

‘21 Grams’: Interdisciplinarity and the Assessment of Quality in the Humanities

Klaus Stierstorfer and Peter Schneck

Abstract In their joint contribution, the president of the German Association for English Studies (Deutscher Anglistenverband), Klaus Stierstorfer, and the president of the German Association for American Studies (Deutsche Gesellschaft für Amerikastudien), Peter Schneck, describe the central motivations behind the decision to actively support the pilot study for the research rating of the German Council of Science and Humanities (*Wissenschaftsrat*) despite some fundamental skepticism among the associations’s members. On the basis of five basic propositions—different in each argument—they both insist that the assessment of research quality in the humanities inevitably requires the central involvement of the disciplines assessed in order to reflect on and formulate the central categories, standards and procedures best suited for such assessments. Such a process must take into account the complexity of research processes and results in the humanities whose qualitative dimensions cannot be fully measured by quantitative methods.

1 Rating Research: Who Needs It, and What Is It Good For? (by Klaus Stiersdorfer)

Research rating and ranking is happening now, at least in German academia in my experience, and it has been growing in the anglophone countries, with which I deal professionally, at an alarming pace and as a kind of *menetekel* for whatever other countries may be planning to do in the future. This is why, and here is my *first thesis*, research rating and ranking cannot be avoided at present. If my first thesis is accepted, then it is worth exploring what it looks like at present in the humanities.

K. Stierstorfer (✉)

English Department, Westfälische Wilhelms-Universität Münster,
Johannisstr. 12–20, 48143 Münster, Germany
e-mail: stiersto@wwu.de

P. Schneck

Institute for English and American Studies, Universität Osnabrück,
Neuer Graben 40, 49069 Osnabrück, Germany
e-mail: peter.schneck@uni-osnabrueck.de

© The Author(s) 2016

M. Ochsner et al. (eds.), *Research Assessment in the Humanities*,
DOI 10.1007/978-3-319-29016-4_16

Most rating and ranking systems I have come across involve any one of the following procedures: peer reviewing of research publications; measuring of quantities of publications; opinion polls on the research reputations of individual institutions and agencies, or any combination of the three. I will not dwell on the latter two as they seem the most obviously inadequate for rating in the humanities, but do want to broach briefly the topic of peer reviewing which is widely seen as the fairest and most reliable tool of the three. The problems I see with it in its current form have, however, to do with fairness and transparency. With most reviewing procedures, the image of the administration of justice attributed to the so-called dark middle ages seems appropriate. There is little transparency in the application of pre-specified criteria; the actual judges (peer-reviewers) are still shielded from the person under review (the defendant) by the inquisitorial screen of anonymity; and the defendant has hardly any means of recourse to plead his or her case when the verdict is negative. This leads to a situation when most researchers in my field, at least where they have the choice, avoid such reviewing processes as the impression (true or not) arising from this black-box juridical system is imputations of favouritism, nepotism and the pursuit of non-scholarly, strategic or political ends under cover of this anonymity. The much-propounded 'blind' or even 'double blind' peer-review really does not mean that justice is iconically blind (as she should be) as to the addressee of her ministrations (projects under review are all too easily attributable in small research communities), but that reviewees are blinded (as they should not be) as to who is their judge and on what grounds their verdict is really passed. Hence, on this ground and many others, my *second thesis* is, current research rating needs improvement if we want to stick to this practice.

How such improvement can be brought about is, of course, the philosopher's stone here, but before its quest is started, the issue of the necessity of rating research in the humanities in the first place must be dealt with. As this is a short statement, the answer suggested here—which is also the prevalent opinion in the Deutscher Anglistenverband and the official position of its presidency and council—is essentially twofold. First, and this is my *thesis number three*, we need research rating because it is there or, more precisely, scholars in the humanities and their societies and associations should get involved in research rating because they are being practiced at the moment; trying to make oneself heard and get involved in establishing the fairest and best practice possible seems reasonable if not logical and unavoidable. Experience has shown that outright refusal to join the discussion does not help to avoid rating and ranking but produces bad, because inexpertly designed procedures.

Why then has research rating been established in the first place? The simple answer is: money. In the progressive commercialization and economization (if that is a word) of our academia, the political focus on money invested in research has been immense, and hence a mechanism for its distribution was sorely needed. On a simple, outcome-oriented economic model, the logical system is to put money where the best outcome is. Hence the idea to measure research outcomes and put most money where the best outcomes can be registered or at least expected. Thus,

research rating is primarily an administrative tool that has to do with investing and distributing limited funds for research. The crux of defining and comparing precisely these outcomes has long been overlooked or neglected. In the most negative reading, the whole process only shifts the problem to another scenario.

Does rating have any benefits for the scholar or researcher in the humanities? My answer is: No, surely not primarily. In a slightly more personal explanation I would stress that I am not interested in knowing whether my colleague X's new monograph is better than mine, and if so how much on a scale from 1 to 10, neither do I need to know whether colleague Y's article in a field I am interested in is rated high or low before I read it as the specific questions I bring to it in my specific research context may differ from quality criteria, nor do I have any desire to be informed whether my publications of the last 5 years are to be graded as 5, 6 or 7 on a scale of 1–10. For purposes of orientation which books and articles to look at in the first place, I have sufficient bibliographic and reviewing tools at hand which are well-established and efficient, even if not easily translatable onto scales from 1 to 10. Thus, my *thesis number four* says research rating is next to useless for the purposes of research itself and time spent on it would be immeasurably better spent on such research.

But, if we cannot reasonably avoid research rating at present, and even if it seems pointless for research, can we gather some lateral benefits from it, although it remains primarily superfluous in the eyes of the researcher? Here my *fifth thesis* is yes, research rating could be devised in such ways that a number of collateral benefits might accrue. Again, a lot of creative thinking could and must go into this question, but I only want to focus on one possible aspect here, that is disciplinary self-reflection. By thinking about criteria how quality of research can be measured and understood, scholars in the humanities will be forced to reflect on their current standards and aims of research and how to define them. This process can help individual disciplines to identify where they stand as a discipline and where they might want to be going in the future, as the steering function of rating procedures can hardly be underestimated. While rating may thus be a good thing for initiating and furthering discussions in disciplines and professional associations such as our Anglistenverband, this does not mean that these guidelines agreed on for the entire discipline are really a good yardstick for individual instances of research. Especially in the humanities we know too well that innovative research is, as Thomas Kuhn, Paul Feyerabend and others have argued, all too often not the kind that is immediately recognizable as such by current disciplinary standards.

Conclusion: Although the benefits seem lateral at best, rating of research is nothing that the humanities can easily avoid at the moment, so it seems better to embrace the discussion leading to its implementation with full commitment in the service of the colleagues for whom we speak in our various associations. The search for a fair, transparent and equitable rating system in the humanities may be a quest for the philosopher's stone, but that does not mean that, under current circumstances, we should not try as best we can.

- Thesis 1: Research rating and ranking cannot be avoided at present.
 Thesis 2: Research rating and ranking needs improvement if it is to be continued.
 Thesis 3: Research rating and ranking is needed because it is there.
 Thesis 4: Research rating and ranking is useless for research itself.
 Thesis 5: Research rating and ranking can produce collateral benefits.

2 ‘Weighing the Soul’ of the Humanities (by Peter Schneck)

Let me begin with a little historical anecdote: On April 10th 1901, Dr. Duncan MacDougall, a medical researcher from Dorchester, Massachusetts conducted an experiment to determine the physical existence of the soul. Placing six moribund patients on specially designed scales, the doctor tried to quantify the soul by measuring the weight of the patient’s bodies shortly before and shortly after their death. Comparing the difference between the two assessments, MacDougall found that each of the patient’s bodies lost precisely the same amount of weight, which was around three-fourth of an ounce, or about 21 g. Since he could think of no other explanation for the difference in weight, the doctor concluded that in the moment of death the soul had left the patient’s body; thus the soul not only existed, it’s weight could also be pinned down rather precisely at 21 g—which is probably less than one would have expected for such a ‘weighty’ phenomena as the soul given its metaphysical significance throughout our cultural and spiritual history.

While MacDougall’s weighing of the soul may be regarded as one of the countless, equally eccentric and futile attempts to measure the immeasurable—an attempt which is symptomatic for a climate of extreme scientific optimism and positivism around the turn of the 19th to the 20th century—it may nevertheless be instructive for understanding the current struggle between those who propose to assess, rate or quantify the quality of research in the humanities with objective methods of weighing and measurement, and those who think that this attempt would amount to a futile ‘weighing of the soul’—that is, an absurd, useless and basically misguided exercise.

The anecdote may be instructive in the context of our discussion for more than one reason, but before I turn to the problem of measuring the immeasurable in the main part of my short remarks, let me clarify a few things from the start.

On the one hand, I am talking to you as a humanities scholar whose teaching and research has been subjected to various forms of quality assessment by an extended number of parties: by other scholars, both from my own field and from other neighbouring fields, by various university administrations and committees, by the review boards of various national and international research funding agencies and institutions, as well as by various assessment boards of the federal state and on the national level. Last, but not least, I have also been asked numerous times to assess myself not by mere introspection, but in a more regulated and prescribed form.

Ever since my performance as a scholar became the subject of a standardized questionnaire for the first time in 1984 at a leading American university, quality

assessment in all its different forms has remained an inescapable part of my scholarly and professional existence.

From this perspective of personal experience as an individual scholar, my feelings towards the continuous increase of assessment processes, the growing repertoire of procedures and protocols, as well as in face of the various institutional and public ratings and rankings in which they result—my sentiments in regard to all this excessive monitoring and controlling could best be described by quoting Elvis Costello: 'I used to be disgusted, now I'm trying to be amused.'

To put it a bit more precisely; even though over the last decades I have come to experience and somewhat grudgingly accept an astounding number of forms of quality assessment and rating processes in the humanities as inescapable, that does not in any way mean I deem them indispensable. On the contrary, as an individual scholar in the humanities, I have increasingly come to doubt and, in fact, severely question both the essential necessity and the positive effect of quantifying ratings and rankings in and for the specific form of research that is being done in the humanities. To put it bluntly: I find it rather hard, if not impossible, to conceive of any process of calculating and expressing in numbers the difference in quality in regard to research in my field that would actually have any impact other than to regulate it (mainstreaming it, prescribing it) by rather artificial measures of comparison.

Thus, the only thing I learned so far from the ongoing and increasing assessment and quantification of research quality in the humanities is this: Whatever can be quantified, will be quantified—and if it hasn't been quantified yet, it will be quantified eventually. So I agree with my colleague Klaus Stierstorfer that if ratings and rankings are here to stay there is hardly a way to avoid them—but that doesn't make them more useful or attractive.

As Werner Plumpe, the president of the Association of German Historians has recently argued with considerable gloom, the sheer pressure of and rush towards ratings and rankings may eventually even reach the unquantifiable soul of the humanities: enforcing quantifying methods on central dimensions of research that cannot and should not be measured and expressed by numerical values only.

There are good reasons to accept some of the more convincing arguments that Plumpe brings forth against rating and ranking procedures in the humanities based on quantification, and I easily agree with most of his criticism and scepticism in regard to the uselessness of quantification for the acknowledgement and assessment of research quality in the humanities. There may also be good reason to subscribe to Plumpe's scepticism that there is a great danger of misinterpretation, or even misuse by third parties, resulting from the suggestive comparability of mere numerical values—something that must be seen as a central concern given the fact that all these numerical values are (increasingly) used as evidence and arguments for the distribution of resources by universities, by the state (both on the federal and the national level) and by third party sponsors like research foundations (both national and international).

And yet there is something slightly uncomfortable and counterintuitive in this well-stated arguments, and even though I share both the reasoning and the sentiment to a certain degree, eventually the conclusions I draw from the current situation are rather different.

In fact, while Plumpe (and the majority of his colleagues in the association of German historians) have emphatically decided not to take part in the preparatory study initiated by the *Deutscher Wissenschaftsrat* (German Science Council), the *Deutscher Anglistenverband* (German Association for English Studies) and the *Deutsche Gesellschaft für Amerikastudien* (German Association for American Studies) have decided to do just that—despite the fact that we share the fundamental scepticism of our colleagues from the history departments about essential aspects of rating and ranking in the humanities per se.

But there are several reasons for this decision, and some of them have already been presented in summarized form by Klaus Stierstorfer. My task in the following parts of these short remarks will be to describe the specific perspective of the association which I represent in respect to the projected study but also in general. This perspective is particularly characterized by the strong interdisciplinary traits of the research that is being done in German American Studies (or more precisely *Amerikaforschung*).

I said there is something counterintuitive or uncomfortable about the complete rejection of the quantification of research quality in the humanities. While there are, as I readily acknowledged, good arguments against quantification as such, these arguments should not (and probably cannot) obscure our perception of the high degree of assessment by quantification that is already in practice in the humanities—in fact, one could argue that it is quantification which dominates the assessment of individual research in the humanities from the very start until the moment when one has successfully become installed by a committee—on the basis of other assessments—as a university professor. In other words, the professional success in the academic field of the humanities is essentially based on ratings and rankings and other accepted assessment procedures within the field. While these procedures are of course not completely based on or expressed in numbers, one cannot overlook or deny the existence and significance of quantification within these assessment practices in the humanities.

This is not meant to be a rhetorical move—I don't think that my colleagues from the history departments would deny the existence of quantification and ranking procedures within their field and as part of their own daily academic practice. Yet while they would readily attest this, they would probably also insist that all this rating and ranking is only done by peers, and based on meticulous and highly reflected methods of reviewing and critical acknowledgment.

However, if there are procedures of assessment involving quantification established in the field as such, it is obvious that the argument against quantification in the humanities is either a universal one—then it either works or it doesn't; and if it does not work because it can never capture the 'soul' that is the real quality of research done in the humanities, then one should drop it altogether: no more grading of research papers, no more graded forms of assessment for doctoral theses on a standard scale (even when using the Latin terms this is still a quantification of quality), no more ranking lists in committees etc.

On the other hand, if the argument is *not* a universal one (and I don't think it is or can be) then the debate should not be about quantification at all, but, rather about consensual standards of comparison and accepted and/or acceptable conditions of

assessment which make the quantified expression of quality not only possible but even desirable for pragmatic reasons (and a number of factors have been named already during our discussions: the sheer increase of scholarship and its ever growing diversity, international competition and funding schemes within the common European research area etc.).

Another aspect that also tends to be neglected in the debate (and I am only talking about the debate about the pros and cons of assessment and quantification of research quality) is the increasing development of new transnational research and study programs, especially on the young researchers level, i.e. joint doctoral programs within the humanities offered and designed by institutions from different countries across Europe. One of the most challenging tasks is to find a common denominator for the assessment and control of the quality of the study programme and the research of the individual researcher. The same is true for international research consortia: there has to be a shared understanding of the quality standards that would guide and make possible the assessment of the research to be conducted. This is an aspect that is of special significance for American Studies as a discipline and a field of research, since in contrast to English Studies (*Anglistik*), American studies has been conceived from the start as a fundamentally interdisciplinary enterprise. In fact, one could argue that American Studies is the name for research done across the boundaries of various disciplines and since its inception this understanding has always led to intense struggles about the proper methodologies, the common concepts, the shared terminology and, last but not least, the commonly accepted standards of quality in research between all participating disciplines.

Therefore, from the perspective of the scientific community involved in research in American Studies in Germany, the participation in the proposed pilot study by the Science Council has both professional, strategic and pragmatic reasons. On the one hand, it presents a calculated step to maintain a central role in the debate and definition of standard criteria and procedures to assess the quality of research done within the discipline. At the same time, it acknowledges the increasing dynamics of collaborative research agendas across disciplines and across national research areas, which are at the heart of the current struggles for standards, criteria and indicators that may be transferable and commonly acceptable at the same time.

In conclusion, one could summarize the motivational aspects that has guided the decision of the DGfA as follows:

- To assure the active participation and indispensable involvement of the field/scientific community in the process of defining standards and criteria of assessment for the quality of research within the field
- To allow for an open and ongoing debate about standards and criteria within the field and across the disciplines ⇒ interdisciplinary research community
- To actively take on responsibility for the development of common standards and criteria
- To make transparent and critically debate existing standards
- To develop common consensual standards across disciplines that meet the requirements and the dynamics of today's interdisciplinary research in the humanities

Let me end with a caveat: The process certainly is not an easy one, and we do not think that we should drop our guard by replacing our healthy scepticism with a naïve trust in the evidence of numbers and graphs. As has been emphasized, the process of arriving at the shared and commonly accepted standards and criteria I talked about can only be a mixture of top-down and bottom-up approaches and perspectives. To return to my initial historical anecdote: Weighing the ‘soul’ of the humanities should not simply be translated into a question of grams and ounces, nor should the wealth and diversity of humanities research be assessed as a *quantité négligeable*.

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

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