

Researchers as evaluators: tasks, tensions and politics

Liv Langfeldt · Svein Kyvik

Published online: 7 October 2010
© Springer Science+Business Media B.V. 2010

Abstract Researchers undertake a number of different research evaluation tasks, taking up a substantial part of their research time—estimated to about one work month per year for a professor. This paper addresses the various evaluator roles and tasks researchers take on, and the tensions they involve. How the research evaluator role may conflict with the researcher role and with societal expectations is discussed, as well as the intrinsic tensions in peer review; including expertise vs. impartiality, evaluators as neutral judges vs. exercise of power and influence, divergent peer assessments vs. the need for unanimous conclusions in peer panels, peer review vs. increase in quantitative indicators, and accountability to society vs. peer review as preserving the autonomy of science. The examination of these tensions provides insight in the political aspects of peer review, and a basis for discussing an agenda for future studies on the role of peer evaluators. Major future challenges for peer review concern how to meet demands for transparency and public accountability, and maintain academic autonomy.

Keywords Peer review · Scientific quality · Evaluation spiral · Academic autonomy · Public accountability

Introduction

Researchers have many different tasks and duties as evaluators of research—they hold different kinds of gatekeeper roles evaluating research undertaken by other researchers (Zuckerman and Merton 1971). These tasks are of fundamental importance to the scientific system, supporting good research to the detriment of poor research. As evaluators, academic staff provide or deny access to opportunities for fellow colleagues to do research, to publish research, and to get tenure or promotion—processes commonly known as *peer review*. The trouble is, according to Ziman (1994: 103), that to do this properly requires “large quantities

L. Langfeldt (✉) · S. Kyvik
NIFU STEP Norwegian Institute for Studies in Innovation, Research and Education,
Wergelandsveien 7, 0167 Oslo, Norway
e-mail: liv.langfeldt@nifustep.no

of the most valuable resource in science: the personal time of the most competent researchers”. Moreover, there is reason to believe that research evaluation has increased in importance and scope. Nowotny et al. (2003) state that the growing emphasis upon evaluation is one of the elements in the transformation of research from Mode 1 to Mode 2.

This paper addresses the various evaluation roles and tasks researchers take on, and the tensions they cause. How much of researchers’ time is spent on evaluation, what is the scholarly and political aspects and importance of the evaluation roles, and how may the evaluation role conflict with the researcher role? A number of interlinked tensions are discussed:

1. The time conflict: time spent on evaluations implies less time for research
2. Expertise vs. impartiality: the closer peer expertise, the more likeliness for bias
3. Dual expectations: evaluators as neutral judges vs. exercise of power and influence
4. Divergent peer assessments vs. the need for unanimous conclusions in peer panels
5. Peer discretion vs. increase in quantitative indicators
6. Dual purposes: evaluations ensuring accountability to society vs. peer review as preserving the autonomy of science

The next section gives an overview of the key features and functions of the various evaluation tasks, while “*Tensions*” is discussed in the following section. In the final section the future of the evaluation role is discussed, and an agenda for further research is suggested. Generally there is lack of data and studies on most aspects of the evaluator role, and in particular on how evaluators are selected.

A variety of evaluation roles and tasks

There are at least nine separate functions of the researcher evaluator role: (a) assessment of doctoral dissertations, (b) selection of new staff and promotion, (c) distribution of resources for research, (d) assessment of manuscripts submitted for publication, (e) reviews of books or the state of the art in a field (book reviews and review articles), (f) assessment of candidates for scientific awards, (g) evaluation of research organisations, (h) assessment of research as input to policy-making, and (i) assessment of future research strategies and priorities. These different sub-roles might be named the examiner role, the staff selector role, the grant distributor role, the referee and editor role, the reviewer role, the prize awarder role, the organisational consultant role, the policy advisor role, and the foresight role. While the first three roles are concerned with the input of human and financial resources to the research system, the five next roles are concerned with the output of research, and the two latter roles with future strategies. The evaluator role is partly performed as one of the duties included in the university position (assessments of doctoral candidates and applicants for university positions), partly as paid extra-work (evaluation work for other institutions), and partly as unpaid work for the disciplinary community (referee of journal articles).

The expectations that academic staff take part in these tasks are above all embedded in norms, seldom in formal rules or regulations. Moreover, they deal with the definition of good and valuable research and are important for the distribution of honour and credibility in the research community. In other words, the evaluator role is voluntary and an opportunity to exercise academic power.

The studies and literature on peer review are extensive, but they deal foremost with review of proposals to funding agencies and manuscripts submitted to scientific journals,

whereas the literature on the many other evaluator roles is more limited. Below we draw on available literature and general insights from the sociology of science, and present the key features of all nine roles.

The examiner: An important aspect of the evaluator role is to examine doctoral dissertations. In most countries, doctoral students are responsible for a substantial share of the total research activity undertaken in universities, and dissertations constitute an important part of the research output. Through the examination process, PhD candidates are certified as researchers and promising research talents are identified. This makes examination of doctoral dissertations one of the most important gatekeeper roles in academia, securing the qualifications of the future researchers. For the involved examiners, dissertation examination is also an important way to keep updated on new research and promising new researchers. An analysis of dissertation examinations suggests that the examiner take a role similar to the reader of any new piece of writing (Johnston 1997: 340). Even if the approach to the text is not too different from that of a normal reader, dissertation examination is one of the more demanding evaluation tasks for researchers. Dissertations are generally long, and a thorough review is expected, often involving both written and oral presentations. Consequently, dissertation examination may be a particularly time-consuming part of the evaluator role.

The staff selector: The assessment of applicants to academic positions is another important task included in the evaluator role. Through the consideration of the promise and limitations of competing aspirants to vacant positions, the members of assessment committees shall select and recommend the best suited person for the position. In addition, the selector role includes the participation in peer-review committees set up to assess applications for promotion to higher rank and tenure. In this role, the academic staff member has to consider whether the applicant lives up to a minimum standard required for associate or full professorship in the discipline and at the particular institution. Within this evaluator role, researchers may impact the future scholarly profile of departments, as well as the staff gender balance (AAUW 2004). These are processes that may cause conflict (Hearn and Anderson 2002).

The grant distributor: A third important sub-role is to take part in committees set up to distribute research grants, or to act as an individual referee of applications for research grants. This is a task of vital importance to the scientific system, because resources are usually in short supply and the applications with the most promising potential should be supported. As the share of resources for research is increasingly distributed through research councils, foundations, and other research supporting organisations where peer review constitute the basis for allocations, the importance of this sub-role has increased. Being a grant distributor implies participating in a zero-sum-game defining the research agenda. The other side of this is the increasing rejection rates of funding agencies, which frustrates both the applicants and the review committees. As a consequence, grant peer review is a disputed system and there have long been suggestions to replace it, for reasons which go beyond its consumption of valuable research time (see e.g. Horrobin 1982; Roy 1984). It is difficult, and not always meaningful, to assess research that is not yet performed. Moreover, biased review may have extensive effects on what kind of research is funded, and what is not. As far as only certain schools of thought are represented on the review panels, conclusions may suffer from cognitive biases (Travis and Collins 1991). Fair negotiation in multidisciplinary review panels is a similar challenge (Lamont 2009). In sum, the review of project applications is notably one of the most contested forms of peer review. Defining which groups and what topics are to be funded, is the implementation of research policy, and the grant distributor is the evaluation task that may most easily be

substituted by other allocation measures, such as direct allocation to institutions based on past performance (see “[The future evaluator role: a research agenda](#)”).

The referee and editor: The sub-role as referee and editor of manuscripts submitted for publication is another important part of the academic role. While referees have the job of assessing the quality and relevance of manuscripts and recommending whether they should be accepted for publication or rejected, the editor makes the final decision based on the advice of a number of referees. Although these are two different functions, the work of referees and editors are both concerned with the selection of manuscripts for publication as books and articles, and consequently a distinct sub-role of the research evaluator (Weller 2001). A related and similar, but less formalised role is the screening of abstracts and papers to scholarly conferences. There is an extensive literature on journal peer review, as well as separate conferences (e.g. Ceci and Peters 1982; Campanario 1998a, b; Speck 1993; Weller 2001). Reviewer bias is a central topic in the discussion.

The reviewer: This function takes the form of either book-reviews or review articles of the state of the art in a subfield, etc. Through exercising this sub-role, the reviewer allocates rewards or sanctions to the scientists concerned through acclaim, criticism or neglect of their articles, books and other scholarly contributions. In most cases, writing reviews can be an integrated part of research—reading and assessing literature as basis for one’s own research.

The prize awarder: Another evaluation function concerns the conferment of scientific awards, prizes and honours. This is not an important sub-role in the sense that many staff members serve on award granting committees. But such duties are normally regarded as prestigious and may take a lot of time for those who are appointed to such committees. Moreover, such evaluation work may give the opportunity to exercise power and take part in the high level politics of science, as illustrated by studies of the work behind awarding the Nobel Prizes in Science (Friedman 2001).

The evaluator of research organisations: The sub-role as evaluator of research organisations (groups, departments, programmes, institutes, and even universities) is not new to academic staff, but has increased substantially in importance over the last two decades as a response to the needs of ‘the evaluative state’ (Neave 1998). Demands for better quality of university research, greater relevance of research to societal needs, and improved efficiency (value for money) have led to an evaluation wave focussing on the organisational level. This sub-role is normally not limited to the assessment of the quality of research of specific units (although such a restricted purpose is not unusual), but also includes considerations and recommendations on how the research could be improved, how the unit should be organised, and how internal and external relations could be developed. There is great diversity between countries in the purpose of these kinds of evaluations, and how they are performed (Hansen 2009).

The policy advisor: Another important task is assessment of research to be used as input to policy-making and regulation. Extensive review work is performed by scientists employed by government agencies within areas such as health and environment. Standing or ad hoc scientific committees serving government agencies also perform this kind of review. Moreover, researchers serving on international bodies as the Intergovernmental Panel on Climate Change (IPCC) belong to this category. The combination of peer review and public policy, however, entails a number of challenges, e.g. concerning divergent assessments and potential biases (Jasanoff 1990: 79 ff, see Tension 4 below).

The foresight viewer: In the 1980s and 1990s, research funding agencies were encouraged and directed by governments to become more strategic in their funding policy through the creation of large-scale R&D programmes aimed at supporting promising new areas of

research. Irvine and Martin (1984) used the term ‘foresight activities’ to describe the techniques, mechanisms and procedures for attempting to identify areas of basic research of strategic potential. In these attempts, academic staff members have been engaged in committees and ‘think-thanks’ to suggest areas for scientific investigation which most likely will provide the knowledge-base for the technologies and industries of tomorrow.

Adding together all the tasks, researchers seem constantly involved in different forms of formal evaluation roles, both as evaluators and evaluatees. An *evaluation spiral* subjecting the same research to repeatedly peer review emerges from the overview above. A project may go through its first evaluation (1) when applying for (one of more) grants, and subsequently when (2) conference abstracts, (3) journal papers, and possibly (4) dissertations and (5) book manuscripts from the project are submitted. Moreover, (6) publications from the projects are the object for review when the authors apply for academic positions, (7) are nominated to awards and prizes, and when (8) their departments or (9) the programmes that funded the project are evaluated. There is also the possibility that (10) one or more reviews of books from the project are presented in scientific journals, and/or (11) that publications from the project are re-evaluated in review articles. If the project entails results of political significance or interesting openings for new research, it may also form part of (12) policy-making processes and foresight studies. Finally, the outputs of the project return to stage one, as researchers’ past achievements are assessed when new grant proposals are submitted—continuing the evaluation spiral. Adding to this, evaluation is an integrated aspect of research. To perform research one needs to constantly assess research methods and results. Hence, in addition to all the formal peer review, there is substantial informal review.

Tensions

There are multitudes of ways in which peer review affects research and the research community, and consequently different ways the evaluator role may conflict with the researcher role.

Tension 1: Time for research vs. time for evaluation

The most obvious conflict between the evaluator role and the researcher role is the *time conflict*. The more time researchers spend on evaluation, the less time there is for research.

With each piece of research being evaluated so many times, how much of a researcher’s time is likely to be spent on formal evaluation tasks? Not much is known about the extent of evaluation tasks. In a survey undertaken in 2001 among all tenured academic staff in Norwegian universities, four of the nine functions of the evaluator role were surveyed: examination of doctoral dissertations; assessment of applicants for vacant positions and promotion to higher rank; referee-work for journals, etc.; and assessment of research organisations. The sub-role that involved most people was referee-work for journals, etc. In the course of a year, two out of three academic staff members engaged in this task. The examiner role and the staff selector role each involved about 40% of the staff members, while less than 20% took part in one or more evaluations of research organisations. In total, about 80% of the academic staff took part in at least one of these four evaluation roles, and those who were engaged used on average close to 17 days on these tasks. The evaluator role took more time for full professors (19 days) than for associate professors (12 days) and assistant professors (8 days). Because the survey did not include the other sub-roles,

the total number of days spent on the evaluator role is higher than the observed figures. A substantial number of university staff members use time on assessments of applications for research grants in the various types of organisations that allocate money for research. Data on the participation in the distributor role were compiled in a similar survey undertaken in 1992. During their career, 40% of the staff had been used as a referee by a research council, and 28% had been a member of a research council, or a member of panels, committees, etc. under a research council. Adding a few days of grant review to the numbers above, an approximate time estimate might be a total average of 20 days per year for the approximately 80% of staff members performing evaluations, and 25 days for full professors (Kyvik and Langfeldt 2004). This is equivalent to four and five normal working weeks. The evaluator role thus takes a lot of time, in particular for full professors.

The question whether the evaluator role has changed over time can only be partly answered by the survey data. For the whole period, only information on examination of doctoral dissertations and staff selection were compiled. The percentage of staff who took part in this work increased from 46 to 58% from 1991 to 2000. The average number of days used for these two sub-roles among those who were involved did not change over time, indicating that the total time used for these purposes by academic staff increased during the last decade. For the other functions, there is reason to believe that the overall pressure on staff has increased. Referee work is hardly being reduced, and an increasing share of resources for university research is distributed through various external research funding organisations, normally involving a substantial number of academic staff. Over the last two decades, the grant distributor role has obviously changed from assessments of individual applications for support, to assessments of applications from research milieus for participation in large research programmes, as well as the establishment of various kinds of temporary centres and networks of excellence. But the function that most likely takes more time than before is the evaluation of research organisations as a consequence of the introduction of new public management reforms with more weight on ex post assessments. A more cosmopolitan profile of the evaluator role is another change revealed by the survey data. In 2000, 22% of academic staff at Norwegian universities took part in evaluation work abroad in contrast to 9% in 1991. In sum, there is reason to believe that the evaluator role has increased in importance and time dedication.

The time conflict is likely to entail some stratification of the evaluator tasks. Some of the most frequently approached researchers cannot possibly take on all evaluation tasks they are invited to. They will have to be selective, limiting their efforts to what they perceive as the most important tasks. The likely result is that the highest ranking academics handle the most prestigious and power-performing evaluation tasks (Cole 1983: 138), leaving the less prestigious and less power-performing tasks to less high ranking and more junior researchers.

Tension 2: Peer expertise vs. impartiality

More serious conflicts between the evaluator role and the researcher role relate to *conflicts of interest*. The most competent peers to evaluate a piece of research are often peers close to those who perform it (Chubin and Hackett 1990: 80), but close peers are disqualified by conflicts of interest regulations. Whether or not they have any identifiable vested interests in the outcome of the evaluation, conflicts of interest regulations may formally disqualify them.

In some cases conflicts of interest is an argument for precautions to ensure the autonomy of science, while in other cases conflicts of interest can be an argument for

alternatives to peer review; for making academia renounce some of its autonomy to avoid setting “the fox to mind the geese”. An example of the first kind is *the interests of pre-publication referees and review writers*. Setting a scientist with commercial interests in pharmaceuticals to review papers on the effects of pharmaceuticals, or to write a review article on it, would endanger the autonomy and credibility of science and call for only ‘pure’ academics to perform such tasks. On the other hand, taking part in the review of proposals to a programme may disqualify the evaluator and the evaluator’s group from applying for grants from the programme. Such disqualifications may conflict with a scholar’s research interests and make researchers more cautious about taking on the review of grant proposals.

Dealing with these tensions may involve restricting the self-governance of the individual research fields, or also limiting the use of peer review for allocating research resources. An example of the first case is the use of foreign mail reviewers and broad panels with no true peers in the research field of the programme. In this way domestic researchers are not in a position to allocate research resources within their own research field (restricting self-governance). An example of the latter case is directing more resources through channels not requiring peer review. This may include allocating public funds for fundamental research directly to the universities based on performance indicators or administrative decision, as well as commissioned research and tender competitions.

Tension 3: Dual expectations: neutral judge vs. promoting research interests

There are also tensions related to dual expectations to the neutrality of evaluators. On the one hand, evaluators are expected to be neutral judges performing impartial and thorough review. On the other hand, evaluators expect to be able to impact what is defined as good research and have a say in how important resources in their field are allocated. These dual expectations leave a negotiable room for the requirements to, and meaning of, “a neutral judge”. “Neutral judge” obviously excludes any personal interests, as formalised by conflicts of interest regulations. But what about scholarly biases, that is, biases relating to the evaluators’ research field, research interests or “school of thought”? In some regards, the notion of a scholarly neutral research evaluator is meaningless. Researchers have different scholarly backgrounds and viewpoints concerning, e.g. specific methods and theories, and the differences in judgments resulting from their backgrounds and viewpoints are likely to be seen as legitimate differences, not as bias (Langfeldt 2002: 67–69). In other words, evaluators are not supposed to be scholarly neutral, they are supposed to provide assessments based on their scholarly discretion. The tensions arise because there are no clear limits to scholarly discretion, that is, no clear borderline between what evaluators may try to impact and what they may not. For example, it may not be easy to distinguish the promotion of appropriate research methods and perspectives, from assessments which support the evaluator’s own research interests and standing.

Tensions between expectations for being a neutral judge and for impacting outcomes may be particularly disturbing when there are no clear standards for selecting evaluators, and when the processes lack transparency. Part of the literature on dissertation examination deals with the lack of national standards for PhD examination (Tinkler and Jackson 2000; Morley et al. 2002) and provides examples of practices that allow for distortions and disruptions. For instance, lack of standards and transparency on how examiners are selected may entail detrimental power relations between the supervisors and the examiners (Morley et al. 2002: 270–271).

Tension 4: Divergent assessments vs. unanimous conclusions

The evaluator role's potential for exerting power derives from scholarly discretion and peer disagreement. There are often divergent scholarly opinions, and participating in peer review provides an opportunity to contribute to the definition of good research, as well as defining the research agenda. Different scholars have divergent assessments and priorities, and peer review processes aim to reach conclusions on the allocations of scarce resources and honour.

There are two major ways of handling such disagreements; through face-to-face discussions, bargaining and compromises between peer reviewers, or by non-experts making decisions based on a number of individual peer assessments. Editorial decision on the publishing of submitted manuscripts to journals is an example of the latter. Editors normally base their decisions on at least two individual expert reviews. If the reviews come to clearly different conclusions, the editor will acquire more reviews. An example of the first way of handling disagreement is decision-making in grant review panels. Whether they base decisions on panel members' assessments only or also on review reports from external experts, grant review panels are supposed to reach unanimous conclusions on the grant applications. Studies indicate that important mechanisms of reaching agreements in these panels are maintaining collegiality and avoiding conflict by deferring to expertise—that is, respecting the assessments of the panel member with the most established proof of competence on the application in question (Lamont 2009: 117)—and that the way the processes are organised may affect the outcome (Langfeldt 2001).

In the review of journal manuscripts, grant applications and most other peer review that allocate scarce resources in the scientific community, divergent assessments are unproblematic and even considered an important part of the dynamics of science. When selecting the peer reviewers, care is taken to cover a broad set of expertise relevant to the research under review, and there are established procedures for reaching a decision. On the other hand, when assessing research to be used in public policy, divergent assessments often cause problems. Solid and non-disputed conclusions may be a requirement for policy decisions as well as for policy implementation. Cases of risk assessment and other issues of national (or international) concern involve different political objectives and interests groups, and the parliament, the media and various NGOs may act as watchdogs. Stakes are high and review processes that leave room for bias or incidental results—results depending on who was selected for a specific review job—are not tolerated. While credibility is more crucial, the potential for bias is also more pronounced in reviews relating to public regulation and policy than in ordinary peer review, as the reviewers' research perspectives, etc. may influence how risks are evaluated (Jasanoff 1990: 76–83). In this way the tasks and challenges of researchers taking on this evaluator role are more delicate than those of ordinary peer review.

Tension 5: Peer discretion vs. quantitative indicators

Along with peer review and the evaluation spiral described above, there is also an increasing use of quantitative indicators of research output. Quantitative indicators are compelling because they are cheap and offer a simple response to calls for greater accountability (Ziman 1994: 103–105). Even when the key quantitative indicators are aggregated conclusions of prior peer assessments (that is, bibliometric indicators), there are tensions between quantitative indicators and peer review, and they can provide different conclusions. Studies comparing the outcome of peer review and bibliometric indicators

have found some correlation between the two, but in many cases the correlation is far lower than what one might expect (Aksnes and Taxt 2004; Bornman and Daniel 2008; van den Besselaar and Leydesdorf 2009). Even though both are built ultimately on peer assessments (directly or indirectly), peer review and bibliometrics are based on very different logics. Peer review involves subtle and tacit judgements and depends on intimate craft knowledge of the work under review (Ravetz 1971: 274). One needs to be up to date on the research frontier and cannot rely on any fixed or clear-cut standards for the assessments. Bibliometrics are, on the other hand, based on counting past peer assessments (number of papers accepted in indexed journals) and peer attention (in most cases restricted to citations in indexed journals). Bibliometric indicators can give important information on a researcher's or research unit's track record, networks and collaboration patterns, but are disputed as evidence of scientific quality. In some cases, peer review and bibliometrics are combined, e.g. in the evaluation of research programmes, and may help to reduce the time researchers spend on evaluation. In other cases quantitative indicators replace peer review, as when allocation of research funds are based on past performance indicators (number of publications, citations, amount of research grants, etc.) instead of review of project proposals. More precisely, when funding authorities are looking for ways to increase accountability and provide productivity incentives in the higher education sector, they replace fixed block grants to universities by funding based on quantitative performance indicators (Sörlin 2007), and not by funding based on peer review.

Tension 6: Autonomy vs. accountability

While peer review is an important control mechanism in the scientific community, it may obstruct accountability to society. Peer review serves as a mechanism of “professional self-regulation that affords scientists a degree of autonomy from scrutiny by the public at large” (Hackett 1997: 57). Scientific quality is one of the least publicly politicised aspects of science. The scientific community commands full autonomy in defining and assessing the quality of scientific research. A key element in academic autonomy is to define only peers as competent to perform quality assessments, including identifying the research frontier, what is of scholarly value and interest, what are adequate methods and theories, and who is competent to perform research. In times when scientific autonomy is contested (Henkel 2007) and when there is little common understanding or overlap of public and academic research interests, demands for public definition of the research agenda are likely to fortify the quest for academic autonomy. Scientists invoke peer review in their own defence (Chubin and Hackett 1990: 5). One example of this is seen in the evaluations of research programmes and fields when peer evaluators comment on imbalances between user oriented and fundamental research, and promote a hands-off policy by recommending more resources to fundamental researcher initiated research (Langfeldt 2002).

The scientific community's ability to protect its autonomy is generally strong. The need of peer expertise to assess scientific quality is obvious, and academics often succeed in defining ‘quality’ as ‘scientific quality’ (primary or only). There is most often a lack of an experienced and committed public to engage in the discussion of the quality of research, and the defined room for non-peer participants in the evaluation of research is clearly delimited. There are, however, obvious differences between the evaluator roles regarding academic autonomy tensions. For instance, whereas the examiner role (assessment of PhD dissertations) entails no autonomy tensions and examiners cannot be substituted by non-peers, the distributor role (assessment of grant applications) is more contested as grant review may be substituted by direct allocations based on e.g. productivity, instead of quality measures.

Moreover, the relevance of non-peer assessments of grant applications may add academic autonomy tensions. Research grant decisions often include other concerns besides scientific quality, e.g. fitness for purpose/to the programme, and the thematic, geographical, institutional and gender allocation of funds, as well as potential societal impacts and ethical considerations (Langfeldt 2001: 827–829). In general, assessments of such concerns do not require strict peer competence. Neither is there an obvious answer to the question of who can best assess the impact of research and the broader societal considerations. Peers may have expertise in identifying potential use of fundamental research, but less so in estimating potential societal benefits or risks, or in measuring the various kinds of impacts of completed research. There is an inherent uncertainty in assessing research, particularly *ex ante* as for grant review, and both scientific and societal merits are uncertain. Peer evaluators may, however, argue that societal merits are more uncertain than scientific merits; that societal effects depend on scientific success as well as several other factors, and consequently prioritise scientific over societal merits. Moreover, any way of organising grant review that opens for non-experts overruling experts, may be seen as illegitimate and be deliberately opposed. For example, in a two stage procedure where peers first provide their judgements and another body including non-peers subsequently makes the final priorities, peers may exclude applications with high societal importance—defining them as low quality applications—before any concerns related to societal importance is considered at stage two.

In conclusion, silently protecting scientific autonomy is perceived as part of the evaluation task in some contexts. Even when peer review is defined as part of the system ensuring accountability for public expenditures, public accountability and ensuring that public funding serves public needs are not the priorities of the peer evaluator (Van der Meulen 1998: 405 ff).

The future evaluator role: a research agenda

The role as researcher evaluator is changing. Over the last two decades it has become increasingly international. Cross-border peer review serves to enhance review quality as well as to avoid conflicts of interest. International peer review is, in itself, also an important part of the internationalisation of research. Open peer review is another trend possibly contributing to more egalitarian review processes. Some journals have introduced more transparent and inclusive on-line peer review processes, including open invitations to comment on submitted manuscripts (Pöschl and Koop 2008).

In addition to the general changes linked to globalisation and transparency facilitated by the Internet, there are some more specific indications of change linked to the tensions discussed above, in particular contested scholarly autonomy, the time spent on evaluation tasks, and increased use of quantitative indicators. Below we discuss an agenda for studies on the research evaluator role addressing these challenges.

Selection of evaluators, time constraints and the stratification of science

The outcome of peer review depends on who is chosen to do the assessments. Yet we know little about how evaluators are selected for the review of manuscripts, grant applications, applicants to academic positions, etc. Research on such selection processes is scant. Moreover, there is little research on time constraints and availability of appropriate expertise. A better understanding of these issues is important for the discussion of the

prospects of the researcher evaluator role and its role in the stratification of the scientific community. Under this heading key research questions include: Who are the most wanted evaluators, for what reasons are they selected to various evaluator roles, and what do the scholars perceive to be the most attractive evaluation tasks—tasks that they are willing to spend time on? Moreover, there is a need for updated studies on how much time is spent on the various evaluation tasks.

Selection processes and time constraints may impact on the work division and stratification of the scientific community in various ways, including the global work division. Researchers are limited in the proportion of time they are willing to spend on evaluations. When more tasks are added to the evaluation spiral, the most senior and frequently enquired scholars will most likely need to prioritise more strictly the evaluation tasks they take on, and allow more junior scholars and/or scholars from a broader range of countries, to be involved in evaluation. In other words, time constraints may reduce the stratification of science. Introduction of open peer review (self-selection of reviewers) may add to this and prevent gatekeeper roles from being held by a small group of distinguished seniors. On the other hand, as far as the most senior/high ranking researchers can choose between evaluation tasks of perceived different importance, they may retain the most politically important ones and leave the more trivial ones for less prominent members of the research community. Research questions here include the limits to how much time a researcher is willing to use for evaluating others' works and how they select the evaluation tasks they agree to take on. For example, when academic autonomy is contested, scholars may be willing to spend more time on evaluation processes to ensure that scholarly quality are not overrun by other concerns.

Autonomy, indicators and non-peers

Other important issues are challenges to scholarly autonomy and the use of quantitative indicators and evaluations not involving peer competence. Low success rates in funding agencies as well as concerns with bias and conflicts of interest in peer review, may bring about measures to limit the researcher time spent on peer review and writing applications, and put more emphasis on quantitative indicators. More concern about societal relevance and accountability of research may result in evaluations by non-peers.

The effect of quantitative indicators and performance based funding on the researcher evaluator role is not evident. On the one hand, increased use of quantitative indicators in the allocation of research funds may reduce the role of peer evaluation—as far as review of project proposals are replaced by performance indicators. On the other hand, performance based funding increases the aggregated importance of peer review. Key indicators in performance based funding are number of peer reviewed publications and success in grant competitions based on peer review.

Quantitative indicators give rise to two set of research questions concerning researchers as evaluators: Firstly, to what extent are the evaluators aware of this aggregated importance of their review work, to what extent does it influence their willingness to take on evaluation tasks, and to what extent does it influence conflicts of interest consideration and use of foreign reviewers? The second set of questions relate to scholars' opinions and preferences concerning indicators and peer review. Do scholars prefer research funding based on quantitative indicators and incentive systems, or peer review of grant proposals? Both peer review of grant proposals and quantitative indicators are frequently criticised by scholars. Peer review is criticised for being conservative, discriminating interdisciplinary and original research, as well as time consuming. As performance based funding is based on the

aggregated outcome of such peer review, it risks reproducing the weaknesses of peer review. Performance based funding is moreover criticised for producing dysfunctional incentives (e.g. priority to quantity over quality). It is still not obvious what is perceived as the most fair, the least conservative and the most efficient way of allocating research grants.

Whereas the introduction of quantitative indicators imply a competitive regime based on (presumably) predictable criteria, the introduction of non-peers in the evaluation processes imply clearer challenges for academic autonomy and criteria. There are several indications that review of project applications are redefined towards more emphasis on public accountability and less on scholarly autonomy. Some funding agencies involve non-peers and broader societal criteria in their review of applications. There are also scholars who challenge traditional peer review and suggest more transparency and a broader set of explicit criteria (Scott 2007), thereby challenging the academic evaluation monopoly. Further, there are cases where broader societal criteria are included in the review of fundamental research and open calls for applications. It is notable that the general review guidelines of the major US funding agency, the National Science Foundation, ask evaluators to assess “the broader impacts of the proposed activity”.¹

There are many arguments against full academic self-governance in allocating research resources; arguments for more emphasis on including external concerns and involving non-peers when setting research priorities. It is held to be a public task to ensure that public funds for research serve public needs or at least some general interests. Moreover, transparency and openness to outside criticisms and responsiveness to public concerns, are prerequisites for the legitimacy of large public research expenditures. We have also seen some demands for citizen participation (democratisation of science), more emphasis on use and impacts, more accountability, and challenges related to complex and controversial issues characterised by a high level of uncertainty (Scott 2007; Jasanoff 2003). Including broad criteria and lay people in assessments may better ensure that research serves overall public needs as well as needs of deprived groups and avoids increased skewness of knowledge (Woodhouse and Sarewitz 2007). Non-academic competences and criteria may reduce insider bias (Martin 2000) and better ensure that a broader set of concerns are taken into consideration (social, economic, environmental, health concerns, etc.) in the allocation of research resources. These are all demands and concerns that underline the politics of peer review, in particular grant peer review, and point towards future challenges in maintaining—or redefining—academic autonomy.

As reviewing grant applications is a zero-sum-game defining the research focus within a programme or research field, this role is likely to be perceived important for protecting scholarly autonomy. How the introduction of non-peers and broader relevance criteria affect the review work and responsiveness to public concerns needs in-depth study. Moreover, both the evaluators’ and the research funding authorities’ conceptions of academic autonomy should be studied.

References

- AAUW. (2004). *Tenure denied. Cases of sex discrimination in academia*. Washington, DC: American Association of University Women Education Foundation and American Association of University Women Legal Advocacy Fund.

¹ <http://www.nsf.gov/pubs/gpg/broaderimpacts.pdf>.

- Aksnes, D. W., & Taxt, R. E. (2004). Peer review and bibliometric indicators: A comparative study at a Norwegian university. *Research Evaluation*, 13(1), 33–41.
- Bornman, L., & Daniel, H. D. (2008). Reliability, fairness, and predictive validity of the peer review process for the selection of research fellowship recipients of the Boehringer Ingelheim fonds. In B. M. Kehm (Ed.), *Hochschule im Wandel. Die Universität als Forschungsgegenstand. Festschrift für Ulrich Teichler* (pp. 365–376). Frankfurt am Main: Campus.
- Campanario, J. M. (1998a). Peer review for journals as it stands today—part 1. *Science Communication*, 19(3), 181–211.
- Campanario, J. M. (1998b). Peer review for journals as it stands today—part 2. *Science Communication*, 19(4), 277–306.
- Ceci, S. J., & Peters, D. P. (1982). Peer review: A study of reliability. *Change*, 14(6), 44–48.
- Chubin, D. E., & Hackett, E. J. (1990). *Peerless science. Peer review and US science policy*. New York: State University of New York Press.
- Cole, S. (1983). The hierarchy of the sciences? *American Journal of Sociology*, 89(1), 111–139.
- Friedman, R. M. (2001). *The politics of excellence. Behind the nobel prize in science*. New York: Times Books.
- Hackett, E. J. (1997). Peer review in science and science policy. In M. S. Frankel & J. Cave (Eds.), *Evaluating science and scientists* (pp. 51–60). Budapest: Central European University Press.
- Hansen, H. F. (2009). *Research evaluation: Methods, practice, and experience*. Copenhagen: Danish Agency for Science, Technology and Innovation. Research: Analysis and Evaluation.
- Hearn, J. C., & Anderson, M. S. (2002). Conflict in academic departments: An analysis of disputes over faculty promotion and tenure. *Research in Higher Education*, 43(5), 503–529.
- Henkel, M. (2007). Can academic autonomy survive in the knowledge society? A perspective from Britain. *Higher Education Research & Development*, 26(1), 87–99.
- Horrobin, D. F. (1982). Peer review: A philosophically faulty concept which is proving disastrous for science. *Behavioral and Brain Sciences*, 5(2), 217–218.
- Irvine, J., & Martin, B. (1984). *Foresight in science: Picking the winners*. London: Pinter.
- Jasanoff, S. (1990). *The fifth branch, science advisers as policymakers*. Cambridge, MA: Harvard University press.
- Jasanoff, S. (2003). Technologies of humility: Citizen participation in governing science. *Minerva*, 41(3), 223–244.
- Johnston, S. (1997). Examining the examiners: An analysis of examiners' reports on doctoral theses. *Studies in Higher Education*, 22(3), 333–347.
- Kyvik, S., & Langfeldt, L. (2004). Tidkrevende bedømmelsesarbeid. *Forskningspolitikk*, 27(2), 15.
- Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Cambridge, MA: Harvard University Press.
- Langfeldt, L. (2001). The decision-making constraints and processes of grant peer review, and their effects on the review outcome. *Social Studies of Science*, 31(6), 820–841.
- Langfeldt, L. (2002). *Decision-making in expert panels evaluating research. Constraints, processes and bias*. Oslo: Unipub. Accessed September 30, 2010, from <http://www.statsvitenskap.uio.no/fag/polit/disputas/fulltxt/langfeldt.pdf>.
- Martin, B. (2000). Research grants: Problems and options. *Australian Universities' Review*, 43(2), 17–22. Accessed September 23, 2010, from <http://www.bmartin.cc/pubs/00aur.html>.
- Morley, L., Leonard, D., & David, M. (2002). Variations in vivas: Quality and equality in British PhD assessments. *Studies in Higher Education*, 27(3), 263–273.
- Neave, G. (1998). The evaluative state reconsidered. *European Journal of Education*, 33(3), 265–284.
- Nowotny, H., Scott, P., & Gibbons, M. (2003). 'Mode 2' revisited: The new production of knowledge—Introduction. *Minerva*, 41(3), 179–194.
- Pöschl, U., & Koop, T. (2008). Interactive open access publishing and collaborative peer review for improved scientific communication and quality assurance. *Information Services & Use*, 28(2), 105–107.
- Ravetz, J. R. (1971). *Scientific knowledge and its social problems*. Oxford: Clarendon Press.
- Roy, R. (1984). Alternatives to review by peers: A contribution to the theory of scientific choice. *Minerva*, 22(3–4), 316–328.
- Scott, A. (2007). Peer review and the relevance of science. *Futures*, 39(7), 827–845.
- Sörlin, S. (2007). Funding diversity: Performance-based funding regimes as drivers of differentiation in higher education systems. *Higher Education Policy*, 20(4), 413–440.
- Speck, B. W. (1993). *Publication peer review. An annotated bibliography*. Westport/London: Greenwood Press.

- Tinkler, P., & Jackson, C. (2000). Examining the doctorate: Institutional policy and the PhD examination. *Studies in Higher Education*, 25(2), 167–180.
- Travis, G. D. L., & Collins, H. M. (1991). New light on old boys: Cognitive and institutional particularism in the peer review system. *Science, Technology & Human Values*, 16(3), 322–341.
- van den Besselaar, P., & Leydesdorf, L. (2009). Past performance, peer review and project selection: A case study in the social and behavioral sciences. *Research Evaluation*, 18(4), 272–288.
- Van der Meulen, B. (1998). Science policies as principal-agent games. Institutionalization and path dependency in the relation between government and science. *Research Policy*, 27(4), 397–414.
- Weller, A. C. (2001). *Editorial peer review. Its strengths and weaknesses*. Medford, NJ: Information Today (ASIS&T Monograph Series).
- Woodhouse, E., & Sarewitz, D. (2007). Science policies for reducing societal inequities. *Science and Public Policy*, 34(3), 139–150.
- Ziman, J. M. (1994). *Prometheus bound: Science in a dynamic steady state*. Cambridge: Cambridge University Press.
- Zuckerman, H., & Merton, R. K. (1971). Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva*, 9(1), 66–100.