In his remarkable account of the rise of the research university, the historian William Clark writes: “A [late 18th century] German…academic had to pass muster with bureaucratic or rationalized criteria for appointment, which included productivity in publication, diligence in teaching, and acceptable political views and lifestyle. But to achieve success, one had also to acquire fame, be in fashion, and display ‘originality,’ a spark of genius, in writings. This became a new type of academic charisma tied to writing for ‘applause’ and ‘recognition.’ The modern academic emerged…from the cultivation of this new legible charisma” (Clark 2006: 3-4). Several aspects of this quote are striking, among them the conceivably oxymoronic notion of academic charisma; does this really accord with our self-image? While arguably the acceptability of our politics or lifestyles has become less of an issue, “productivity in production” and “diligence in teaching” remain key to both our careers and our discipline, as do institutional interests in and
mechanisms for monitoring our performance. How our “charisma” becomes “legible,” that is, how our accomplishments are recognized, measured, and valued, has changed considerably over the past two centuries, as have the institutions and actors engaged in such supervision. And, finally, the tensions between creativity and commensurability, between research agenda and regulatory regime, remain strong, although the contours of those tensions may change dramatically.

In this talk I want to explore something of the contemporary context within which anthropological charisma is being recognized, assessed, and acted upon. As all of you know far more directly and thoroughly than I, the emerging “European convergence” and the institutions and practices that it both depends upon and engenders, are complex and deeply consequential. I am truly indebted to Andres Barrera and his colleagues on the organizing committee, to the University, to the various groups funding this gathering, and to all of you for making these conversations possible. I know I will be learning a great deal and am very appreciative of the opportunity to add a bit to our discussion of the current state and possible trajectories of our field in Spain and in Europe.

In this talk I’ll start with a brief participant-observer account of the International Benchmarking Review of UK Social Anthropology conducted in 2006 for the Economic and Social Research Council/UK (hereafter ESRC), Association of Social Anthropologists and the Royal Anthropological Institute (ESRC 2006). I chaired the review committee and was principal author of the final report. Today I’ll not be talking about the manifest accomplishments of British social anthropology that we noted. I’ll rather outline some of the methodological and interpretive dimensions of how the review was conducted. I’ll then turn to a discussion of current debates concerning the value and
limitations of the two principal assessment methodologies, peer review and formal
metrics (especially citation analysis and journal impact factors) – and pointing towards
some strategies for more appropriate and helpful policies. A remarkable – and
remarkably consistent – range of reports and analytical articles has been emerging over
the past few years concerning academic evaluation. I’m hoping that several of the
publications might be helpful for you, both for intradisciplinary conversations and as a
resource in talking with academic managers. Finally I’d like to suggest that, to use
Andres Barrera’s phrase, the particular “epistemological and methodological strengths”
of our shared discipline can help us contribute generatively to these critical debates.

Benchmarking UK Social Anthropology

In 2005 the ESRC, ASA, and RAI collaborated in setting up a Steering
Committee and working to “benchmark the quality and impact of Social Anthropology
research in the UK against international standards” (ESRC 2006: 3). The Steering
Committee then invited a number of non-UK anthropologists to conduct the review.
Eight of us (from Australia, Norway, Mexico, and the US) constituted the committee,
which I chaired. The core objective of this review, to quote the letter of September 27,
2005, from the Steering Committee inviting the visiting panel’s participation, was “to
benchmark UK Social Anthropology against international standards,” that is, to gain a
clearer and empirically grounded understanding of how the quality, visibility, impact, and
broader contributions of British Social Anthropology figured within the field worldwide.
A further goal of such benchmarking reviews, of which this was the first, was to consider
both likely and possible future trajectories for the field within the UK and, particularly,
“to identify ways of enhancing capacity and help promote future research agendas.”

These goals imply a range of crucial questions: How good is contemporary Social Anthropology in the UK, what is its international standing, and what particularly significant strengths, foci, and gaps are evident? What kinds of evidence are available, and how can such evidence most effectively be explored and used in providing an illuminating and useful assessment? How might the discipline most effectively be developed, and what courses of action might be suggested for multiple actors: departments and the profession, university administrators, funding agencies, and current and future users of anthropological knowledge? While this benchmarking review was commissioned by the Economic and Social Research Council and was intended to be useful in ESRC’s strategic planning, both ESRC and the committee members clearly hoped that it would be of interest and value to other funding agencies, stakeholders, administrators, and colleagues within the discipline.

Two further general elements of our charge should be noted. First, this benchmarking was to be a review of the field as a whole within the United Kingdom. As our charge noted, it was “important to emphasize that the discipline as a whole is being assessed, not individual departments or just ESRC investments.” We were asked not to rank or otherwise measure departments against each other but to locate UK Social Anthropology within the discipline worldwide. This marked a real and significant departure from earlier, more routine forms of comparative assessment and from the differential rewards that such practices resulted in for departments. While the general structure of our review drew in large part on visits to selected individual departments and on supplementary materials from other departments and users, our goal was consistently
to get a stronger, more subtle, and empirically better grounded sense of the field as a whole.

A second key element of the benchmarking was that we were specifically asked to pursue it as a qualitative exercise. As our charge stated, “The qualitative nature of the review is particularly important. It will help balance the UK Government’s increasing use of metrics, especially citation counts, to make judgments about research impact and standing.” Or, as Ian Diamond noted in his comments at the Steering Committee meeting of August 10, 2005, relying “on qualitative rather than quantitative measures [was necessary] in order to complement the assessment made by the Research Assessment Exercise (RAE) and avoid some of the problems which ESRC has already identified with existing quantitative indications such as citation indexes.” As I will discuss at some length later, ESRC was not alone in its concerns regarding such quantitative measures. Indeed there has been an explosion in the critical literature since the time of our study.

The benchmarking review of Social Anthropology was the first such qualitative review to be conducted for ESPC. It was necessarily something of a prototype exercise, then, and, as our discipline of Social Anthropology is definitively associated with qualitative research methods and especially with ethnography, the long-term, intensive description and analysis of complexly embedded relationships between, on the on hand, local and broader contexts and, on the other, the specifics of social and cultural practice, Social Anthropology was clearly an appropriate starting point. The panel conducted its work at a time of considerable transition in UK academia, as a new system of figuring and meeting costs was just being introduced. The steering committee commissioned a very useful briefing paper from Dr. David Mills (Mills 2006) providing current data on
disciplinary research funding, staff and student demographic trends, and career paths in both the academy and non-academic professions, as well as a very useful introduction to the RAE system and the results of the 2001 RAE. Mills’s paper and the other literature that it indexed were invaluable in giving us a sense of the contours of the field. When compared, for example, with the US, it is clear that UK Social Anthropology had over the past twenty or so years developed an empirically rich, reflective, and critical literature on the field itself within the context of higher education and research policy and practice, one deeply grounded in both social and institutional contexts.

As an international group, our panel was able to view the UK scene within an international comparative framework; what might be taken as routine and expectable by our UK colleagues could be – and often was – quite striking and distinctive to panel members. We also could, crucially, rely on our own expertise – years of reading in the field, conducting research, meeting with colleagues both in our home countries and internationally – in working towards a picture of the field in the UK.

A further aspect of our external viewpoint had to do with the possibilities of reframing and in so doing extending the semantic range of key terms in the benchmarking process itself. In part this is because rethinking assumedly transparent and stable terms is a hallmark of anthropological analysis. In larger part, however, it is because what we learned in the course of our various campus conversations made it clear, when it comes to evaluations, one size does not fit all. When considered vis-à-vis the social sciences more generally, Social Anthropology is distinctive along several dimensions. For example, the time required for design, preparation (including language learning), fieldwork, analysis, and publication is considerable and likely to be longer than for other fields; expectations
concerning rates of productivity and research “turn around times” should accordingly be
tailored to the field’s temporalities. While the time required for anthropological research
is considerable, the financial scale for funding such research is often much lower than in
other social sciences (and certainly in the natural sciences). In financial terms, Social
Anthropological research recurrently provides a bigger bang for fewer bucks, a factor that
should be taken into account when thinking about what kinds of risks a funding agency
might be willing – or eager – to take. How one might weigh questions of impact and
value might similarly vary across fields, a point to which we will return below. We are
not claiming here that Social Anthropology is singular in its singularity. This is rather to
argue that review measures and practices take into account the local particularities of each
discipline or interdisciplinary cluster – and to think through strategies appropriate for and
likely to be revelatory about each field.

The steering committee also laid out a clear set of criteria and topics
recommended for structuring both our meetings with departments and our final report. In
addition, we were provided with responses to a questionnaire distributed to Heads of
those departments we were unable to visit, as well as the results of a use or stakeholder
survey. Perhaps most significant in terms of background knowledge, however, were the
breadth and depth of knowledge of anthropology that the eight members of the review
committee brought to the benchmarking; we represented a considerable variety of topical
and regional expertise and were all quite familiar with the contributions of British Social
Anthropology, both historically and in the present, to the development of our field.

The central element of the review consisted of visits to twelve departments of
Social Anthropology, departments that had been selected by the steering committee. Each
of the twelve departments had provided its own background materials for the committee. In each of these visits we worked to explore the three core topics of central concern to ESRC and the steering committee: research issues, research capacity, and impacts on policy and practice. We were able to spend three or four hours at each site, meeting separately with academic staff (or some portion thereof) and with postgraduate students and postdoctoral fellows. Formats varied from school to school, with some departments making fairly elaborate presentations and others moving more directly into conversation with the committee. In both kinds of meetings our colleagues were consistently articulate, reflective, and forthcoming. We found the presentations helpful but considered the discussions particularly illuminating, especially those with postgraduate students and postdocs whose visions of and commitment to the field seem particularly significant when thinking about its future course.

In reflecting upon both our oral presentation to the Steering Committee and our written draft report, the panel returned to the methodological questions with which we initially framed our work: how might such benchmarking reviews most comprehensively, subtly, and accurately reflect the contours of a particular discipline? I want here to highlight several qualitative criteria that emerged during our review. First, intellectual pluralism seemed central to the remarkable accomplishments of the field, and we would imagine that it is also central to imagination and creativity in other disciplines as well. Such reviews should pay attention to a range of topics and methods and to the interrelationships among them, and especially to synergies in research and funding. Such variety – and, at times, such apparent and concomitant incommensurability across projects and programs – should be approached as a strength, rather than regarded as a
bother to measure. Such generative pluralism can be documented if not measured in the usual sense, and it is clearly key to UK Social Anthropology’s success. A further qualitative criterion had to do with the ongoing and long term influence of UK Social Anthropology internationally through the training of non-British students and employment and collaborative research abroad, particularly in Europe. It is clear that this is an area in which the field has long excelled, and it was crucial that such transnational training and collegial work be taken into appropriate account in any review.

We also wanted in our report to note that, in our view, many earlier styles in assessment were very problematic not because they were quantitative per se but because they were restrictively, reductively so. Turning to one or two numerical measures, for example, citation indexes, as proxies for scientific and scholarly excellence is a deeply flawed approach, a point to which I’ll return. Taking a much wider range of more subtle, countable evidence into account, however, can complement qualitative findings and help substantiate and refine them. In terms of research capacity, for example, such variables as the proportion of staff in the field who are considered research intensive (especially when viewed comparatively across fields) are critical. Better and meaningful tracking of the post graduate careers of Ph.D.s well beyond their first employment would help speak to both training and impact. A recently released study of US Ph.D.s in six social science fields five and more years after their degrees (Nerad 2008) provides both a substantively rich account of contemporary careers and of the fit – or misfit - between preparation and professional trajectory and an invaluable model of principled and illuminating quantitative review. I should note that in our view appropriate and relevant quantitative
data can both inform and complement more qualitative findings, but that they could not replace it.

Several procedural elements of a broadly qualitative review seemed particularly important to the panel. First, any measures and evaluations of a discipline should be understood vis-à-vis the specific contexts and characteristics of that discipline. One size doesn’t fit all, and the meaning of any measurement is certain to be discipline-sensitive. Second, proxies can be mischievous; letting one variable stand for an entire universe often leads to real misunderstanding. And any measurable variables should be considered not only in relation to each other but also, more significantly, as a complement to broader qualitative findings. Third, comparative perspectives, whether cross-nationally within the same discipline or across disciplines within a specific national context, are invaluable, in large part because of the unexpected and revelatory moments of recognition or, occasionally, surprise that they afford. Finally, central to the success of a large-scale qualitative review such as the UK Benchmarking exercise is that it involves a group of colleagues in serious inquiry, reflection, and consultation. Our process was truly deliberative, one in which we were very fortunate in our interlocutors – within the panel itself, at the ESRC, ASA and RAI, and, especially, among UK social anthropologists. Such collaborative deliberation not only allowed for but required taking time, context, disciplinary cultures, and current practice into serious and subtle account. A crucial characteristic of any qualitative review is the necessity of such shared thinking and argument, a kind of productive deliberation all too easily circumvented by reliance on the reductive measures characteristic of the more usual review.
Meetings and metrics

I want to begin this section with a quote from Professor Eric Thomas, Chair of the Research Policy Committee, Universities UK. In his foreword to a report concerning the likely shape of a new framework for assessing and supporting university research in the sciences after the 2008 Research Assessment Exercise has been concluded, Thomas writes “It is widely expected that the ratings will initially be derived from bibliometric-based indicators rather than peer review. These indicators will need to be linked to other metrics on research funding and on research postgraduate training. In a final stage the various indices will need to be integrated into an algorithm that drives the allocation of funding to institutions. The quality indicators would ideally be capable of not only informing funding but also providing information for higher education institutions and stakeholders. They are also expected to be cost-effective to produce and should reduce the current assessment burden on institutions” (Evidence 2007: 2). While this vision of a wholly algorithmic future is unlikely to be realized even in the UK, the virtues of clarity, comparability, and cost effectiveness it is taken to afford resonate strongly with current desires – if not always with practice – in higher education and research institutions in much of the world. Thomas also juxtaposes the practices of peer review to a system of indicators and metrics, a distinction echoed in a recent US National Research Council report on strategies for assessing science: “Two generic decision strategies are available for assessing scientific progress and setting research priorities: (1) applying analytic techniques, such as benefit-cost analysis, bibliometric analysis, and decision analysis, and (2) using deliberative processes, such as those of expert review panels” (Feller 2007: 2).
In contrast to the Universities UK report, however, Feller and his colleagues argue compellingly that “(b)ecause of the uncertainties about the reliability and validity of all existing methods, we recommend a strategy for decision making that relies primarily on processes and secondarily on methods…Analytic techniques for quantifying scientific progress can provide useful input to decision-making deliberations, but they should not be used as substitutes for the necessary judgment” (Feller 2007: 2).

In this section of my talk I want to sketch out some aspects of both peer review and “analytic” approaches to evaluation and take note of some of the recurrent and emerging criticisms of each approach. To anticipate my discussion – and to reassure you that I hope to be counseling not despair but a sense of principled possibilities – the algorithmic model proposed by Universities UK and others has catalyzed very strong, empirically rich and technically sophisticated resistance. I’ll close with brief accounts of future directions for scholarly and scientific evaluation.

Turning first to peer review, “the practice by which the worth of research is evaluated by those with demonstrated competence to make a judgment” (British Academy 2007: ix), I should note that there is a great deal of variety in what is being reviewed, who the relevant actors are, how the review is conducted, and what the consequences might be. Collegial consultation, whether pursued long distance with individual respondents or taking place in face-to-face meetings, figures centrally in the allocation of postgraduate, postdoctoral and other Fellowships (where those being reviewed are unlikely actually to be the peers of those doing the reviewing) and of research funds through the proposal evaluation process. Peer review is also central to the evaluation of manuscripts for possible journal and book publication, constitutes a critical
element in individuals’ personnel evaluations, and is key to departmental program
reviews. The UK RAE is perhaps the largest scale example of his last variety of peer
review. Peer review combines individual judgment with collaborative deliberation,
whether pursued through the mail, online, or over several days in a windowless
committee room. Indeed, all of us here are most likely overly familiar with peer review,
as we have been applicants, examinees, and assessors many times over. At the same time,
the very familiarity of the practice in its many forms can make it invisible save as a
source of exasperation, tedium, and occasional satisfaction. The historian of science
Mario Biagioli (2002: 11) has asked “Why do we tend to perceive peer review as either
good or bad, helpful or obstructive, but not as one of the fundamental conditions of the
construction of academic knowledge and the construction of its value?...it is puzzling that
academics do not seem to recognize peer review as their distinctive kind of discipline,
that is, as something that is simultaneously repressive, productive, and constitutive of
their knowledge.” I’ll first try to lay out some of recurrent critiques of peer review and
then turn briefly to Biagioli’s questions concerning the deeper constitutive roles peer
practice may play.

Peer review based at government agencies has been the subject of considerable
critical study. Some literature has concentrated on particularly flagrant ethical abuses
(Bell 1992), but most scholarship has been concerned with more everyday structural and
procedural problems within the system (Chubin 1990; Cole and Cole 1981; Roy 1985).
Peer review is sometimes seen as providing too much opportunity for self-interest, given
the centrality of individual position taking. Over time concerns about fairness
(Government Accounting Office 1994) have alternated with a sense that panelists are
often not fully up to the task, as they might lack the degree of specialized knowledge and sophistication required to appraise cutting edge science or scholarship responsibly. A perceived dichotomy between generalists and real experts has proven particularly significant in trying to explain the particular challenges of interdisciplinary panels (Feller 2006; Lamont 2006).

A particularly salient and consequential term in critiques of peer review is subjectivity, that is, the key role of individuals in coming to their own conclusions, even in contexts of intensive consultation and joint consideration. There is frequently a tension between this sense of individual judgment and concerns that such decisions make sense, that is, reflect appropriate outcomes reasonably reached. Reviewer subjectivities are rarely in free play, as agencies very actively work to shape both written responses and panel discussions. At both the National Institutes of Health (NIH) and the National Science Foundation in the US, for instance, great attention is paid not only to the general criteria that should guide commentary and discussion but also to the more specific rating systems in use. The most recent NIH self-study of peer review (National Institutes of Health 2008; Pollick 2008), for example, draws heavily upon work in psychometrics to suggest making available a wider array of criteria that must be addressed by both panelists and external reviewers. Panel meetings often begin with calibration exercises, that is, with trial scorings to see how one’s ratings and scores align with those of others. One recurrent concern about subjectivity is that some individuals may always give high scores while others give low ones; calibration is one strategy to bring panelists into closer accord. On some kinds of panels even more explicit opportunities for self-monitoring and regulation are provided; on postgraduate fellowship panels, for example, participants are
routinely informed about how many applications they have read and what their average scores are. This ongoing struggle with subjectivity, to balance individual judgment with concerns for fairness, reasoned discussion, and good science and scholarship, is a hallmark of peer review in practice.

Two other elements figure in criticism of peer review. One has to do with its closed nature; only a few scholars or scientists are involved, and such small numbers in themselves might accentuate the influence of particular individuals. With the advent of electronic resources, a number of scholars have argued that larger scale peer review online would make for better results Stevan Harnad (2000; 2008), one of the most active advocates for such open reviewing, has argued that “(t)he traditional closed peer-review process is hampered by limited competence and the sometimes conflicting of editors and referees. It frequently leads to a retardation and loss of scientific information.”

A final very significant concern is that highly innovative grant proposals are either actively stifled (Roy 1985) or for other reasons do not rise to the top. My own ethnographic research on funding panels suggests that the premium placed on amity around the committee table and a commitment to comparative discussion, one which particularly innovative and idiosyncratic proposals might make difficult, may indeed keep some of the most creative work from being supported (Brenneis 1999; Brenneis 2004).

I want briefly to touch further on my own work and note three key aspects of the funding committees in which I worked as a literal participant-observer. First is that peer review is deeply embedded in and shaped by complex social relationships. “Peer review” can actually seen as multidimensional, as “we (as panelists) were both reviewing the
work of our peers and, in our discussions, concerned with being peers…Participation in such decision making made one, for the moment at least, an equal. In peer review we jointly constituted an ephemeral peership among ourselves as well as vis-à-vis those whom we were evaluating” (Brenneis 1999: 141). Further, panel discussions on the interdisciplinary panel on which I served for three years were marked by a remarkable degree of deference across disciplines; “our discussions were generally characterized by amiability and a willingness to listen to what others had to say. Ironically, this willingness to listen often served to limit what one might say or how strongly one might be willing to say it…the highly collaborative nature of panel work makes disagreement difficult” (Brenneis 1999: 142). Such ongoing social process is likely to be no surprise to social anthropologists, but it has rarely been noted elsewhere in the literature.

A second comment has to do with the struggle with subjectivity noted above. As panelists we learned how to become disciplined readers and actors. (Ewald 1991: 32) has written of the norm as a “principle of communication, a highly specific way of reducing the problems of intersubjectivity.” To varying extents – most markedly in the postgraduate fellowship panels – panelists are plunged into a normalization process. In so doing we “acquire a new language in different ways: by attempting to standardize those criteria by which we evaluate proposals; by participating in a fiction of objectifiability engendered, at least in part, by the negotiation of what constitutes [licit and relevant] “information;” and by the adoption of fungible categories of discussion and comparison” (Brenneis 1994: 32).

Finally, as panelists, applicants, or referees we are often engaged with a variety of documents, whether a recommendation form or a scoring sheet. These are, to use Richard
(Harper 1998) term, mundane documents, ones that engender routine responses, both from those filling them out and from subsequent readers. These documents are born in the work of staff and the recommendations of committees, circulate among and are given specific substance by individual scholars, and go on to figure centrally in the decisions made at other meetings. At the same time, they and their consequences remain, in large part because of their very ordinariness, “analytically invisible” (Brenneis 2006: 42). My analysis of the social lives of such mundane forms has focused on several dimensions including the forms themselves as both texts and frame and the kinds of “doings with documents” (Harper 1998: 3) that recurrently take place within an institution and the ways in which documents both derive from and help constitute the work of that institution. These forms both require certain kinds of response and make others unlikely – or, at times, unthinkable. To borrow another term from Harper, this time from his research with Abigail Sellen (Sellen and Harper 2002: 16-18), these forms have very particular affordances: they enable and perhaps even require some activities and efface the possibility of others.

In its recent examination of peer review for assessing work in the Humanities and Social Sciences, the British Academy concludes that “(p)eer review has its critics, who allege that it is costly, time-consuming and biased against innovation. None of these criticisms is entirely without force, but the Working Group concluded that there were no better alternatives and that often the criticisms were directed at deficiencies of practice rather than the principle of peer review” (British Academy 2007: ix). Other institutions, for example the US National Institutes of Health in the 2007 review, also argue that peer review remains the strongest option, especially if practices are reshaped to encourage
more focused and consistent deliberation. And, as noted above, the US National Research Council sees combining deliberation with appropriate and multiple analytical strategies as optimal.

What, then, do the metrics that Universities UK wants to have replace peer review entail? And why do some consider it such an attractive alternative to peer review? While other quantitative data such as average years to degree and research dollars awarded often figure in such metrics, at their heart lies bibliometrics, “using counts of journal articles and their citations” (Evidence 2007). Such indicators measure raw productivity (without making judgments of quality) but also count instances of citation by other scholars, which is taken as a proxy for the paper’s quality. Significantly, only citations within two calendar years of an article’s publication are counted, a very brief window indeed. The Universities UK report argues that “(c)itations between papers are signals of intellectual relationships. They are a natural, indeed essential, part of the development of the knowledge corpus. They are therefore valuable as an external index about research because they are produced naturally as part of ‘what researchers do’ and because they are related naturally to ‘impact’ and ‘significance’” (Evidence 2007). A key metric is the journal impact factor, a score for individual journals “created in the 1960’s as a way to measure the value of journals by calculating the average number of citations per article over a specific period of time” (Adler 2008). These impact factors are frequently used for ranking journals; they also are frequently taken as a way of measuring the quality of individual scholars’ work (depending on where it is published). The data for these citation statistics are all drawn from the Thomson Scientific (formerly ISI) indexes that cover “8,700 of the most prestigious, high impact research journals in the world”
(Evidence 2007). While 8,700 may sound like a large number, it represents a very small proportion of world journals in the sciences, let alone beyond them.

The attractiveness of this approach, in the words of a recent critical report, lies in the fact that “(t)hose who argue for this simple objectivity believe that research is too important to rely on subjective judgments. They believe citation-based metrics bring clarity to the ranking process and eliminate ambiguities inherent in other forms of assessment. They believe that carefully chosen metrics allow us to compare all parts of the research enterprise – journals, papers, people, programs, and even entire disciplines – simply and effectively, without the use of subjective peer review” (Adler 2008).

Apparently simple, clear, systematically related and ideally suited for commensurability, bibliometric assessment is also routinized, does not require the investments of time inevitable in peer review, and is relatively inexpensive. And the assumed transparency of bibliometrics is also wholly compatible with contemporary managerial concerns for accountability and audit: they appear to promise fairness and openness.

Not surprisingly, such metrics have engendered considerable criticism; some of the strongest and most convincing arguments against reliance solely upon this approach have come from mathematicians, engineers, and other scientists. As part of my goal in this talk is to suggest relevant resources, I want to highlight three very recent – and very compelling – assessments of evaluative practices, including bibliometrics. The first is the 2007 report, A Strategy for Assessing Science (Feller 2007), the work of a committee organized by the US National Institute on Aging. This report notes that bibliometrics (and the broader field of scientometrics of which it is a part) “were developed originally for exploring the working of the scientific enterprise, that is, as descriptive and analytical
tools, not as evaluative or predictive ones” (Feller 2007). The report further argues that, however carefully current shortcomings in the approach might be addressed, “the evaluative meaning of bibliometric comparisons requires interpretation.”

A second major document is the September 2007 British Academy Report, Peer Review: the Challenges for the Humanities and Social Sciences. In contrast to the solely metrics-based approach proposed for the sciences, it has been announced that “research in the arts, humanities, social sciences, and mathematics will be assessed by a light-touch process, based on peer review but informed by metrics” (British Academy 2007). The report makes two recommendations specifically concerning bibliometrics. The first, more directly relevant recommendation is that “(c)are should be taken to ensure that any metrics employed reflect the distinctive nature of the humanities and social sciences research and do not have an adverse effect on the quality of the work that they are seeking to measure,” further suggesting that possible modifications of the Thomson Scientific indices be considered.

A final major report (released on June 18 of this year) was written by a committee representing the International Mathematical Union, the International Council for Industrial and Applied Mathematics, and the Institute of Mathematical Statistics (Adler 2008 ). Clearly prompted by the Universities UK policy, this exceptionally cogent paper provides a rigorous, well-documented and daunting critique of the proposed move to metrics. Some of the criticisms echo those in the British Academy report; the two year window for citation counting, for example, proves as inappropriate for mathematics as it does for the humanities. There are also detailed analyses of gaps and errors in the statistical claims being made and of the limitations of specific measures (Adler 2008 ).
But most interesting for me as an anthropological reader was what the authors considered a key question: “Those who promote citation statistics as the predominant measure of research quality do not answer the essential question: What do citations mean? They gather large amounts of data about citation counts, process the data in order to derive statistics, and then assert the resulting assessment process is “objective.” Yet it is the interpretation of the statistics that leads to assessment, and the interpretation relies on the meaning of citation, which is quite subjective” (Adler 2008). They provide a brief but bracing discussion of the sociology of citations, noting that authors cite others for many reasons beyond recognizing an intellectual debt. The authors argue that “citation-based statistics can play a role in the assessment of research, provided that they are used properly, interpreted with caution, and make up only part of the process” (Adler 2008).

In closing I want to highlight one dimension of bibliometrics that particularly struck me as a linguistic anthropologist. That is that Universities UK seems to be seeking a language for description, comparison, and evaluation that is, to borrow Ewald’s phrasing, “a language of precision and certainty, a language without puns, stylistic features or interference” (Ewald 1991), in short a language of pure reference. It is also a language without context, as it transcends the specifics that it claims to represent. And it is, I think, a language into which we should resist translation. As Stefan (Collini 2003) has suggested, it is critical that we recurrently argue that “intellectual activity can, for the most part, be judged but not measured.” Beyond this, however, we can work against the development of evaluative procedures that encourage the generation of too much information of a particular, decontextualized type, as they make “transferable skills an objective…reduc(ing) what makes a skill work, its embeddedness” (Strathern 1997: 14).
No field has wrestled with questions of the deeply contextualized nature of social, cultural, and scholarly practice than anthropology; it is a key aspect of our academic charisma. And no group is better suited than this one for deliberatively, imaginatively, and knowledgeably bring the strengths and singularities of our field to bear on the complexities of the “European convergence.”

References

Academy, The British

Adler, Robert, John Ewing, and Peter Taylor
2008 Citation statistics: A report from the International Mathematical Union: International Mathematical Union.

Bell, Robert

Biagioli, Mario
Brenneis, Don

Brenneis, Donald

Brenneis, Donald

Chubin, Daryl E., and Edward J Hackett

Clark, William
2006 Academic charisma and the origins of the research university. Chicago: University of Chicago Press.

Cole, Jonathan, and Stephen Cole

Collini, Stefan

ESRC

Evidence, Ltd.
2007 The use of bibliometrics to measure research quality in UK higher education institutions: Universities UK.

Ewald, Francois

Feller, Irwin

Feller, Irwin, and Paul C. Stern, eds.

Harnad, Stevan

2008 Validating research performance metrics against peer rankings. In Inter-research ethics in science and environmental politics.

Harper, Richard

Health, National Institutes of

Lamont, Michele, Gregoire Mallard, and Joshua Guetzkow

Mills, David

Nerad, Maresi
2008 Social Sciences PhDs - Five+ years out (A national survey of PhDs in six fields). Seattle: University of Washington.

Office, General Accounting

Pollick, Amy
Roy, Rustum  
1985  Funding scince: the real defects of peer review and an alternative to it.  
Science, Technology, and Human Values 10:73-81.

Sellen, Abigail J., and Richard Harper  

Strathern, Marilyn  
1997  From improvement to enhancement: an anthropological comment on the audit culture (Founder's Memorial Lecture, Girton College).