of local university administrations providing the requested information for the rating. However, efforts to reduce the ratings to quantitative indicators and to develop a more parsimonious method were only partially successful. Riordan *et al.* (2011) found that only half of the variance in research quality ratings could be explained using quantitative data on peer reviewed publications, third party funding, and reputation. Auspurg *et al.* (2015) were able to further enhance explanatory power by reliance on more fine-grained data and the inclusion of self-conducted citation analyses but also did not succeed in perfectly reproducing ratings. Both studies hence illustrate that informed peer review actually added another, important component to the rating.

On the other hand, this contribution made by informed peer review has also been questioned and reframed as a subjective component impairing the objectivity of the rating. In face of the absence of clear guidelines how reviewers should weight the different available indicators, these critical voices do at least not fully lack substance. Although group discussions caused a certain degree of homogenization, every reviewer could, in principle, use his own evaluation standards to reach an assessment. One potential solution (see Riordan *et al.* 2011) to this problem would be to provide reviewers not only with quantitative indicators but also with a rating suggestion based on a weighting formula which is fixed ex ante (e.g. based on analysis of previous ratings). In a second step, the reviewers can adjust this proposal taking into account flaws of indicators and qualitative information. As a consequence, the time consumed by reading the submitted literature and screening projects – both previously reviewed by peers – could be reduced significantly. Another way to deal with the subjective component of ratings is to make divergence of positions explicitly visible in published ratings instead of reducing them to a seemingly precise point estimate of research quality (see Neidhardt 2006) – an insight also gained from the Swiss project «Developing and Testing Research Quality Criteria in the Humanities» presented in the preceding sections.

5. Discussion

What can we learn from these two projects for the evaluation of research performance in sociology? While the German research rating pilot study shows that an evaluation of sociological research using informed peer review is possible, it also points out weaknesses: first, the criteria for research quality are quite ambiguous and the weighting of the different sub-criteria is not clear; second, in order to compensate for the first issue, the ratings were discussed between reviewers before giving a final score which artificially increases inter-rater reliability and makes a power game likely; third, the criteria were specified top-down through the Council and specified by a small group of sociological experts, which may lead to mainstreaming. The results of the Swiss project «Developing and Testing Research Quality Criteria in the Humanities» provide solutions to these weaknesses by applying a framework for developing quality criteria that can readily be used in sociology: first, it names clearly specified quality criteria for research, most of which can be transferred easily to the field of sociology, thus making a quality judgement transparent; second, it suggests a measurement approach that clearly links indicators to criteria and thereby follows the insights of Thorngate *et al.* (2009) who state that judging something overall is usually inconsistent and not adequate for judging merit while judging sepearately according specified criteria reveals more reliable results[6]; third, it presents a method how to develop quality criteria in a bottom-up procedure, which is very important in SSH disciplines in order to account for different paradigms and lines of research, since there is no vision of linear progress of research but a coexistence of different lines of thought.

Therefore, for an evaluation of sociological research we suggest the following procedure:

1. Develop consensual quality criteria for sociological research surveying all evaluated researchers in order to include all lines of thought in the criteria. As a basis the humanities' quality criteria can be used as point of departure.
2. Search for indicators measuring the consensual criteria for research quality and let the scholars rate the usefulness of the indicators.
3. Create an evaluation sheet using the consensual quality criteria and indicators.

[6] In their words, the following steps are necessary to judge merit adequately: «Avoid the three 'I's: informal, intuitive, and inconsistent. Instead, decide what components or features of applications are important for judging merit, and judge each application by these features only. Make your judgement of each feature separately. Write down your separate judgements. Then add them up, weighing them in the same way for each application. The prescription is mechanical, likely more time consuming and less exciting than flying by the seat of your gut feelings. But it will produce noticeably fairer and better judgements of merit.» (Thorngate *et al.* 2009, p. 26)

4. Add criteria from other stakeholders, i.e., the other criteria for research performance included in the research rating (i.e., promotion of young scientists, knowledge transfer) and search indicators for the criteria and add them to the evaluation sheet.
5. Apply an informed peer review evaluation procedure similar to the German research rating using the above-mentioned evaluation sheet. In contrast to the German pilot study, reviewers' reading should be restricted to a reasonable amount of effort.
6. Do not provide overall ratings only but provide results for single criteria. If overall ratings are produced, use a transparent weighting procedure. However, it has to be kept in mind that research units have different goals or missions and thus, a single weighting might favour some missions over others, thus structurally discriminating some research units.

6. Conclusion: A Discipline-specific Approach to Research Evaluation for Sociology?

In times when society is facing huge global challenges, such as global warming and environmental catastrophes, and the hopes are high that such problems are solved by technological means (Beck 1992)[7], the SSH are not in the focus of public discussions. Especially, the critical and uncomfortable questions SSH research often poses is not high on the political agenda. That does not mean that sociology and other SSH disciplines have to give in and join in the neo-positivist hype of parametrically measurable research progress nor is it adequate for SSH disciplines to pout and refrain from any accountability to the society. Quite the contrary, the task of SSH disciplines, and especially so sociology, is to critically examine current hypes and mainstreams and analyse them according to the societal situation, detect spurious truisms and blind technological faith and propose alternative ways of how to guide university research in order to optimally serve societies in the long term. However, this critical role does not free the SSH from providing evidence for their effective and efficient use of public resources.

How can this balancing act between accountability on the one hand and critical thinking and opposition to blind faith in parametrically controlled research policy on the other hand be resolved? We think that the two projects presented in this article might point into a fruitful direction. The project on quality criteria in the humanities provides a framework for the development of quality criteria that are based on the notions of quality of the scholars. This is important since only the scholars can really judge what is good or important research; while it is also legitimate and important that the notions of quality of other stakeholders are taken into account, the scholars'

[7] See, e.g., the statement of then-president of the U.S. George W. Bush that „the way to meet this challenge of energy and global climate change is through technology, and the US is in the lead" in a speech to the US Global Leadership Campaign in 2007 (The Climate Group 2007). While Bush is not president anymore, his view is still widely shared in the public as well an in politics.

notions of quality are of utter importance to avoid negative steering effects or the so-called perverse effects of evaluation procedures.

Furthermore, the project showed that there is a quite large mismatch between the notions of quality of university administrators and evaluators on the one hand and the scholars on the other hand, at least for the humanities (see Hug *et al.* 2013; Ochsner *et al.* 2012). The fact that until now almost no evaluation procedures include scholars in the development of the evaluation protocol and if they do only a few scholars are involved and the criteria are defined in a top-down manner points to a problematic situation: The evaluations are not related to the scholars' notions of quality and, hence, the scholars cannot identify with the evaluation procedure. Therefore, it is no surprise that evaluation is met with opposition. The project shows that the scholars are ready to talk about research quality and to contribute to the development of quality criteria and the construction of measurement instruments.

However, the project has not yet been implemented to evaluate research. Here, the research rating pilot in sociology provides valuable insights: It shows that it is possible to evaluate sociological research using informed peer review exploiting qualitative and quantitative data. However, the analysis of the research rating pilot study also revealed the following issues: such an enterprise demands a huge effort by all involved stakeholders; differential weighting of indicators by the reviewers; and social processes potentially influencing the final rating. The latter two pitfalls were partially countered by plenary discussions of the peers to equal out a final rating; but a residual risk of gaming by peers, murky indicators and heterogeneous use of rating criteria certainly remains. Thus, the two projects complement each other: While the research rating shows that an evaluation is possible, the project on quality criteria in the humanities provides a solution how to overcome the problem of the somewhat arbitrary use of the criteria. It furthermore provides a framework how to define the criteria bottom-up instead of top-down thus enhancing the acceptance of the procedure in the research community.

We therefore suggest a combination of the two projects presented in this article for an adequate evaluation of sociological research that follows six points: (1) developing discipline-specific quality criteria based on the scholars' notions of quality using a bottom-up procedure; (2) search for indicators measuring these criteria and that are accepted by the scholars; (3) create an evaluation sheet using the criteria and indicators; (4) add the criteria for research performance (other than research quality) from the German research rating to the evaluation sheet; (5) apply an informed peer review procedure similar to the German research rating; and (6) do not present an overall rating but provide the results for the single criteria since missions of institutes might differ and a certain weighting might favour certain missions.

Finally, we want to point out that informed peers might not only contribute to research assessments by providing quality ratings. They could additionally give valuable input, e.g. by highlighting promising areas of specialization, potential for cooperation within and between the departments, or providing insights into strengths and weaknesses of research units. Evaluation is certainly much more than just a summative tool of control – why not use the different potentials it offers?

7. References

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

Andersen, H., Ariew, R., Feingold, M., Bag, A. K., Barrow-Green, J., van Dalen, B., . . . Zuidervaart, H. (2009). Editorial: Journals under threat: A joint response from history of science, technology and medicine editors. *Social Studies of Science* 39(1): 6-9. doi: 10.1177/03063127090390010702

Auspurg, K., Diekmann, A., Hinz, T., Naef, T. (2015). Das Forschungsrating des Wissenschaftsrats für die Soziologie in Deutschland revisited. [The Research Rating of the German Council of Science and Humanities on Sociology Revisited] Soziale Welt 66(3), forthcoming.

Beck, U. (1992). *Risk society. Towards a new modernity*. London: Sage.

Bell, D. (1973). *The coming of post-industrial society: A venture in social forecasting*. New York, NY: Basic Books.

Butler, L. (2003). Modifying publication practices in response to funding formulas. *Research Evaluation* 17(1): 39-46.

The Climate Group. (2007, June 1st). President Bush sends mixed signals on Climate Treaty ahead of G8. The Climate Group News Blog. Retrieved from http://www.theclimategroup.org/what-we-do/news-and-blogs/President-Bush-sends-mixed-signals-on-climate-treaty-ahead-of-G8/

Dávidházi, P. (2014). Exploring paradigms and ourselves. In P. Dávidházi, ed., *New publication cultures in the humanities. Exploring the paradigm shift*, 9-18. Amsterdam: Amsterdam University Press.

Delbecq, A. L., Van de Ven, A., and Gustafson, D. H. (1975). *Group techniques for programm planning. A guide to nominal group and Delphi processes*. Glenview: Scott, Foresman.

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

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

Geschwind, L., and Larsson, K. (2008). *Getting Pole Position - Pre reform research strategies in the humanities at Swedish universities.* Working Paper Series in Economics and Institutions of Innovation No. 140. The Royal Institute of Technology.

Gläser, J., Laudel, G., Hinze, S., and Butler, L. (2002). *Impact of evaluation-based funding on the production of scientific knowledge: What to worry about, and how to find out*. Expertise for the German Ministry for Education and Research. Retrieved from http://www.academia.edu/3065913/Impact_of_evaluation-based_funding_on_the_production_of_scientific_knowledge_What_to_worry_about_and_how_to_find_out

Hayek, F. (1937). Economics and knowledge. *Economica* 4: 33-54. Retrieved from http://www.virtualschool.edu/mon/Economics/HayekEconomicsAndKnowledge.html

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

Hicks, D. (2004). The four literatures of social science. In H. F. Moed, W. Glänzel and U. Schmoch, eds., *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems*, 473-496. Dordrecht: Kluwer Academic.

Hug, S. E., Ochsner, M., and 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., and 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. Retrieved from http: //www.psh.ethz.ch/research/publications/ijelp inpress.pdf

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

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

Lack, E. (2008). Einleitung – das Zauberwort "Standards". In E. Lack and Markschies, C., eds., *What the hell is quality? Qualitätsstandards in den Geisteswissenschaften*, 9-34. Frankfurt a.M.: Campus.

Lane, R. E. (1966). The decline of politics and ideology in a knowledgeable society. *American Sociological Review* 31: 649–662.

Lariviere, V., Gingras, Y., and Archambault, E. (2006). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics* 68(3): 519-533.

Lawrence, P. A. (2003). The politics of publication. Authors, reviewers and editors must act to protect the quality of research. *Nature* 422: 259-261.

Linstone, H. A., and Turoff, M. (1975). *The Delphi method: Techniques and applications*. Reading, MA: Addison-Wesley.

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

Münch, R. (2009). Die Konstruktion soziologischer Exzellenz durch Forschungsrating [The construction of sociological excellence through research rating.]. *Soziale Welt* 60(1): 63-89.

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics* 66(1): 81-100.

Neidhardt, F. (2006). Forschungsrating der deutschen Soziologie durch den Wissenschaftsrat. *Soziologie* 35: 303-308.

Neidhardt, F. (2008). Das Forschungsrating des Wissenschaftsrats. Einige Erfahrungen und Berichte. *Soziologie* 37: 421-432.

Neidhardt, F. (2009). Über Nachteile von Vorteilen. Ein Kommentar zu Richard Münch: „Die Konstruktion soziologischer Exzellenz durch Forschungsrating", in: Soziale Welt 60: 63-89. *Soziale Welt* 60(3): 325-333.

Nussbaum, M. C. (2010). *Not for profit. Why democracy needs the humanities*. Princeton, NJ: Princeton University Press.

Ochsner, M., Hug, S. E., and Daniel, H.-D. (2012). Indicators for research quality in the humanities: Opportunities and limitations. *Bibliometrie - Praxis und Forschung* 1, 4. URN: urn:nbn:de:bvb:355-157-7.

Ochsner, M., Hug, S. E., and Daniel, H.-D. (2013). Four types of research in the humanities: Setting the stage for research quality criteria in the humanities. *Research Evaluation*, 22(2): 79-92. doi: 10.1093/reseval/rvs039

Ochsner, M., Hug, S. E., and 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 Erziehungswissenschaften* 17(Supplement 6): 111–132. doi:10.1007/s11618-014-0576-4

Peric, B., Ochsner, M., Hug, S. E., and Daniel, H.-D. (2012). *AHRABi. Arts and Humanities Research Assessment Bibliography.* ETH Zurich. Retrieved from http://www.psh .ethz.ch/crus/bibliography/

Peters, M. A. (2007). *Knowledge economy, development and the future of higher education*. Rotterdam: Sense Publishers.

Plumpe, W. (2009). Stellungnahme zum Rating des Wissenschaftsrates aus Sicht des Historikerverbandes. In C. Prinz and R. Hohls, eds., *Qualitätsmessung, Evaluation, Forschungsrating. Risiken und Chancen für die Geschichtswissenschaften?*, 121–126. Historisches Forum. Berlin: Clio-online. Retrieved from http://edoc.hu-berlin.de/e_histfor/12/

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

Riordan, P., Ganser, C., and Wolbring, T. (2011). Zur Messung von Forschungsqualität. Eine kritische Analyse des Forschungsratings des Wissenschaftsrats [Measuring the quality of research – A critical analysis of the Forschungsrating of the German Wissenschaftsrat]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 63(1): 147-172.

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

Stehr, N. (1994). *Knowledge societies*. London: Sage.

Thorngate, W., Dawes, R. M., and Foddy, M. (2009). Judging merit. New York, NY: Psychology Press.

UNESCO (1998). *World Declaration on Higher Education for the Twenty-first Century: Vision and Action. And: Framework for Priority Action for Change and Development in Higher Education*. Paris: UNESCO. http://www. unesco.org/education/educprog/wche/declaration_eng.htm Accessed on 4. April 2015.

UNESCO (2005). *Toward knowledge societies. UNESCO world report*. Paris: UNESCO. Retrieved from http://unesdoc.unesco.org/images/0014/001418/141843e.pdf

Välimaa, J. and Hoffman, D. (2008). Knowledge society discourse and higher education. *Higher Education* 56(3): 265–285. doi:10.1007/s10734-008-9123-7

Weingart, P. (2008). Was ist gesellschaftlich relevante Wissenschaft? In A. Schavan, ed., *Keine Wissenschaft für sich. Essays zur gesellschaftlichen Relevanz von Forschung*, 15-24. Hamburg: Edition Körber Stiftung.

Weingart, P. (2015). Nostalgia for the world without numbers. *Soziale Welt* 66(3), forthcoming.

Wissenschaftsrat. (2004). *Empfehlungen zu Rankings im Wissenschaftssystem. Teil I: Forschung*. Köln.

Wissenschaftsrat. (2008a). *Pilotstudie Forschungsrating. Empfehlungen und Dokumentation*. Köln.

Wissenschaftsrat. (2008b). *Forschungsleistungen deutscher Universitäten und außeruniversitärer Einrichtungen in der Soziologie*. Köln.