Forum

A Discipline Refuses

Rating Academic Research Performance in Germany

John Bendix
History, University of Basel

Introduction

The ability to conduct academic research is partly a function of the time available for it, especially relative to teaching and administrative obligations.[1] For the last decade, both the number of students enrolled at German universities and the number of full-time professors has remained at about the same level. The number of *wissenschaftliche Mitarbeiter*, doctoral-level research assistants, however, has increased by more than half, and the number of *Lehrbeauftragte*, those on temporary teaching contracts, has increased by three-quarters. There is thus no lack of personnel to help professors meet teaching or administrative obligations, or to assist on research projects.[2] Nevertheless, and particularly in the humanities, German professors complain about their teaching burdens, about added administrative tasks their universities place upon them,[3] and about what they see as new pressures to bring in funding or produce results.[4] That the Historikertag, the biannual meeting of German historians, had "Boundaries" (2010) and "Resources–Conflicts" (2012) as the overarching themes for its last two meetings seems in keeping with this sentiment.[5]

Major changes to the landscape of German higher education began with unification and the resulting complex integration of East German academic endeavors into West German formats, but have been intensified more recently by other developments, including the pressure across Europe to reform university curricula and degree structures (the "Bologna" reforms). Not only are German universities to become more European,[6] but the entire drive to internationalize has become so central "that it can no longer be separated from questions concerning the reform of study programs and

German Politics and Society, Issue 103 Vol. 30, No. 4 Winter 2012
doi:10.3167/gps.2012.300404

study structures, as well as from the reform of the higher education institutions and the entire higher education and science system."[7] Accompanying this have been efforts to rank universities, whether across Europe (Leiden University's Center for Science and Technology Studies (CWTS rankings) or more globally (Shanghai rankings since 2003; Times Higher Education since 2004).[8]

Within German universities, the advent of management consultancy promoting performance-driven norms of efficiency has led to a vocabulary, as well as university policies, that emphasize outcomes and accountability.[9] Higher education is seen increasingly as a competitive marketplace, and this is reflected in well-publicized, indicator-based, and commercial efforts to help students choose where to study. The best-known of these is the Centrum für Hochschulentwicklung (Center for Higher Education Development, CHE) ranking, which is addressed below.[10] Another part of the new competition has been engendered by the Excellence Initiative, a German government effort since 2005 to encourage high quality, cutting-edge research and create graduate programs for young scholars.[11] The use of *leistungsorientierte Mittelvergaben* within universities, which internally (re-)allocates some resources based on performance, has also increased. Evaluation is at the heart of all these developments, and can be said to have started after unification with the effort to establish whether what had been done in East German universities met West German standards.

Evaluation efforts have proliferated in the meantime,[12] so much so that they have been labeled *Evaluitis*, an "epidemic disease that has particularly befallen research science."[13] Not a few researchers feel frustrated by the sheer number of assessments they are now subjected to,[14] and what they see as recurrent demands to prove adequacy despite a conviction this has been more than sufficiently demonstrated already. As Jena University sociologist Hartmut Rosa puts it:

> When I finished my *Habilitation*, I thought: well, that was the very last exam of my life. I soon realized how wrong I was. Actually, with every evaluation, every grant proposal I write, I get scrutinized all over again. And the yardstick used, increasingly, is no longer my entire career as a researcher but instead only what I've done in the last two or three years. A university teacher must constantly prove that he is entitled to have his post.[15]

Accountability demands, in other words, raise doubts among researchers about the wisdom of the scrutiny itself. The frustration extends to the Excellence Initiative. Faculty members need to draft and submit these proposals, which to Humboldt University historian Jörg Baberowski simply cuts into research time:

> We just need time, to read and do research, and sit at our desks, and not be on a treadmill where we write proposals that are completely pointless about things that in the end professors themselves don't even research because they appoint some doctoral student in a research cluster to do it.[16]

Though expressed here by individuals, skepticism about evaluation is a *leitmotif* that runs through the opposition of the Association of German Historians (VHD) to having their discipline be evaluated (or "rated.")

Still, the discursive landscape[17] is rather sharply divided between those who question the sense in trying to externally evaluate academic research, and those who are interested in refining evaluation methods.[18] It is a conflict, in one characterization, between the Humboldtian ideal and New Public Management.[19] Those interested in refining the methods of evaluation have a tendency to let:

> the reasons for evaluation as well as the difference between assessment and measurement sometimes fade a bit into the background, while the critics barely address the "how" of evaluation, often argue scientific research is a special case, and have a tendency to decouple (and try to immunize) the research system from its environment.[20]

The critics might be interested to discover that those engaged in refining evaluation methods do not do so blindly but instead are concerned to establish whether the use of an evaluation is meant to be instrumental, conceptual, interactive, legitimating, or tactical. Linked to it is the question whether its function is formative (aimed at improvement or modification) or summative (assessing past performance). Distinctions are also made between engaging in audits, controlling, quality management, performance measurement, monitoring, or evaluation. In light of the reactions, the questions that the refiners of methods raise about the effects of evaluations, whether in terms of intent, anticipation, explication, significance, or assessment, might even be of some relevance to the critics.[21]

Yet among those assessed, the tendency is to disregard such analytic nuances and functionally, or without adequate differentiation, define evaluation simply as an "after-the-fact assessment of the performance of an organization or person by external experts."[22] That is accompanied by a not small degree of anxiety that the ultimate purpose of evaluation is to discipline scholars and researchers.[23] Lending weight to this fear are recent sociological studies that suggest rankings either gradually transform the evaluated "into entities that conform more closely to the criteria used to construct rankings" or that evaluations serve in the "construction of self-reproductive status hierarchies."[24] In both cases, the examining tool used alters the subject examined,[25] which itself is a reason that self-reflective subjects being examined resist such tools.

The Wissenschaftsrat (WR), the German Council of Science and Humanities, has recently summarized the changes as well as cross-pressures that bear on higher education and the researchers in it.[26] The focus of the discussion below is on pilot efforts the WR has undertaken in recent years to "rate" (make assessments, comment on or give grades to but not directly rank) research conducted in Germany in specific disciplines. The overarching idea has been to identify the "strengths and weaknesses as well as the profile of the discipline in the respective institution."[27] In terms of evaluation process, this effort is noteworthy for involving the disciplines being evaluated from the beginning in the assessment process itself.

It is also remarkable for failing to convince History, the first humanities discipline it approached, to participate. This can be seen as a first, and quite public, shot across the bows marking the beginning of open resistance—as opposed to private grumbling—to efforts trying to evaluate research in the humanities. It is difficult to know how isolated an act this is, however, because English and American Studies agreed to participate in the WR's pilot program instead.

What follows is an examination of the rating itself as well as the reasons given for the refusal to participate. In part, I suggest History[28] resists outside evaluation[29] because it more highly values its own disciplinary-specific practices and standards in evaluating its own products, and places its autonomy to do so above other claims. In that sense it can be said to have struck a blow for what it considers to be the Humboldtian ideal, with which other, smaller, disciplines might heartily agree.

Nevertheless, I argue that those practices resist easy articulation. They are embedded in the socialization involved in becoming a historian, and the disciplinary self-image upholds an ideal of craftsmanship, a theme explored at the end of my considerations. Evaluation perceived as coming from "outside," and using a managerial language of output, resources, and productivity—or framed in perceived quantitative terms—deeply offends both disciplinary self-image and socialization. Yet, because these embedded practices are so implicit in what it means to be a practicing historian, the reasons given for opposing evaluation are fragmented, disparate, and inchoate. In the end, it remains a challenge for historians to provide a positive articulation of what the evaluation of research in their discipline means to them, couched in terms that could be understood by non-historians. Even the suggestion that it might be in their interest to provide such an articulation may well be seen as intrusive.

The following is organized into three parts. The first gives an overview of the WR's intent and actions. Prior to approaching History and Electrical

Engineering, as examples of a humanities and a technical discipline, the WR pilot completed evaluations of a natural science (Chemistry) and a social science (Sociology). Some sociologists participated later in the main historians' discussion platform, and the results of the WR sociology evaluation were publicly available, so historians had an idea of what they could expect should they participate. For this reason, some discussion of the sociology pilot results is provided.

The second part summarizes themes that emerged in the positions taken both for and against participating. Positions in support were in the minority. In the opponents' camp, some objections were made to the proposed (or assumed) methods, others to (presumed) interests being served. More often, objections reflected the self-image of the discipline, with many historians remarking on the need for evaluation to be rooted in the peculiarities, specificities and practices of the study of history. From that perspective, evaluation should not be conducted in a disinterested, general, global manner by those who did not "represent the interests of any specific discipline, institution or organization"–which is how the WR characterizes itself.[30]

The third part suggests reasons for the historians' rejection. One concerns the frequent pairing nowadays of calls for accountability with assertions of autonomy. I argue here that there is a fundamental difference in how researchers, especially in the humanities, understand these two notions from how outsiders–including academic administrators or politicians who hold the purse strings–understand them. That difference rests on how historians think about the very nature of their products and how they should be evaluated. The WR, in the end, foundered on the shoals of a discourse inside the discipline, a discourse incompatible with the discussions of evaluation conducted outside the discipline.

The wider context involves institutional considerations of autonomy and accountability. On the one hand, scholars at German universities have long claimed a right to be free from outside interference. Yet this has gone hand-in-hand:

> with the persistence of an institutional relationship between the university and the state that was both utterly non-autonomous and characterized by more or less total dependence on both the regulatory and providential tutelage of the state over the university.[31]

In the current era, the demands for university autonomy from the state grow louder, so the exercise of control shifts from the tutelage of the state over universities to increased tutelary functions a given university exercises over its internal affairs. That gives university leadership a freer hand:

> to encroach somewhat upon the professor's unlimited right to his or her own academic agenda. At the same time, the unfettered exercise of the individual faculty member's autonomy is seen as undermining the very degrees of freedom that the university has gained in the battle for greater autonomy from the state.[32]

The implication is that at the limit, from the university administration's vantage point, an individual scholar's interpretation of what autonomy means runs counter to the university's interests.[33]

The Wissenschaftsrat's Good Intentions

The Wissenschaftsrat makes recommendations about higher education and research to the German federal and Länder governments, and serves as a national policy advisory council. It has existed for more than half a century, and is an example of successful cooperative federalism, inasmuch as it brings representatives of the federal government, the state governments, academics, and research institutions together, without any one group able to outvote the other. In fact, consensus is prized, and decisions the WR reaches internally are often by large majorities.

More relevant is that its suggestions have often become enacted government policy in recent years. On the WR's recommendation, the research conducted in former East German universities was evaluated, and since that time, the WR has increased its interest in evaluation, leading to recommendations to introduce B.A. and M.A. degrees (2000), provide better support for up-and-coming scholars (2001), and reform how doctoral students are trained (2002), among others. Wissenschaftsrat recommendations also lay behind introducing the Excellence Initiative (2005), which led at least one commentator to argue the WR "increasingly plays the role of initiator and moderator of institutional change in science."[34] That very role as a major motor of change may well be a reason historians did not trust it.

After a lengthy preparatory phase and careful consideration of the methods to be used, the WR announced in 2004 that it intended to carry out a ratings exercise. Organizations participating in the WR, as well as the professional associations in chemistry and sociology, proposed reviewers, who once selected were appointed to assessment boards for these two disciplines by a steering group of the WR's science commission.[35] There was an explicit interest in having the assessment boards represent disciplinary subfields in as broad a manner as possible, and an interest in finding reviewers with international experience. Experts from the Netherlands,

Austria, and Switzerland also participated on the assessment boards, which had fifteen and sixteen reviewers, respectively, on them.[36] In the case of Sociology, the process began in November 2005 with the appointment of the assessment board, and after developing indicators, collecting and correcting data, and completing the analysis, the results were published in April 2008, thirty months later.

In the course of its preliminary discussions, the steering group suggested "the quality of research should be the core criterion of research rating and be assessed in a more differentiated way than the other criteria."[37] Initially, there were to be nine criteria, but in the discussions this was whittled down to three dimensions–research, promotion of young researchers, and knowledge transfer–associated with six more specific criteria. For the research dimension, these criteria were research quality, impact, and effectiveness/efficiency; the criterion for the promotion of young researchers was identical with the dimension itself; and for the knowledge transfer dimension, the two criteria were transfer to other areas of society and promotion of the public understanding of science. Dimensions and criteria were meant to be broad and applicable to rating all disciplines, while the information to be provided to the assessment panel(s) would be discipline-specific. In terms of the humanities, a WR committee formed subsequent to the historians' refusal argued that the "selection of criteria and the classification or assignment of indicators should be discipline-specific, if possible only using those indicators that do not create undesirable incentives and that cannot be manipulated."[38]

The entity to be assessed was the Forschungseinheit, the research unit at the institution participating in the pilot. That institution also identified the research units to be assessed. The Wissenschaftsrat did not actually define what such a research unit was, though it was meant "to allow research quality assessments at a lower level than that of entire institutions" and be from a level below that of a faculty, which in German universities is typically called a seminar, institute, or department. Functionally, when "research units" were subsequently identified in the Chemistry pilot, they comprised "six senior scientists … including three professors" on average. In Sociology, however, "nearly 75 percent of all units comprised only one professorial chair," which meant a single professor together with his or her graduate students and research assistants.[39]

Because one tends to think of a "research unit" as something larger in organizational terms, such as a team working in a laboratory, or in thematic terms, such as having a focus on a particular research area, it is worth examining the reasons the WR gave for identifying research units in

this manner, not least because it sheds light on structural aspects that make evaluation and comparison difficult. A number of these aspects, here described for Sociology, would be true of History as well. First, most Sociology seminars, institutes, or departments at German universities are structured by instructional task, not by research foci. Second, about one-third of the institutions identified research in Sociology taking place in two or more faculties, reducing the chances that those conducting sociological research would cooperate to form a larger research unit. Third, much of the research engaged in was multidisciplinary, so diffuse boundaries made it difficult to identify research that could be unambiguously assigned to the discipline of sociology. Finally, as it was known that the results of the pilot study would be publicized, "tactical considerations"–the term is used by the WR itself[40]–might have played a part in determining which entities universities chose to identify as research units. In other words, both organizational aspects (where research was being conducted in the university) and the nature of the research being conducted (multidisciplinary, or not wholly in Sociology itself) militated against defining a "research unit" as being much more than a professorial chair, at least for sociology.

An evaluation based on research units identified this way can neither determine how good an institute's research is collectively nor can it even meet the WR's own ambition of assessing the "strengths and weaknesses as well as the profile of the discipline in the respective institution."[41] All it can do is produce a set of separate judgments about the strengths and weaknesses of each individual research unit–which for Sociology meant three-quarters of the judgments were being made about individual professors. The mode chosen was to state that of seven or eight Sociology research units at a given university, for example, the weakest was judged "satisfactory" and the strongest "good." In the end the Sociology pilot evaluated research performance in 254 research units at fifty-four universities and three institutes. The range of research units was from one to ten (average 4.5) per university, each research unit had an average of 5.7 researchers (1.5 of them professors), and overall, 376 professors participated in the rating (21 percent of them women).[42]

The criteria related to the research dimension were graded on a five-point scale (unsatisfactory, satisfactory, good, very good, excellent), and because many universities listed more than one research unit, research quality–the first, and arguably the most important criterion–was evaluated as a range (e.g., satisfactory to good) in order to encompass all the research units. As information on knowledge transfer and promoting public understanding was very heterogeneous and virtually without quantita-

tive indicators, these were reported on a simpler ordinal scale (below average, average, above average). One can readily make this rating resemble a ranking. If one compares research quality in sociology at Dortmund with that at the Free University (FU) of Berlin, Dortmund's research units range from unsatisfactory to good, while the FU Berlin's range from satisfactory to excellent: The entire scale is higher at the FU Berlin, so its quality of research "ranks" higher than Dortmund.[43]

Assessment was to be based primarily on qualitative, peer assessment of publications, augmented by quantitative indicators. In other words, this was intended to be "informed peer review."[44] The WR-specified reviewers were to look at research results with respect to "significance, degree of innovation, originality, timeliness, and both national and international recognition with respect to the breadth and influence of the questions posed, both in one's own area of research and in other disciplines."[45] Overall evaluation results reflected "the judgment of a group of evaluators, based on publications, various qualitative and quantitative indicators, and background information on each facility," but "in the end, it was the analysis of the content of the research–by reading the publications submitted–that played the key role."[46]

The publications in Sociology for the period assessed (2001-2005) were first culled from existing social science databases. The results were then sent to the research units for correction and amendment, a process yielding a list 50 percent longer, a reflection of the absence of single, comprehensive, bibliographic source for the discipline.[47] Each research unit had an average of sixty-three publications, which works out to 2.7 publications per researcher per year, a figure comparable to other studies of academic productivity. Research units themselves selected the publications to be evaluated, with acceptable publication types including journal articles (including e-journals), contributions to edited volumes (including to thematic issues, compilations and *Festschriften*), literature reviews, and book reviews. Monographs, however, could only be submitted in the form of extracts no more than fifty pages in length and ephemeral literature was explicitly excluded. That research units selected which publications were to be evaluated, even if they were only allowed to submit three publications on average, is a feature lacking in the UK's research evaluation exercises, for example. A major point of contention to historians, however, was the fifty-page limit for monographs, and their stance on this issue is an important key to their self-image.[48]

Each research unit in the Sociology pilot was assigned two rapporteurs, selected based on their particular areas of expertise. The rapporteurs were

given information that they were to use in their assessments. Such data was both quantitative–the number of articles in peer-reviewed journals, number of reviewed third-party financed projects, total number of publications by type(provided in both absolute and relative terms)–and qualitative–the submitted publications themselves, a list of publications, a list of third-party funded projects, and a self-description of strengths and weaknesses. As in the UK Research Assessment Exercise, the criteria for judging research quality were originality and significance, though the suitability of methods used substituted for the criterion "rigor" used in the UK.

Rapporteurs evaluated individually, though they worked together if they disagreed, and reasons for their judgments were presented to the assessment board. The WR has subsequently come to the conclusion that "a standardized evaluation of the quality of publications in the area of the humanities is hard to imagine for the time being," and later interviews conducted with members of the Sociology assessment board indicated evaluators were "only given vague guidance about how they were to deal with the results of examining the literature and the quantitative data in arriving at their assessment."[49]

The major work of evaluating research units, including judging publications, therefore rested on the shoulders of two evaluators, and though the assessment board as a whole voted on the final rating of a research unit, careful studies of comparable academic decision-making situations suggest internal decision-making dynamics directly affect the scores given.[50] Board members may engage in deferential behavior to preserve collegiality and not question the judgment of those deemed more competent and knowledgeable about a particular subject. Evaluators may drift towards homophily and find that work excellent which "most looks like their own work."[51] Decisions may be made in rapid succession simply to be able to finish scoring in the limited time available for the meeting. Such dynamics make it difficult to fully adhere to declared criteria or standards, but if the gold standard is peer judgment, the modes used to arrive at judgments may seem less important. It is indisputable, however, that such decision-making dynamics were at work in the WR Sociology assessment board meetings.[52]

Overall, the grades given to the 254 research units in sociology followed a Gaussian distribution. The mean was just below "good," with 9 percent of the research units judged unsatisfactory, 24 percent satisfactory, 38 percent good, 18 percent very good, and 4 percent excellent (7 percent could not be evaluated). If one aggregates to create only the categories below mean, mean, and above mean, then 33 percent of the research units

were below the mean, 38 percent at the mean, and 22 percent above the mean (again, 7 percent not evaluable). This prompted the higher education editor of the *Frankfurter Allgemeine Zeitung* to conclude that "on average, mediocrity reigns" in German sociology.[53]

Responses to the Wissenschaftsrat Proposal to Rate History

Having examined a natural science and a social science, the WR next wanted to rate a technical and a humanities discipline. The pilot study of electrical engineering, the technical discipline, launched in 2009 would be completed rapidly, the final report appearing in June of 2011.[54] But history balked, with the Association of German Historians (VHD) taking an executive decision at the end of June 2008 to not participate in the pilot study.[55] That decision was widely supported in the discipline; English and American Studies thereupon agreed to participate in its stead.

In 2009, the WR established a working group, likely in reaction to the historian's decision, to formulate specific recommendations for evaluating research in the humanities.[56] Unfortunately, discussion within the VHD and in various history institutes have not been made public, so the following draws largely on submissions to the website *H-Soz-und-Kult*, the key internet information clearinghouse for historians.[57] *H-Soz-und-Kult* understands itself as a community platform run "by researchers and for researchers," and its moderators are at the History Institute of the Humboldt University in Berlin. In mid2009, more than a year after the VHD's decision, the site launched a discussion forum on the topic "Measuring Quality, Evaluation, Rating Research. Risks and Opportunities for the Discipline of History?"[58] Some contributions to this forum were solicited, and not all came from historians.[59]

Voices in Approval (and the CHE-ranking)

A number of proponents of participating came from the ranks of those who had been directly involved in the WR's rating process. Advantages, in the view of two sociologists who served on the Sociology assessment board, included evaluating individual and decentralized research units rather than an institution as a whole. Ratings also emphasized output per researcher, seen in the context of other burdens, such as teaching, that reduced time for research, rather than overall productivity. It was also considered good that the WR focused on evaluating publications rather than, as in German and Swiss universities, using the amount of third-party

funding acquired or spent as an indicator: better to rely on the qualitative judgment of peers than on automatically generated, quantified data.[60]

One organizer of the ratings project at the WR itself noted that because "research can only be properly assessed by researchers," a key benefit was to have the evaluation process explicitly be shaped and steered by researchers themselves, including the professional association of that discipline. Another advantage was that indicators were not automatically converted into grades in the scoring process, but instead resulted from having "numerous quantitative and qualitative aspects be considered together." History had been selected for its "overall size, internal complexity, interdisciplinary connections as well as its many non-university research institutions."[61]

Those historians who saw something of value in the ratings adopted a resigned tone. Rankings and ratings were here to stay, history professor Lutz Raphael argued, so the only viable option was to participate, which he himself had done as a member of the WR's scientific committee since 2007. By taking part, one might be able to exert some influence, "expand the room to maneuver, the room to autonomously shape such evaluation processes."[62] Besides, if one did not support this particular rating exercise, far worse or more amateurish modes of evaluating already in circulation would be imposed.

This last was an allusion to the CHE rankings of academic programs. The Centrum für Hochschulentwicklung was founded in 1994 as a limited liability company by the German Rector's Conference and the Bertelsmann Foundation, a multinational media corporation that provides much of the funding.[63] The first CHE rankings, of economics and chemistry programs, appeared in 1998 and were produced in conjunction with the *Stiftung Warentest*, an independent organization (and foundation) that tests and compares goods and services to help consumers make informed choices. CHE rankings are meant to help students choose their university wisely. The overt intent is to provide information about the strengths and weaknesses of departments, but commercial motivations to sell more publications containing these rankings lie behind it, a motivation common to comparable commercially produced ranking lists.[64] CHE rankings today survey thirty-four disciplinary areas and courses of study, and draw on information about teaching, resources, and research at dozens of German universities. The CHE also creates a "research ranking" for specific disciplines at German universities, though it only uses three indicators to do so for humanities disciplines: third-party funding, the number of PhDs granted, and the total number of publications.[65]

The "rankings and ratings are here to stay" argument was echoed in *H-Soz-und-Kult* by Dieter Müller-Böling, a cofounder of the CHE.[66] Expansion in the number of German universities and increased specialization led to decision-making constraints that left few alternatives to participating in ranking. In a changing higher education system, transparency had to be increased to improve the steering and competitiveness of the institutions themselves. Indeed, the German Rector's Conference had concluded the entire system needed to be reformed, and that ranking should form a part of that reform. To keep higher education institutions from being politically coerced by state-run education bureaucracies,[67] and to bring dynamism into the system, it had been urgently necessary to establish what the research output actually was. Quantification was not an end in itself but rather an "informational process to permit comparative judgments to be made." Müller-Böling saw the development of distinctive profiles and competition among universities as a positive effect of CHE rankings, as they led to strategic decisions at the faculty or university level.

There is a direct connection to historians here, because the CHE began publishing rankings for humanities disciplines in 2002. At the time, they received "lively cooperation in the design of rankings and indicators from the discipline of history."[68] Yet a mere seven years later in July of 2009, a year after refusing to participate in the WR rating, the VHD decided to no longer participate in CHE rankings. In its justification, the VHD noted that its experience had led it to conclude that the CHE rankings were "alien to the discipline," and that what was gathered and reported did not allow for "acceptable information" to be generated about the capabilities and capacities of German history institutes. The refusal was signed by twenty-six history institutes across Germany. The VHD remained open to "discipline-specific" procedures as an alternative to the "currently dominant general evaluations,"[69] though it provided no information about what this meant.

Voices in Opposition

Objections to the WR rating came in two main forms. The first looked at what was coming from without, and raised questions both about the specific methods used and about the more general purposes evaluation might be playing. The second, more inward-looking, commented on the nature of the discipline as historians themselves understood it.

As to the specific methods, objections were made to how data that would form the basis of a rating would be collected, aggregated and evaluated (no mention was made of the process used for Sociology). There was also disbelief, if not incomprehension, expressed that reading a brief, fifty-

page excerpt from a book could allow anyone to evaluate how original a research effort it was.[70] The question who might be selected as evaluators raised concern as well, with a clearly expressed preference for not having emeriti but instead those still actively engaged in research on assessment panels.[71] It also was not clear whether locational factors could adequately be taken into account, and whether a one-shot rating could in any manner properly reflect research that spanned many years.[72] There was concern over added expenses such a rating would incur.[73]

In a press release, the VHD summarized a number of further objections: "the unclear criteria and the unforeseeable consequences of a rating conceived as a pilot project with an open outcome means it cannot be determined exactly how representative it is or what its scope will be."[74] Historians at the Institut für Zeitgeschichte (IfZ) in Munich put it more sharply:

> the [WR] ratings procedure is supposed to establish comparability and be transparent as well as increase the "international visibility" of the discipline. In fact, however, it flattens specific strengths and weaknesses, distorts the profiles of the research institutions and generates misleading perceptions. The unidimensionality of the grades misses the diversity and differentiation in the research landscape.[75]

A major complaint was not even specific to the WR: we are already suffering from this new disease *Evaluitis*, and are ill from being over-evaluated. A more subtle variant argued that "our history institute" was already being regularly evaluated, including for aspects the WR did not consider, such as the degree of internationalization, and changes proposed by such regular evaluations were already being implemented. The WR's effort was therefore superfluous, as it gathered less information and yielded results inferior to existing evaluations, plus it came at a cost and effort hard to justify.[76]

Underneath these objections lay a deep suspicion of what end evaluations in general, and the WR ratings specifically, served. As the VHD's president put it, because WR ratings came in "parametric" forms, they appeared suited for making political decisions.[77] The categorizations, if not numbers, they furnished made politicians feel they were informed and could make decisions based on them. Yet, if politicians were trying to foster "excellence," that was something far more than simply applying principles of economic competition to academe. Indeed, it could hurt the discipline to participate in such a rating if it then formed part of a competition over resources, and if the WR thought research could be steered through ratings, they were deluded. All that would be reinforced was strategic behavior destructive of academic culture, especially in the humanities.[78]

The nature of the discipline was felt to pose particular challenges for evaluation. History was a heterogeneous, changing and dynamic disci-

pline, filled with niches and sub-fields, and one long suffused with tensions between critique and reconstruction.[79] That made it hard to understand in toto, particularly if an evaluation followed a standardized procedure. At the very least, it would call for a large number of evaluating peers to encompass the heterogeneity. The discipline was conceptually fragmented, and the specificities of certain sub-disciplines such that an appropriate representation of the numerous theoretical approaches for the purposes of evaluation would be difficult to achieve.[80] Indeed, if one agrees that the conceptual structure of the humanities currently emphasizes specificity, inter-subjectivity, perspective, verbal expression, reflexivity and universality,[81] it is hard to see how the humanities can be evaluated at all. History also has a division of labor, with some engaged in discovery and the exploration of new areas, others fleshing out well-established areas, as when an international research result did not yet have a German counterpart. Using the criterion of "originality" would clearly favor scholarship of the first kind of history research to the detriment of the second.[82]

Evaluation, if carried out at a particular point in time, could not adequately or properly understand research output in the discipline. It might take four or six years to produce one's first qualifying work, in the form of a dissertation or *Habilitation*, several more to produce major studies, and perhaps another seven or even ten years for "innovative outputs to be communicated and to spread through the discipline" through reviews, discussion, and conferences. A snapshot taken in a particular year would yield a distorted picture: a rating or evaluation would have to be repeated at regular intervals and under identical conditions to accommodate this time dimension.

Still, the image of the lone historian producing insights in splendid isolation was not wholly accurate, even if it reflected a fond self-image of heterogeneity or was underscored organizationally by individual university chairs in history. Cooperative research efforts did exist, as at the IfZ in Munich or in larger collaborations to create products like the *Monumenta Germaniae Historica*. The argument that third-party funding amounts which supported such collaborative efforts was a poor measure, or that research was not organized in teams–an argument why "research units" were the wrong entity to evaluate–was therefore not entirely correct.

It was true, however, that the discipline, and humanities scholarship more generally, was internationalizing. Being "internationally visible" had become more important, though problematic because the discipline continued to be dominated by the "primacy of the perspective of national history," as Lutz Raphael characterized it. That affected the development of

sub-fields, the particular foci of study chosen, and schools of thought. The result was often that the national language was used to present research results to a largely national audience. In response, Raphael suggested one not give foreign (e.g., English) language publications pride of place in the discipline, and that one separate the use of international standards for assessing academic work from the question of international visibility or international networking.

Still, historians differed over the role of the international. Ulrich Herbert argued that an international consensus did exist over what scholarly standards were. His example was the evaluation of an academic text, where the, in essence universal, standard for judgment was "the breadth of knowledge of the materials, the familiarity with the literature, the analytic acuity, the resourcefulness and originality in the research undertaking, the plausibility of the judgment, and finally the aesthetics of the language used in composing the text." Though consensus might be difficult to reach over how to specifically measure research performance, surely one could not pretend that "the achievement of a historian or a literary scholar … be treated as literally immeasurable."[83]

Observations on Autonomy and Disciplinary Self-Image

Increasing evaluation activity is one aspect of numerous other changes to the German higher education landscape. Growing calls for accountability from universities accompany these changes, whether this is taken to be accountability to students, funders, industry, parliaments or the general public. The pressure to demonstrate accountability has increased as funding has become more competitively awarded. In the past, funding for research often came as part of the basic grant to a department, but now it must be applied for and publication success is an important indicator in grant decisions.

The historians' response can hence be understood as asserting autonomy in an era that ever more loudly calls for accountability. Yet, in this context, different players mean different things when using these terms, with autonomy in the historians' view an assertion that "our" (historians') way of evaluating is superior to "their" (the WR's, the CHE's) way. To better understand this, it is helpful to understand the self-image and the sensibilities historians' have, including for evaluating the work of fellow historians.

Raphael's argument that those who do not participate willingly in an evaluation will have participation in that evaluation thrust upon them has

a sophisticated counterpart in the pairing of autonomy and accountability. As a researcher sees it, the autonomy to conduct research as he or she judges fit (and on a freely chosen subject, free from interference) carries with it an accounting of what was done (or is planned), including justifying the incurred expenses, to the employing university, grant-giver or government agency. Accountability can be understood more abstractly as well, as when it is seen as owed to the profession, to peers, or even to society at large, as in "working through" and coming to terms with the past–a subject of particular weightiness for postwar German historians.[84]

But a politician in a German state who must vote on allocating money to a university in that state sees the trade-off between autonomy and accountability quite differently. If the university is to be given more autonomy, a demand which has been growing louder in Germany,[85] then the state handing over some of its control wants an accounting from the university administration: this is taxpayer money, after all. By placing more control over the budget into the university's hands and taking it out of the hands of the Land's education ministry, accountability becomes defined relative to political choice. How much money, the politician wants to know, how much control over the state's resources, should we grant universities compared to resources we allocate to other institutions? What sort of accounting will accompany the increased autonomy we grant?

This sense of accountability is institutional, but to a researcher at the university it seems all bookkeeping and bureaucracy. A humanities scholar whose vision of accountability is to provide a project description, including for whom the results will be valuable, of his or her research in return for being left in peace to get on with the research (i.e., autonomously) has trouble in a context that sees accountability as that of the university as a whole to the state or the public at large. That scholar would like, in Thomas Widmer's terms, to decouple or immunize the research system from its environment. To the researcher, one might say, autonomy looms large; to the politician, the funder, the public, accountability looms far larger–and what autonomy and accountability refer to in each case is different or almost inimical.

This helps situate some of the historians' opposition. Accountability demands being made on researchers are resisted partly because of a sense, as Jörg Baberowski puts it, that those demands are interfering with the ability to get valuable and important research done. Not only that, but outsiders want information in forms that is hard or time-consuming to supply, or who want to judge what is being accomplished based on criteria alien to the discipline. Institutionally, the VHD is accountable to the

researchers and institutes it represents, but to whom the WR is accountable seems unclear to at least some historians.

Still, accountability games only work when there are resources to distribute, and if there is little or nothing, the game can soon turn cynical. We researchers pretend to give you–administrators, politicians, the public–what you want, namely an accounting or signs of accommodating behavior, such as publications in the "right" journals. You, in turn, pretend to give us something in return, namely resources either inadequate or symbolic. In a seemingly endless era of belt-tightening, historians may simply be saying the returns they foresee from participating in a pilot rating exercise, one which carries no consequences for university finances,[86] may just not be worth the time or effort.

In its current self-image, the discipline does not serve the interests of the powerful, and a significant intellectual strand in history has argued for a more than a generation that the discipline should serve, and study, the exact opposite: those who have not been powerful. The interest in the underdog view of the world may be an undercurrent of the opposition too. Certainly it is of a piece with the view that the discipline uses methods of "self-referential observation and evaluation"[87] that by definition will be incongruent with methods of observation and evaluation coming from outside the discipline.

One can start with the simple fact that without much reflection, the WR assumed History was a humanities discipline. Yet there have been repeated efforts by historians themselves to argue that History is, or also is, or has aspects of a social science.[88] The question whether History ought to be placed in the humanities or in the social sciences is not new,[89] nor is it a question confined to Germany. The National Foundation on the Arts and the Humanities Act, a law passed in the U.S. in 1965, for example, placed History in the humanities, while the National Research Council, in a study of doctoral programs a generation later, instead placed it among the social sciences.[90] In its listing of disciplines, the 1965 Act put History between Literature and Jurisprudence, while the 1995 study listed it between Geography and Political Science. If one considers the breadth of what has been published under the label of "History," neither placement seems wholly inappropriate. But the issue is not where History should be located. Rather, the issue is whether the characterization of what it is and does comes from inside or from outside of the discipline.

One of the objections historians raised concerned the selection of evaluators, and yet the WR went to considerable effort to identify evaluators acceptable to the disciplines of Chemistry and Sociology, and certainly

did not question the premise that evaluators should be very familiar with the discipline they were to evaluate. In point of fact, in the Sociology pilot, some of the best-known, active, and respected sociologists in the country did serve on the assessment board, which raises an interesting point about peer judgment. Judgments reached by those not in history may be rightly feared by historians for being made in ignorance. But what of judgments made by those who are anything but ignorant? Does their objection hide a greater fear that the judgments reached by experienced, respected scholars in the discipline might be appropriate—and, much as in Sociology, result in a finding that mediocrity rather than excellence rules?

Rejecting participating in the WR pilot may rest on a more subtle basis, and not so much be a rejection of evaluators themselves as it is of the Wissenschaftsrat and what it is assumed to stand for. As a nationwide council not beholden to any particular discipline or institution, the WR talks to national and state governments and recommends adopting national policies. The VHD talks to individual historians and university history institutes scattered across the country and furthers the interests of a particular discipline. The WR's universality hence stands in contrast to the historians' particularism—and universal and particularist claims tend to be incompatible.

One can take this a step further. In more highly valuing their autonomy to determine their own standards, and related to the objection of reading only a fifty-page excerpt from a monograph, one can suggest that historians, or humanities scholars more generally, have the sensibility of artists. They want to be encountered in their entirety, and demand readers do the work of reception and not merely of consumption. They want readers, or reviewers, to plow through the evidence of their contemplation, not just flip through a few pages in search of pre-determined idea of "quality." They are creators who lose themselves in their work and want to be appreciated as craftsmen honored for their work ethos and their serious engagement with their creations. Like craftsmen, they are immersed in the "dignity of their objects," as Ulrich Herbert put it, and in the sense Max Weber described, are not merely doing a job but exercising a vocation, or even living out a calling.[91]

Publications, while they contribute to a scholarly discourse, then are understood by their creators in the Romantic, holistic manner that "a work" was understood, as a piece of art or craft. Such "work" expresses individuality, and is a struggle to express. It may only express the artistic vision of the self, but it hopefully also says something to others, even much later in time. Producing such a work of art or craft also can have a

kind of obsessional, driven energy behind it that can lead to "rewriting a sentence again and again to get its imagery or rhythm just right."[92]

The consequence for evaluation is that only a "true" colleague, meaning a colleague similarly dedicated and committed to the craft, one who understands and values the enterprise in the same kind of way, is capable of judging what is produced. As this is very much an individual struggle to express oneself, at the limit no peer will ever be fully able to comprehend the product, and must rely instead on a general understanding of what is at stake or what has been accomplished. The only genuine, acceptable, judge of a craftsman's work, can thus only be another craftsman, one who can draw on their own experience in exercising the craft.

Conclusion

The expressed purpose of the WR rating was to establish the profile of a discipline in a given institution, and more generally to characterize performance as strong or weak. In that sense, it is of a part with trying to promote excellence, though doing so, if the term is to retain any meaning, will perforce highlight what is less than good. By definition, there can only be a limited amount of excellence.

One reading of the Sociology pilot results is that for every above average research unit, there are three to four merely average or below average. In a competitive era, being judged average is not what one wants to hear. A little self-defensively, some historians opined that if the real intent was to identify the high-flyers, then "it was anyway known in the guild who was brilliant–and a rating was therefore a waste of effort."[93]

One can take this a little more abstractly as well to argue that at its core, history, and humanities more generally, constantly question structures and assumptions. What is produced by the discipline is not a report of a completed lab experiment but instead reflects approaches and insights gained as part of a dialogic process. The purpose of such insight is to find better terms, more accurate contextualization, new vantage points, multiple interpretive possibilities. Evaluation then needs to be more than an evaluation of the particular quality at a particular point in time of that particular insight, but instead be of the contribution made to the debate–and debates in historiography can stretch over a rather *long durée*. History is suffused with the insight that every generation must come to its own understanding of the past, making the object of study shifting and prismatic. It undermines being able to assert clearly what knowledge is (still)

canonical in the discipline, and creates genuine difficulties in reaching a comprehensive understanding of where the discipline actually is.

What unifies history, in a recent characterization, is not the "notion that the field is (or can be) unified around a common theory." Instead, "what is shared is agreement on what constitutes good historical craftsmanship, a sense of 'careful archival work'" that is based on "certain shared values about commitment to doing certain kinds of work."[94] If true, then it is only those who share the guild's values who can properly evaluate their fellow craftsmen, and that means not just reading a fifty-page excerpt but considering the whole, both process and the outcome.

The critique of using brief excerpts ultimately reflects a fundamental difference in perspective, and an irreconcilable difference between two concepts of quality. On the one side stand those who search for excellence, for better methods and tools. They want to create "a system that works correctly and their impulse to reform reflects something about all craftsmanship: to reject muddling through, to reject the job just good enough, as an excuse for mediocrity."[95] This is the impulse that spurs ratings, evaluations whose aim is to uncover weaknesses, and ever more subtle or complex ways to assess.

On the other side, one has "claims of practice" made by craftsmen themselves, which "encompasses pursuing a problem in all its ramifications," a pursuit that takes patience and time. From this corner comes the claim that a snapshot taken of research is simply inadequate, which is one of the historians' critiques. Richard Sennett is quite categorical about the result when the two are compared, though: "The reformers' desire to get things right according to an absolute standard of quality cannot be reconciled with standards of quality based on embedded practice."[96]

Scholars in Germany used to be judged by reviews of their work by other scholars, with respect shown by inviting those one respected to give a talk, even by deciding to apply for a job where they worked. Those were signals of esteem, ways to mark whether that scholar, or his–this is the past we are talking about–institute was worth anything. Today, "the belief has arisen that if one could only organize the evaluation of research cleverly enough, then one would be able to establish the significance of scholarly work through calculations expressed numerically, and through stable, formal, procedures."[97]

Still, how the discipline of history will render accountability to those who stand outside it remains unanswered. Is it possible, coming from outside, to find a way to measure quality or performance in a manner the (humanities) discipline being evaluated agrees with? The historians' refusal,

along with their evident lack of eagerness to suggest alternate measures they might find acceptable, suggest the answer is no. That leaves one with two discourses, the one held within and the one that comes from without. For now, these remain discourses in genuinely different languages.

JOHN BENDIX received his Ph.D. in Political Science from Indiana University, Bloomington, and currently teaches courses on U.S. (and Swiss) politics at the University of Zürich. He also works as a German-to-English translator/editor, specializing in social science and humanities material. He previously taught at Aachen, Bamberg, Göttingen and Innsbruck, and at Bryn Mawr, Haverford, Penn, and Reed. The present article is drawn from a set of case studies he recently wrote for a University of Basel project entitled "Entwicklung und Erprobung von Qualitätskriterien für die Forschung in den Geisteswissenschaften," research funded by the Conference of Swiss University Rectors.

Notes

1. I would like to thank Professor Christian Simon and Professor Martin Lengwiler for the time they have taken to lend their insights to this paper. All translations are my own; all websites cited were current as of midJune, 2012.
2. From 2001 to 2010, the number of full-time professors rose by 10 percent (37,600 to 41,500); in the same period, the number of *Lehrbeauftragte* increased by 75 percent (47,900 to 84,100). Added burdens are often ascribed to rising numbers of university administrators, but their number only rose by 19 percent (64,000 to 76,000) in the same time period. The ratio of administrators to full-professors 1.7 in 2001 and 1.8 in 2010) has remained unchanged. Figures from Federal Statistics Office, *Bildung und Kultur. Personal an Hochschulen* (Fachserie 11, Reihe 4.4, 2010), 24. On *wissenschaftliche Mitarbeiter*, see Jan-Martin Wiarda, "Enorm leidensfähig," *Die Zeit*, 8 Dezember 2011.
3. Such as to submit proposals to the Excellence Initiative (see note 11).
4. In a 2011 interview, Jena University sociologist Hartmut Rosa said: "Now when I want anything from my university administration, the first question is: What have you brought to the university recently in research funding? It has become problematic that our work is measured practically only now in terms of figures–third-party funding, number of PhD students graduated, publications." See "'Jeden Tag schuldig ins Bett:" Burn-out bei Professoren," *Die Zeit Online*; available at: www.zeit.de/2011/45/Burnout-Interview-Rosa.
5. Of course, much presented at these meetings is unrelated to higher education, but it is interesting that one can find quite self-reflective panels such as "Das organisierte Disziplin als Forschungsproblem: Perspektiven auf eine Geschichte des Historikerverbandes" (2012) or "Zeitgeschichtliche Forschung über Fächergrenzen und die Grenzen den Fachs" (2010).

6. Helga Welsh, "Higher Education in Germany. Fragmented Change amid Paradigm Shifts," *German Politics and Society* 28, no. 2 (2010): 53-70; Jürgen Schriewer. "Bologna und kein Ende: Die iterative Konstitution eines europäischen Hochschulraums," in Themenportal Europäische Geschichte (2006); available at http://www.europa.clio-online.de/2006/Article=146.
7. See European Commission, "Higher Education Institutions' Responses to Europeanisation, Internationalisation and Globalization: Final Report," (2005), 112; available at http://cordis.europa.eu/documents/documentlibrary/100124101EN6.pdf.
8. http://www.leidenranking.com/ranking.aspx; http://www.shanghairanking.com/index.html; http://www.timeshighereducation.co.uk/world-university-rankings/; accessed 2 October 2012.
9. For an excellent overview, see Andrae Wolter, "From the Academic Republic to the Managerial University: The Implementation of New Governance Structures in German Higher Education" (2006); available at http://www.boeckler.de/pdf /stuf_proj_leitbild_wolter_2007.pdf. See also Rosalind Pritchard, "Trends in the Restructuring of German Universities," *Comparative Education Review* 50, no. 1 (2006): 90-112; on the global position of German universities, David Baker and Gero Lenhardt, "The Institutional Crisis of the German Research University," *Higher Education Policy* 21 (2008): 49-64.
10. Because the historians' association rejected participating in the CHE a year after rejecting the WR, and because they had participated in CHE rankings up until then, this ranking system is relevant to the discussion.
11. Barbara Kehm and Peer Pasternack, "The German 'Excellence Initiative' and its Role in Restructuring the National Higher Education Landscape, in *Structuring Mass Higher Education: The Role of Elite Institutions*, eds. David Palfreyman and Ted Tapper (New York, 2008), 113-127.
12. One indicator is the proliferation of academic journals–Scientometrics, Research Evaluation, American Journal of Evaluation, Evaluation Review, Evaluation–devoted to the topic. Another is the increasing professionalization of evaluation practitioners themselves, and the increasing international adoption of the "Program Evaluation Standards" codified in 1981 by the Joint Committee for Standards for Educational Evaluation in the United States. For an example of such adoption, see Wolfgang Beywl and Thomas Widmer, *Handbuch der Evaluationsstandards*, 2. ed. (Opladen, 2000).
13. Bruno Frey, "Evaluierungen, Evaluierungen… Evaluitis," *Perspektiven der Wirtschaftspolitik* 8, no.3 (2007): 207-220; available at http://www.bsfrey.ch/articles/462_07.pdf.
14. In an 2011 interview, part of a Swiss research project with which the author is involved, one very senior academic in the humanities complained he had been involved in six evaluations already–and the year was only half over.
15. Hartmut Rosa (see note 4).
16. "Historiker: Exzellenzinitiatve hält Forscher vom Arbeiten ab," Radio interview, 14 June 2012; available at http//www.dradio.de/dkultur/sendungen/thema/1784082/. The government's efforts have not convinced some academics: "Despite the Excellence Initiative, the ability of German universities to compete internationally has not improved." Johannes Balve, "Ein Modell mit Zukunft. Kritische Anmerkungen zur heutigen deutschen Universität," *Forschung und Lehre* 2 (2012).
1. See Ralf Adelmann, "'Oh, oh, oh, let's count some more: Hochschulrankings als mediale Form," *Zeitschrift für Medienwissenschaft* 4, No. 1 (2011): 178-182; available at http://doi_10.4472_zfmw.2011.0014.pdf.
18. Among the prominent authors here, one can list Henk Moed, Anton Nederhof, Anthony van Raan, Hans-Dieter Daniel, or Wolfgang Glänzel.
19. Wissenschaftsrat, "Empfehlungen zur Bewertung und Steuerung von Forschungsleistung," Drs. 1656-11 (2011); available at http://www.wissenschaftsrat.de/download/archiv/1656-11.pdf

20. Thomas Widmer, "Evaluationsansätze und ihre Effekte: Erfahrungen aus verschiedenen Politikfelder," in *Wissenschaft unter Beobachtung: Effekte und Defekte von Evaluationen*, eds., Hildegard Matthies and Dagmar Simon (Wiesbaden,2008), 267.
21. These lists are drawn from Thomas Widmer's excellent survey (see note 20), 273-276.
22. Frey (see note 13), 207. He notes that the key elements, and what irritates, is the "after the fact" assessment–in other words the often summative rather than formative character of evaluation–and that it is "external" experts who are scrutinizing.
23. See Martina Röbbecke, "Evaluation als neue Form der 'Disziplinierung'–ein nicht intendierter Effekt?" in Matthies and Simon (see note 20), 161-177.
24. Wendy Espeland and Michael Sauder, "Rankings and Reactivity: How Public Measures Recreate Social Worlds," *American Journal of Sociology* 113, no. 1 (2007): 33. See also Richard Münch, "Stratifikation durch Evaluation: Mechanismen der Konstruktion von Statushierarchien in der Forschung," *Zeitschrift für Soziologie* 37, no. 1 (2008): 60-80. For an analytic/personal view, see Helmut Wiesenthal, "Evaluation als Organisationslernen" (2007); available at http://www.hwiesenthal.de/downloads/evaluation_wzb.pdf
25. The Heisenberg principle, in other words, applied to the humanities…
26. See Wissenschaftsrat (see note 19), 9.
27. Wissenschaftsrat, "Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften," Drs. 10039-10 (2010), 22; available at http:// www.wissenschaftsrat.de/download/archiv/10039-10.pdf.
28. Or at least one faction within it that has risen to prominence (Werner Plumpe, personal communication).
29. Not just with the WR, either; a year later, it also ceased cooperating with the CHE rankings.
30. Wissenschaftsrat, "Organisation and procedures" (2011a); available at http:// www.wissenschaftsrat.de/1/about/organisation-structure-and-methods/.
31. Hans Weiler, "Ambivalence and the Politics of Knowledge: The Struggle for Change in German Higher Education," *Higher Education* 49 (2005): 182.
32. Ibid., 183. Weiler knew whereof he spoke, as he served for six years (1993-1999) as the first rector of the newly constituted Viadrina University in Frankfurt/Oder.
33. On a personal note, having myself experienced Paul Feyerabend's anarchic approach to teaching at Berkeley, which was in keeping with his intellectual stance against method, I sympathize with what university administrators sometimes have to put up with.
34. See Andreas Stucke, "Der Wissenschaftsrat," in *Handbuch Politikberatung*, Svenja Falk et al., (Wiesbaden, 2006), 252. For an interesting, personal view, see Herbert Gassert, "Wirtschaft im Dialog: Als Externer im Wissenschaftsrat," *Gegenworte* 6 (Herbst 2000).
35. The WR's science commission has thirty-two members directly appointed by the German president, eight of whom are eminent public figures proposed jointly by federal and state governments. The remaining twenty-four are scientists whose names are proposed by the Helmholtz Association, the Fraunhofer Society, the Max-Planck Society, the Leibnitz Association, the German Research Foundation, and the German Rectors' Conference. The science commission thus reflects the interests of numerous disciplines, institutions, organizations and political entities. Germany conducts a great deal of technical and scientific research in (more than 200) non-university institutes organized under umbrella networks like the Helmholtz or Leibnitz Associations, but few of these institutes are devoted to the humanities or the social sciences.
36. Wissenschaftsrat, "Steering Group Report on the Pilot Study Research Rating in Chemistry and Sociology," (2008a); available at http:// www.wissenschaftsrat.de/download/archiv/pilot_8893-08_steering-group.pdf.
37. Ibid., 13.
38. Wissenschaftsrat (see note 27), 21.
39. Wissenschaftsrat (see note 36), 17.
40. Wissenschaftsrat, "Pilotstudie Forschungsrating Soziologie. Abschlussbericht der Bewertungsgruppe," (2008b): 18; available at http://www.wissenschaftsrat.de/ download/

Forschungsrating/Dokumente/Grundlegende_Dokumente_zum_ Forschungsrating/8422-08.pdf

41. Wissenschaftsrat (see note 27), 22.
42. Wissenschaftsrat, "Forschungsleistungen deutscher Universitäten und ausseruniversitären Einrichtungen in der Soziologie. Ergebnisse der Pilotstudie Forschungsrating des Wissenschaftsrats," (2008c), 11-15; available at http://www.wissenschaftsrat.de/download/Forschungsrating/Dokumente/Pilotstude_Forschungsrating_Soziologie/pilot_ergeb_sozio.pdf; Wissenschaftsrat (see note 40), 18.
43. Wissenschaftsrat, "Results of the Pilot Study for a Research Rating in Sociology: Tables," (2008d); available at http://www.wissenschaftsrat.de/download/archiv/pilot_8665-08_Sociology.pdf.
44. Gerhard Fröhlich, "Informed Peer Review–Ausgleich der Fehler und Verzerrungen?" in *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung*, Hochschulrektorenkonferenz (2006), 193-204 (); available at http://hdl.handle.net/10760/8838.
45. Wissenschaftsrat (see note 27), 25.
46. Wissenschaftsrat (see note 42), 5, 19.
47. A further lack was the reason why no comprehensive bibliometric analysis was conducted. Analysis of the 739 publications submitted to the WR pilot discovered only 9 percent of them were in journals used for creating citation indexes: bibliometry simply would not have reflected actual publications produced in sociology; Wissenschaftsrat (see note 42), 16.
48. Each research unit could submit two publications, three if it had four-six researchers, and four if seven-nine. The mean number of researchers per unit was 5.7, hence three publications, and the total publications submitted divided by total research units (739/254) also yields just under three publications; Wissenschaftsrat (see note 42), 16.
49. The first quote is from Wissenschaftsrat (see note 27), 26; the second quote is from Patrick Riordan, Chrisian Ganser and Tobias Wolbring, "Zur Messung von Forschungsqualität. Eine kritische Analyse des Forschungsratings des Wissenschaftsrats," *Kölner Zeitschrift für Soziologie* 63 (2011): 147-172, here 152.
50. Stefan Hirschauer, "Peer Review Verfahren auf dem Prüfstand," *Zeitschrift für Soziologie* 33, no. 2 (2004): 62-83; Michèle Lamont, *How Professors Think: Inside the Curious World of Academic Jjudgment*(Cambridge, 2009), 138-141.
51. Lamont (see note 50), 231.
52. Riordan et al. (see note 49), 160-61.
53. He went on to observe that since two-thirds of the participating institutions had at least one unit that was "research-weak," this could mean sociology professors were recruiting only those "whose light is dimmer than their own." See Jürgen Kaube, "Durchschnittlich herrscht Mittelmaß," *Frankfurter Allgemeine Zeitung*, 22 April 2008; available at http://www.faz.net/artikel/C31373/soziologie-rangliste-durchschnittlich-herrscht-mittelmass-30098244.html.
54. Wissenschaftsrat, "Ergebnisse des Forschungsratings Elektrotechnik und Informationstechnik," Drs. 1372-11 (2011b); available at http://www.wissenschaftsrat.de/download/archiv/1372-11_Ergebnisse_ETIT.pdf.
55. Werner Plumpe, "Forum Qualitätsmessung: Stellungnahme zum Rating des Wissenschaftsrates aus Sicht des Historikerverbandes," *H-Soz-u-Kult*, 18 May 2009; available at http://hsozkult.geschichte.hu-berlin.de/forum/id=1101&type= diskussionen.
56. Wissenschaftsrat (see note 27).
57. The acronym stands for *Humanities– Sozial und Kulturgeschichte*. The site is affiliated with H-Net, an international network of humanities and social science researchers; the German website has existed since 1996.
58. Rüdiger Hohls and Claudia Prinz, "Forum Qualitätsmessung: Editorial 'Qualitätsmessung, Evaluation, Forschungsrating. Risiken und Chancen für die Geschichtswis-

senschaften?'" *H-Soz-u-Kult*, 12 May 2009; available at http://hsozkult.geschichte.hu-berlin.de/forum/id= 1098& type=diskussionen.

59. Claudia Prinz and Rüdiger Hohls, "Qualitätsmessung, Evaluation, Forschungsrating. Risiken und Chancen für die Geschichtswissenschaften?" *Historisches Forum* 12 (2009); available at http://edoc.hu-berlin.de/histfor/12/.
60. Jürgen Gerhards and Gert Wagner, "Forschungsrating: Es geht um die Qualität" (2008); available at http://www.academics.de/wissenschaft/forschungsrating_es_geht_um_die_qualitaet_30823.html.
61. Elke Lütkemeier, "Forum Qualitätsmessung: Das Forschungsrating des Wissenschaftsrats," *H-Soz-u-Kult*, 25 May 2009; available at http://hsozkult.geschichte.hu-berlin.de/forum/id=1105&type=diskussionen.
62. Lutz Raphael, "Forum Qualitätsmessung: Probleme und Chancen der Forschungsbewertung im Fach Geschichte," in: *H-Soz-u-Kult*, 20 May 2009; available at http://hsozkult.geschichte.hu-berlin.de/forum/ id=1106&type=diskussionen.
63. Angela Borgwardt, *Rankings im Wissenschaftssystem: Zwischen Wunsch und Wirklichkeit* (Bonn, 2011), 28. From 1999 to 2004, these rankings were published in *Stern*; since 2005, they have appeared in *Die Zeit*. See CHE-HochschulRanking 2011; available at http://www.che-ranking.de/cms/?getObject=50&getLang=de.
64. Ben Wildavsky, "Why We Need College Rankings," (2010) Zocalopublicsquare.org; available at http://www.youtube.com/watch?v=OS2wbgf-Uus.
65. A point made by sociologist Hartmut Rosa as well (see note 4). These three indicators are calculated both in absolute (per year) and relative (per researcher) terms, leading to separating programs into high, middle, and low categories. A table at the beginning of the CHE report on Anglistik highlights those programs with the highest combined indicators. Next to it is a separate table listing which programs have a high reputation among peers in the discipline. How this reputation is established is not reported, though one Anglistik professor at a German university remembers being called by the CHE and asked: "If you had a child you were sending off to study in your discipline somewhere else in Germany, where would you recommend they go?" See Section C (Anglistik) in Sonja Berghoff et al. "CHE-Forschungsranking deutscher Universitäten 2008," CHE-Arbeitspapier 114 (2008); available at http:// www.che.de/downloads/CHE_AP114_Forschungsranking_2008.pdf.
66. Müller-Böling's views noted here are summarized in Hohls and Prinz (see note 58) and cited in Borgwardt (see note 63), 28, 30.
67. The point supports the analysis Weiler provides (see note 31).
68. Sonja Berghoff, "Forum Qualitätsmessung: Das CHE ForschungsRanking in den Geisteswissenschaften," *H-Soz-u-Kult*, 12 June 2009; available at http://hsozkult.geschichte.hu-berlin.de/forum/id=1123&type=diskussionen.
69. Simone Lässig, "Forum Qualitätsmessung: Stellungnahme des VHD zum CHE-Ranking der deutschen Geschichtswissenschaft," *H-Soz-u-Kult*, 23 Sept 2009; available at http://hsozkult.geschichte.hu-berlin.de/forum/id=1171 &type=diskussionen.
70. Bernhard Gotto, "Forum Qualitätsmessung: Stellungnahme von Mitarbeitern des IfZ zum Vorhaben eines 'Forschungsratings' der Geschichtswissenschaften," *H-Soz-u-Kult*, 16 June 2009; available at http://hsozkult.geschichte.hu-berlin.de/forum/ id=1122& type=diskussionen.
71. Plumpe (see note 55).
72. Raphael (see note 62).
73. Simone Lässig, "Forum Qualitätsmessung: Stellungnahme des Verbandes der Historikerinnen und Historiker Deutschlands (VHD) zum Pilotprojekt des Wissenschaftsrates 'Forschungsrating in den Geisteswissenschaften,' *H-Soz-u-Kult*, 7 July 2009; available at http://hsozkult.geschichte.hu-berlin.de/forum/id=1172&type=diskussionen.
74. Lässig (see note 73).
75. Gotto (see note 70).
76. Ibid..

77. Plumpe (see note 55).. For a good overview of his own thinking, see "Ratings fördern das strategische Verhalten: Fragen an Professor Werner Plumpe, Vorsitzender des Deutschen Historikerverbandes," *Forschung und Lehre* 8 (2009): 570-571.
78. Hohls and Prinz (see note 58); Plumpe (see note 55)..
79. As in Ernst Schulin, *Traditionskritik und Rekonstruktionsversuch* (Göttingen, 1979)..
80. Plumpe (see note 55); Lässig (see note 73); Raphael (see note 62).
81. Radegundis Stolze, "Geisteswissenschaften als Forschungsparadigma: Rezension von *Humanities. Was Geisteswissenschaft macht. Und was sie ausmacht* von Marcus Beiner,;" Available at http://www.relue-online.de/neu/2009/04/geisteswissenschaften-als-forschungsparadigma/. This is my own view, not one raised by the historians, though I daresay some would agree with it.
82. Raphael (see note 62). The next three paragraphs are a distillation of his views.
83. Ulrich Herbert, "Forum Qualitätsmessung: Ulrich Herbert und Jürgen Kaube: Die Mühen der Ebene. Über Standards, Leistung und Hochschulreform,: *H-Soz-u-Kult*, 14 May 2009; available at http://hsozkult.geschichte.hu-berlin.de/forum/ id=1100&type=diskussionen. Herbert served on the WR from 2001 to 2007; his stance here may have been part of a difference of opinion with his WR colleague Lutz Raphael.
84. Charles Maier, *The Unmasterable Past: History, Holocaust, and German National Memory* (Cambridge,1988).
85. See Otto Hüther et al., "Hochschulautonomie in Gesetz und Praxis" (2011); available at http://www.academics.de/wissenschaft/hochschulautonomie_in_gesetz_und_praxis_51015.html.
86. Stefan Lange and Jochen Gläser, "Performanzsteigerung durch Selektivität? Erwartbare Effekte von Forschungsevaluationen an deutschen Universitäten im Lichte internationaler Erfahrungen," *Dms–der moderne Staat* 2 (2009): 411-432.
87. Hohls and Prinz (see note 58).
88. Jürgen Kocka, "Historische Sozialwissenschaft. Auslaufmodell oder Zukunftsvision?," (Oldenburg,1999).
89. Michael Oakeshott, "History and the Social Sciences" (1936); available at http://www.michael-oakeshott-association.com/pdfs/ mo_history_social_sciences.pdf.
90. The Act is at 20 U.S.C. §952, the study is Marvin Goldberger, Brendan Maher and Pamela Flattau, eds. *Research Doctorate Programs in the United States: Continuity and Change* (Washington, 1995).
91. See Max Weber, *The Protestant Ethic and the Spirit of Capitalism*, trans. Talcott Parsons (London, 1992) The *locus classicus* for historians is Marc Bloch. *The Historian's Craft* (New York, 1954). His original 1949 French title, however, was *Apologie pour l'histoire ou Métier d'historien.*
92. Richard Sennett. *The Craftsman* (London, 2008), 243.
93. Tilmann Warnecke, "Historiker boykottieren Forschungsranking," *Der Tagesspiegel* , 10 July 2009; available at http://www.tagesspiegel.de/wissen/evaluation-historiker-boykottieren-forschungsranking/1554428.html.
94. Lamont (see note 50), 80, 85.
95. Sennett (see note 92), 51.
96. Ibid., 52.
97. Kaube (see note 53).