

Rating Research Performance in the Humanities: An Interim Report on an Initiative of the German *Wissenschaftsrat*

Christian Mair

Abstract The author, a professor of English linguistics at Freiburg University, was a member of the German Council of Science and Humanities (*Wissenschaftsrat*) from 2006 to 2012 and, in this capacity, was involved in this advisory body's rating and assessment activities. The present contribution focusses on issues arising in the rating of research output in the humanities and is informed by his dual perspective, as planner and organizer of the ratings undertaken by the *Wissenschaftsrat* and as a rated scholar in his own discipline, English and American Studies.

Over the past decade, rankings—whether home-grown or international—have had a profound impact on higher education in Germany, although the way in which they are being used tends to reveal a degree of tactical short-termism if not downright cynicism. Institutions which come out on top rarely question the procedures by which the welcome result has come about, but are happy to make the most of the free advertising provided. Those not placing so well do not take the result as a motivation for systematic self-study, but rather look to convenient quick fixes which, they hope, will enable them to move ahead in the league tables the next time around.

Within the academic community, rankings have become an informal mechanism of reputation assignment which is not entirely unproblematical but which—at least so far—has had few tangible consequences in terms of structural reform or strategic planning. In wider society, rankings may have some influence on students' and parents' choices of institutions and programmes, though there is as yet no evidence that they are a crucial factor in such decisions, which is probably not a bad thing, either, as the criteria which rankings are based on usually have no very direct bearing on the needs of first-year undergraduates.

In this situation, the German Council for Science and Humanities (*Wissenschaftsrat*), decided to carry out an analysis of the extant rankings in 2004. Its main finding was that the systematic, comparative and often quantitative assessment of research performance had come to stay, but that the methods and criteria employed by the

C. Mair (✉)

English Department, University of Freiburg,
Kollegiengebäude IV, Rempartstr. 15, 79085 Freiburg, Germany
e-mail: christian.mair@anglistik.uni-freiburg.de

various rankings were usually not fully transparent and that, moreover, the relevant academic communities had little say in how they were framed (Wissenschaftsrat 2004). The *Wissenschaftsrat's* suggestion for improvement was to develop a rating system in which research output in a particular field would be evaluated comparatively on the basis of criteria developed in consultation with the relevant research community.

As such a rating exercise involved substantial preparation and considerable investment of labour from all parties concerned, pilot studies were deemed essential. The concept was first put to the test on a nationwide scale in the fields of chemistry and sociology—and proved generally workable in both fields, despite their very different objects and methods of investigation (Wissenschaftsrat 2008). Encouraged by this, in 2008 the *Wissenschaftsrat* decided to carry out two further pilot studies, which were supposed to conclude the test phase, and then make the new instrument available on a large scale. The disciplines selected for this second phase of pilot studies were electrical engineering and informatics, on the one hand, and history, on the other. While the engineering pilot was successfully completed in June 2011 (Wissenschaftsrat 2011), the history pilot ended in a deadlock between the *Wissenschaftsrat*, representing the advocates of measuring research output in the humanities, and the *Verband deutscher Historiker* (Association of German Historians), representing the research community to be rated. As some of the debate was conducted in the culture pages of major national broadsheets, it generated an amount of publicity which, at least for the *Wissenschaftsrat*, was not entirely desirable in such an early phase of testing the new instrument.

On the other hand, it is the high profile that this episode gained which makes it instructive and interesting beyond its immediate academic-political context. In the remarks which follow I shall therefore take it as a starting point for a discussion of the particular difficulties—objective and subjective—surrounding the comparative measurement and evaluation of research output in the humanities and to present the *Wissenschaftsrat's* line of argumentation on this important issue.

In principle, there is no reason why a rating exercise as envisaged by the *Wissenschaftsrat* should be offensive to scholars' sensibilities in the humanities. After all, in its critique of the current situation, the *Wissenschaftsrat* points out the superficiality and lack of transparency of most existing *rankings* and makes the point that any instrument used to measure research performance needs to fit the discipline it is applied to. The *ratings* which the *Wissenschaftsrat* (Wissenschaftsrat 2004, pp. 33–43) suggests as the appropriate alternative are supposed to:

- be conducted by peers who understand the discipline they are evaluating,
- apply criteria specific to the field being evaluated,
- evaluate research output in a multi-dimensional matrix rather than a simple rank list,
- differentiate between achievements of individual 'research units' representing the field at a particular institution.

The last-mentioned criterion in particular should be welcome to scholars in the humanities, who define their research agenda very much as individuals and would resent their achievement to be levelled into departmental averages in a rating exercise.

While the preparation for a rating may involve a certain degree of nuisance and the rewards may be uncertain, the overall design features should find a sympathetic audience among humanities scholars. As a principle, informed peer review is accepted in the humanities as in other academic fields. It determines what gets published or who gets selected for positions, and at conferences or similar forums humanities scholars certainly enjoy the opportunity of showcasing their work and benefit from constructive criticism and advice extended by peers as much as anyone in academia.

What then is the cause of the hostility towards the rating exercise articulated by German historians (or at least their spokespeople in the association)? At least in part, I would contend, the conflict was due to a communication problem. Rankings and ratings, including the *Wissenschaftsrat*'s, tend to be presented in a discourse of administrative control and neoliberal new public management which makes many scholars in the humanities suspicious from the very start. Their main experience with this discourse has so far been gained in the defensive rather than the offensive mode. Strategic planning of research has been experienced as increasing regimentation, increasing pressure to produce largely bureaucratic documentation and—in the extreme case—withdrawal of personnel and resources. That the humanities stand to gain from strategic planning—for example through improving career prospects for young scholars or claiming their due place in expensive digital infrastructure projects—has been less obvious by comparison. In this situation, any type of ranking or rating is thus likely to be considered as part of an unhealthy trend towards the bureaucratization, commercialization and commodification of higher education.

Let me briefly illustrate the type of miscommunication I have in mind with one of the *Wissenschaftsrat*'s own formulations. Both internally and in several external presentations it has defined the purpose of the rating exercise as 'Unterstützung der Leitungen bei strategischer Steuerung durch vergleichende Informationen über Stärken und Schwächen einer Einrichtung' [supporting administration in its strategic planning by providing comparative information on strengths and weaknesses of a unit] (see *Wissenschaftsrat* 2004, p. 35, for a published version). Putting things in this way is certainly not wrong, but—in view of what has been said above—clearly not the best way of enlisting the support of the scholars whose participation is required to make the exercise a success. While the formulation allows us to infer the threats that may accrue from under-performance, it is not very explicit on the rewards to be derived from co-operation, both in terms of a particular field and the individual researcher. Researchers in the humanities are generally individualists and therefore sceptical about higher-level strategies of promoting or regimenting their scholarly creativity. They are competitive but not necessarily in the corporate sense of championing their institution. Successful teams are more likely to be composed of scholars working in different places than of colleagues belonging to the same department.

In his public debate with the *Wissenschaftsrat*, Werner Plumpe, the renowned historian and president of the German Historians' Association at the time, emphasizes exactly these points in his critique of the proposed rating (Plumpe 2009). Quan-

tification and standardization, he claims, may suggest the simplicity that political decision makers in university administration and higher-education bureaucracies crave, but this simplicity is a spurious illusion [in his own words (Plumpe 2009, p. 123): ‘teilweise quantifizierte, immer aber parametrisierte Informationen für politische Diskussions- und Entscheidungsprozesse, die gemessen an der Realität des Faches unterkomplex [sind]’]. An even bigger illusion is the assumption that success in research is the result of stimuli set in the system or advance planning of other kinds [‘Illusion, Wissenschaft lasse sich parametrisch durch das Setzen bestimmter Anreize steuern’] (Plumpe 2009, p. 123). According to Plumpe, a standardized rating is not merely useless but counter-productive, because it encourages scholars to focus on meeting the targets of the system rather than the often different standards of professional integrity and scholarly excellence [‘Herausbildung und Verfestigung strategischer Verhaltensweisen, die zumindest in den Geisteswissenschaften die akademische Kultur zerstör[en]’] (Plumpe 2009, p. 123). In short, the field of history does not owe it to itself or anyone else to take part in such a problematical project:

Das Fach habe es aber weder nötig noch sei es im eigenen Interesse verpflichtet, die gefährlichen Illusionen der derzeit politisch hegemonialen Strömungen zu bedienen.

[Neither self-interest nor external necessity forces the community to pander to the current hegemony’s dangerous illusions.] (Plumpe 2009, p. 123)

As we see, the opposition is comprehensive and formulated with considerable rhetorical investment. A compromise between the Historians’ Association and the *Wissenschaftsrat* was not possible. While the opponents of rating could claim a victory and were in fact heralded as champions of academic freedom in some of the press reportage, the *Wissenschaftsrat* found itself in a bit of a fix. In an atmosphere thus charged, it would have been futile to just move on and approach another field in the humanities to enlist its co-operation. The way out of the impasse was the creation of a working group bringing together a wide range of scholars in the humanities—from philosophy through literature and linguistics all the way to area studies, including the *kleine Fächer*, highly specialized areas of enquiry such as cuneiform studies or Albanology, which in the German system are frequently incorporated as micro-departments consisting of one professor and one or two lecturers or assistants. This interdisciplinary working group was expected to assess the suitability of the *Wissenschaftsrat*’s proposed rating to the humanities and suggest modifications where it held them to be necessary.

The present author was privileged to be part of this working group and can testify to the open atmosphere of discussion which made all participants aware of the wide range of research methods and theoretical frameworks found in the contemporary humanities. Most members of the group eventually (though not initially) accepted that rating research output according to the *Wissenschaftsrat*’s model was possible in the humanities, might even have beneficial side effects for maintaining and developing quality in the individual fields, and be a means of securing the humanities’ general standing in the concert of the other disciplines. Intense disputes, however, arose every time concrete and specific standards of evaluation had to be formulated. Early drafts

of the recommendations contained fairly contorted passages on the relative merits of the traditional scholarly monograph as against the co-authored paper in a peer-reviewed journal, on the need to encourage publication in English while safeguarding the continuing role of national languages as languages of scholarly publication, and so on. About half way through the proceedings, participants realized that the best way to solve these issues for the time being was to defer them, i.e. to state the problem but to expect the solution to emerge from subsequent discussions in the individual research communities concerned. The recommendations thus grew slimmer, but improved from meeting to meeting as discussants realized that they had to aim for a mid-level of abstraction and leave the concrete fleshing out of standards to the discipline-specific experts. In a slight departure from existing *Wissenschaftsrat* rating conventions, the following three dimensions of evaluation were proposed (*Wissenschaftsrat* 2010, p. 20):

- Forschungsqualität [quality of research]
- Forschungsermöglichung [activities to enable research]
- Transfer von Forschungsleistungen an außerwissenschaftliche Adressaten [transfer of research achievement into non-academic domains].

To accommodate possible slower rates of maturation of research results and slower dissemination and reception, the standard five-year cycle of assessment was extended to seven years. It will be a major challenge to rating exercises based on these recommendations that qualitative measures were prioritized over quantitative ones. Thus, for the assessment of research quality, each ‘research unit’ will be asked to submit the five publications from a relevant seven-year period which are considered most important. The technical designation ‘research unit’ is intended to make possible reporting at a contextually appropriate level intermediate between the individual researcher and an institutionalized administrative unit such as a ‘department’ or an ‘institute’. In a traditional German humanities context, this level would typically be understood to be the ‘*Professur*’, i.e. the professorial ‘*Lehrstuhl*’ or chair comprising the professor and his or her assistant(s). Discussions in the working group suggested that some academics would be quite happy to dispense with this intermediate layer in practice and submit five publications per professor, thus defining the relevant unit of documentation as the individual advanced researcher. Clearly, those responsible for the next pilot study will take the opportunity to clarify this contested issue against the background of their discipline.

The most salient feature of the proposed procedure when compared to rating in the natural sciences is that quantitative information, such as number of publications, will play an ancillary role only. This is justified, though, in view of the fact that standard quantitative indicators such as impact factors or citation indices are only marginally relevant in the humanities. One additional dimension of evaluation which it was judged necessary to include in rating research quality similarly defies quantification, namely a researcher’s scholarly reputation. In view of reputation’s auratic and intangible nature, those members of the working group who would rather not have included it as a criterion will probably take consolation from the fact that it will not have the same importance for all disciplines and certainly not for all individuals.

One of the more convincing ways of measuring reputation was considered to be taking note of the award of prestigious research prizes, such as the German Research Foundation's (DFG) Leibniz Award. Those who advocated considering reputation emphasized that it was not something which lapsed in the seven-year time-window relevant for measuring performance.

The term *Forschungsermöglichung*, not conventionally established, was used as a cover for activities which did not necessarily result in research publications by the principal investigator, but promoted research activities in a wider sense. Typical examples would include contributions to the development and maintenance of important research infrastructures, such as digital text archives or linguistic corpora, acquisition of external funding for research teams providing career opportunities for young researchers, etc. The distinction between the two dimensions of quality and enabling was felt necessary as (a) the mere fact that research in the humanities was funded by external grants did not mean that it was necessarily of high(er) quality and (b) across virtually all humanities disciplines the individual researcher was considered to be in a position to produce first-rate research unaided by teams or expensive infrastructure.

Transfer was expected to take forms appropriate to the individual disciplines, ranging from involvement in exhibitions and museums (art history) via in-service teacher training (foreign languages) to consulting activities (philosophical ethics).

As I briefly hinted at above, it is also very interesting to note the points on which the general recommendations are silent. They do not pronounce on the relative merit of different formats of publication, such as the article in a refereed journal, the article in a volume of conference proceedings, or the monograph. What constitutes an effective or prestigious place of publication is a question for individual disciplines to decide, and linguists' answers will certainly be different from historians'. Personally, I found this attitude of tolerance a little too generous as I am convinced that publishing cultures in all humanities subjects are in a state of transformation. The bad news is that too much is published, and too little is read, but the good news is that in many disciplines informal hierarchies of publishing outlets are emerging which may not be as rigorously enforced as the impact-factor-based reputation hierarchies in the natural sciences, but nevertheless provide orientation to scholars as to where they should strive to publish in order to ensure a maximum audience for their findings.

Another important point the recommendations are silent on is language(s) of publication. Research in the humanities is informed by culture- and language-specific traditions of academic writing, and most scholars in the humanities consider multilingualism an asset in their practice. Arguably, however, our current practices and the academic language policies currently advocated do not promote the most intelligent kind of academic multilingualism in the humanities. Knee-jerk reactions to combat the spread of English and promote academic publication in the respective national languages will usually find favour with the public but are potentially harmful. Consider the following example. A German specialist on the Portuguese language with interesting results on the specificities of Brazilian as against European Portuguese has three theoretical options: (a) publish the findings in German and guarantee dissemination in the peer group most relevant to his or her career, (b) publish in Portuguese

and thus reach the speakers of the language itself, and (c) publish in English to reach the global community of experts on Portuguese. Each of the strategies will potentially lose some readers: people interested in the Portuguese language not reading German (a), general linguists with no particular fluency in Portuguese (b), and people interested in the Portuguese language unable to read English (c). To compound the issue further, the strategy adopted will partly determine the use made of the findings. Publication in German or English will attract additional readers with no specific interest in Brazilian Portuguese as such, but with an interest in the standardization of pluricentric languages in general (e. g. Canadian English vs. United States English, or convergence and divergence between Standard German as used in Austria, Switzerland and Germany). Publication in German may lead to more intensive popularization of the findings among the small group of German-based teachers of Portuguese as a foreign language. These are merely some of the legitimate motivations which guide writers in the choice of languages for publication.

Conceivably, publication in German or Portuguese might also be employed for less than honest purposes, for example as a convenient method to get away with the unreflected use of traditional philological methods by insulating one's work from potential criticism articulated by a now largely English-speaking international community of 'modern' general linguists. But then again, this very Anglophone global linguistic establishment could be accused of cultural imperialism, which for example indeed manifests itself often in refusing to recognize important innovations until they are made available in English. Given the complexity of the politico-linguistic terrain in the humanities, researchers need more support than they are getting now. For example it is much better to fund the translation of excellent work published in languages other than English than to force researchers who are not entirely confident in their language skills to write in English themselves.

The labours of the working group have had one immediate positive result. The group's recommendations have made it possible for the relevant professional associations in the field of English and American Studies to participate in a pilot study. The panel started work in March 2011. Its findings were published in November of the following year (Wissenschaftsrat 2012). The results of the research rating *Anglistik/Amerikanistik* will eventually help determine whether the *Wissenschaftsrat's* approach to measuring research output in consultation with the relevant communities will have a future as a routine tool in the German system of higher education.

If the pilot study turns out to be successful, English and American Studies in Germany will take the rating exercise as the external stimulus to undertake the necessary critical stock-taking that every department needs at intervals. Owing to the safeguards described above, researchers can rest assured that their output is measured against criteria developed by their peers. In the full concert of disciplines in the university, scholars in English and American studies will not have to plead that their subject represents a special case—a strategy which may bring short-term rewards but which is sure to marginalize a field in the long run.

In marketing the rating exercise to the community, both the *Wissenschaftsrat* and the professional associations will be well advised to rephrase the definition

quoted above (‘Unterstützung der Leitungen bei strategischer Steuerung durch vergleichende Informationen über Stärken und Schwächen einer Einrichtung’) as:

Unterstützung der Einrichtung bei Standortbestimmung und Weiterentwicklung durch vergleichende Informationen über Stärken und Schwächen der Leistungen der Forscherinnen und Forscher am Ort.

[Supporting the unit in its efforts to assess its position and develop its potential by providing comparative information on strengths and weaknesses of research carried out locally.]

Understood in this way, the rating exercise can become part of a dialogue between scholars and the other stakeholders in the academic system: administrations, funding authorities, other (and sometimes competing) disciplines and, not least, the educated public whose support the humanities need more than other subjects in order to survive and prosper.

If this sounds too good to be true, consider the following three alternative scenarios which might result from a successful pilot study. It is the year 2027, and we are going through the preparations for the second routine rating for English and American Studies in German higher education (after two seven-year cycles: 2014–2020, 2021–2027).

The first scenario is the dystopian one. Status hierarchies and the peculiarly strong German fixation on the professorial chair¹ will still reign supreme, and we will witness a replay of a heated debate which took place in the 2010 meetings of the working group: ‘Is my colleague allowed to report a publication by his assistant, just so he can boost his standing in the rating?’ Assuming that there are two ‘chairs’ in English linguistics in a department, the chief motivation of each chairholder to take part in the rating will still be the hope that each one will turn out the better one of the two (rather than both putting on a good show jointly, in the interest of their department and university, and—not least—for current and prospective students). Among the publications reported we will find a 500-page tome titled *Morphologische Kreativität im nigerianischen Englisch: Neologismen aus der Presse*, published in German, by a German academic vanity press, with a subsidy, and a print run of 150, only five of which are sold outside Germany. This notwithstanding, it is cited as a ‘magisterial treatment of its topic, well written and with many interesting case studies’.

This, on the other hand, is the utopian scenario. While the pilot rating (2012) stirred up a lot of furore at the time, the first routine exercise in 2020 added modifications to reduce the burden on evaluators and evaluatees, thus increasing acceptance in the community. By 2027, ratings have become socially embedded practice in the academic community, including the humanities, and apart from mild irritation caused by the inevitable bureaucratic requirements, the general response is positive – along the lines of ‘good thing somebody is taking note of the research we’re doing here’, ‘well, they’ve politely pointed out the weaknesses that, to be honest, we have been aware of ourselves—in fact, they’ve given us free expert advice’ and ‘good thing we know where we stand this time, and good thing we’ve improved since the last one’.

¹Consult the web for the collocation ‘member(s) of my chair’ and observe how much of the material emanates from the .de top-level national domain.

Neither of the extreme scenarios is likely. As an optimist, I hope for a moderately positive reception of ratings in the humanities. Colleagues will actively embrace ratings as an opportunity to showcase their achievement, but, as in the pilot study, researchers will groan at the tedium of compiling the self-report, and this will be echoed by assessors' groans at the tedium of some of the writing they will have to read.

Acknowledgments I would like to thank Dr. Elke Lütke-meier and Dr. Veronika Khlavna, of the *Wissenschaftsrat's* head office, for reading and commenting on an earlier version of this paper. I have profited greatly from their long involvement in the *Wissenschaftsrat's* rating projects.

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

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

References

- Plumpe, W. (2009). Stellungnahme zum Rating des Wissenschaftsrates aus Sicht des Historikerverbandes. In C. Prinz, & R. Hohls (Eds.), *Qualitätsmessung, Evaluation, Forschungsrating. Risiken und Chancen für die Geschichtswissenschaften?* (pp. 121–126). Historisches Forum. Berlin: Clio-online. Retrieved from http://edoc.hu-berlin.de/e_histfor/12/.
- Wissenschaftsrat. (2004). Empfehlungen zu Rankings im Wissenschaftssystem. Teil 1: Forschung. Köln: Wissenschaftsrat. Retrieved from <http://www.wissenschaftsrat.de/download/archiv/6285-04.pdf>.
- Wissenschaftsrat. (2008). Pilotstudie Forschungsrating. Empfehlungen und Dokumentation. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/download/Forschungsrating/Dokumente/FAQ/Pilotstudie_Forschungsrating_2008.pdf.
- Wissenschaftsrat. (2010). Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften. Köln: Wissenschaftsrat. Retrieved from <http://www.wissenschaftsrat.de/download/archiv/10039-10.pdf>.
- Wissenschaftsrat. (2011). Forschungsrating Elektrotechnik und Informationstechnik. Einzelergebnisse der Universitäten und ausseruniversitären Forschungseinrichtungen. Köln: Wissenschaftsrat. Retrieved from http://www.wissenschaftsrat.de/download/archiv/1328-11_Ergebnisdarstellungen.pdf.
- Wissenschaftsrat. (2012). Ergebnisse des Forschungsratings Anglistik und Amerikanistik. Köln: Wissenschaftsrat. Retrieved from <http://www.wissenschaftsrat.de/download/archiv/2756-12.pdf>.