2020

WHITE PAPER FOR SCIENCE POLICY ANNUAL MEETING

RESEARCH QUALITY

- DO WE SELECT THE BEST PROJECT PROPOSALS, SCIENTIFIC STAFF AND STRATEGIC INITIATIVES?
CONTENTS

Introduction P. 2

Recommendations P. 6

1 Peer review in the awarding of competitive research funds is under pressure P. 8

2 Assessment of research quality in scientific staff hiring P. 22

3 Strategic initiatives and breakthroughs P. 30

References P. 38

Further reading and notes P. 40
How do we identify research excellence? Can research quality be credibly assessed by non-experts, e.g. via various quantitative measurements? It is important to know whether research quality can be assessed in a reasonably objective manner, and if so how, since the answer is of critical importance for the decisions taken every single day in the world of research. We have therefore chosen ‘research quality’ as the topic of the Royal Academy of Sciences and Letters’ White Paper and 2020 Science Policy Annual Meeting. Our selection of the ‘research quality’ issue reflects the fact that it is not only timely and a topic of constant discussion among active researchers. It also reflects the discussion raised in a number of current Danish and foreign publications dealing with this and closely related issues.¹

Research quality is assessed every day around the world, and these kinds of assessments have taken place as long as there has been research. In this situation, the question is not so much whether research quality can be assessed at all, but rather, how to ensure that such an assessment creates a reasonably comprehensive and accurate picture. The problem, therefore, is far from simple: ultimately, it is about how society ensures that the research being carried out within the Danish universities is of high and, preferably, increasing quality. It is in everyone’s interest that the best research is supported. This applies to both those whose research is assessed and those funders who invest in the research. If inappropriate methods are used to assess research, or if recognized research assessment methods are used inappropriately, society will not obtain optimal benefit from the limited research funds available. It is therefore important that the various actors in the field at the political level, in the research funding sphere - both public and private - and in the universities, are aware of any weaknesses in the current research quality assessment practice.

In our discussion, we mean by research primarily the research that is conducted within the Danish universities. In general, it is more difficult to assess the quality of basic research, whereas research that targets a specific practical problem can often be assessed on the basis of whether the problem is being solved. The term (good) research quality is used in the White Paper in the broad sense of research that makes a difference over the short or
long term. This means that good or outstanding research is that which generates new knowledge that is - or proves to be - important to others than the individual researcher, whether they be peers, researchers from other disciplines or outside ‘recipients’ of the research results.

There is a clear time perspective in assessments of research quality. Few would dispute the fact that the quality of a research product can be reasonably and objectively assessed decades after publication. By that time, in most cases one knows whether the research has made a difference, either by showing new pathways or (probably more typically) by making an important contribution to the body of knowledge. But after such a long time, quality assessments have no consequences for anyone other than the prestige of the individual researcher and their institution. It is therefore understood that when we in this White Paper discuss whether quality can be assessed, our interest is in an assessment in the here and now - not in 100 years.

One factor that is important to keep in mind is: Who makes the assessment? A distinction must be made between whether research quality is being assessed by experts within the relevant field or by non-experts. In the latter case, for example, the assessment could be based on measuring the number of publications, number of citations, journal ranking, etc. Another factor that is important to be aware of is that it will generally be easier to assess the quality of research that is directed toward solving a specific practical problem than the quality of pure curiosity-driven basic research. Having said this, it is perhaps most important to keep in mind in the discussion that the disciplinary fields are extremely diverse. One cannot assume a priori that there is a uniform answer to the White Paper’s implied question about whether research quality can be assessed in a reasonably objective way. It may well be that the answer depends on whether the research has been carried out in the field of, say, Medicine, Egyptology, or Theoretical Physics.

In light of these considerations, the purpose of this White Paper and the Annual Meeting is more to stimulate discussion and raise issues than to
provide cut and dry answers. Nevertheless, during our deliberations and discussions in the preparation of the White Paper, we have allowed ourselves to elaborate some recommendations.

We have chosen to structure the White Paper into three chapters, each of which focuses on different areas where quality assessments of research play a crucial role: Chapter 1 deals with the assessment of applications for competitive research grants. Chapter 2 discusses assessment of research quality in connection with hiring of scientific staff, typically those applying for positions as university lecturers or professors. Finally, Chapter 3 discusses assessment of research quality in relation to strategic initiatives and breakthroughs.

With best wishes for good reading and a good debate at the annual meeting on March 24.

On behalf of the Science Policy Committee,
Royal Danish Society for Sciences and Letters

Jeppe Dyre
CHAIRMAN (ROSKILDE UNIVERSITY)

Morten Rievers Heiberg
(UNIVERSITY OF COPENHAGEN)

Thomas Sinkjær
(LUNDBECKFONDEN)

Ole Wæver
(UNIVERSITY OF COPENHAGEN)

Birgit Schiøtt
(AARHUS UNIVERSITY)

Frederik Stjernfelt
(AALBORG UNIVERSITY)
1. We recommend that research funders and institutions continually, critically evaluate their selection processes - including the use of peer review - in order to increase innovation and diversity, as well as, reduce bias and burden on applicants, reviewers and the funds/institutions themselves.

2. Inspired by recent initiatives in Switzerland, New Zealand, and Germany, we propose that one or two public foundations launch test trials for awarding funds on the basis of a grant lottery among qualified applicants, possibly modified so that a small percentage of absolutely unique applications are granted funds outside the lottery.
3. We propose that the universities ensure that the composition of all hiring committees for scientific staff contain a clear majority of associate professors and professors with specific expertise in the relevant field.

4. We recommend that strategic focus areas at the national level be identified by the preparation of a catalogue of ideas based on a broad consultation process. The process must include all relevant stakeholders, from researchers to interest organisations, who can submit proposals that are subsequently considered by one or more expert groups.
We have a strong reluctance to admit uncertainty and indeterminacy in human affairs. Rather than accept the limits of reason, we prefer the rituals of reason

(JON ELSTER, 1989)
The world’s total annual investment in research and development of just over DKK 13,000 billion (OECD, 2017) contributes significantly to the world community. In Denmark alone, the total research investment by the public sector, private foundations and companies exceeds DKK 60 billion annually, of which approximately DKK 23 billion is invested in research conducted at public research institutions (Forskningsbarometeret, 2018). These massive investments have naturally led to an increasing desire to understand and improve the way in which research fulfils society’s needs.

There has been less focus, however, on understanding the actual selection of the research that receives support. The selection by colleagues, known as ‘peer review’, is used predominantly in the awarding of competitive research funding, both in Denmark and internationally. According to the latest prognoses, the competitive research funding system will soon comprise more than 50% of the total research funding managed by Danish public research institutions (DFIR, 2019). Ismail et al. (2009) have estimated that more than 95% of UK medical research funding has been awarded on the basis of peer review assessments.

Peer review assessment in the evaluation of research applications is a recognized meritocratic mechanism that puts the academic argument to its ultimate test in order to discover potential, identify quality, and define who best deserves to be supported (Roumbanis, 2019; Norn et al., 2019).

THE PEER REVIEW PROCESS

The vast majority of research applications are evaluated by peer review panels consisting of recognized researchers with experience in assessing research projects in a broad field of study. The panels typically consist of 5 to 15 specialists depending on the number of applications and the breadth of the subject area. Each application is often assigned two to four panel reviewers who have individually read the application prior to the panel meeting and have given the application a number of comments and numerical scores.
Subsequently, the reviewers in the assessment panel will meet jointly to discuss and make decisions on which applications deserve support, in accordance with criteria established in advance. The benefit of having several panel members assess each application is that a ‘360 degree review’ is typically conducted with a view to: the research’s novelty value, the potential to actually carry out the proposed project idea and the academic level of the applicant and the professional environment in which the project is to be carried out. Assessment panel members often highlight different elements of an application, and through the professional discussion among the panel members, it becomes clear whether the application has shortcomings that make it unworthy of support. It is in this process, where the quality of the application is often translated into a numerical score, from where emerges a differentiation between those applications that will obtain financial support and those that will not.

Overall, the assessment panels work well in rejecting applications with obvious shortcomings. The assessment process is also good at identifying the few absolutely, extraordinary, outstanding applications (typically less than 5% of all applications). The problem lies with the applications that are just as good (perhaps even better) that the assessment process does not ‘catch’. In a situation where the proportion of successfully funded applications is low compared to the total number of applicants, which is the case for the vast majority of the competitive research programs today, it is a matter of concern if the evaluation system cannot distinguish the good application from the outstanding, within a pool of high quality applications. Studies have shown that in highly competitive situations, i.e., with low success rates for applications, for some application processes with success rates of 20%, there emerges a systematic bias in terms of, for example, subject area, institution, gender, race or professional status. This bias seems to occur regardless of how much effort has been made by those in the assessment panels to select the best projects for support (Fang and Casadevall, 2016).
HOW RELIABLE, EFFECTIVE AND FAIR IS THE ASSESSMENT PROCESS IN PRACTICE?

The evaluation of scientific ideas is inherently uncertain because both bias and risk are embedded in the final result. Philip Tetlock, for example, has shown that experts not only did poorly in attempting to predict the future, but also overestimated their own ability to do so (Tetlock, 2005). Is the peer review system capable of recognizing potentially transformative research? Probably not, because intense competition for funds encourages both reviewers and applicants to be risk averse. The very structure of the assessment system encourages consensus and thus the suppression of innovation of the kind that can lead to scientific breakthrough (Fang and Casadevall, 2016). Or, as Nobel laureate Roger Kornberg has said:

In the present climate especially, the funding decisions are ultraconservative. If the work that you propose to do isn’t virtually certain of success, then it won’t be funded. And of course, the kind of work that we would most like to see take place, which is ground-breaking and innovative, lies at the other extreme (ibid.).

Based on a review of approximately 100 studies on the use of peer review to identify which research applications should receive funding, Guthrie et al. (2018) found a notable lack of evidence for the effectiveness of peer review in the awarding of grants, considering the central role of peer review in the entire process (see table 1). Based on the available studies, they find a bias against innovative research. There is also evidence that, at best, peer review assessment is only a weak predictor of where research investments should be placed and that the professional judgment varies significantly among reviewers.
<table>
<thead>
<tr>
<th>EVALUATION QUESTION</th>
<th>GENERAL CRITIQUE</th>
<th>PARTICULAR CRITICISM(S)</th>
<th>IS THE CRITICISM VALID?</th>
<th>STRENGTH OF THE EVIDENCE BASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is peer review an effective system for awarding grants?</td>
<td>Peer review does not fund the best science</td>
<td>It is anti-innovation</td>
<td>Yes</td>
<td>Suggestive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It does not reward interdisciplinary work</td>
<td>Unclear</td>
<td>Suggestive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It does not reward translational/ applied science</td>
<td>Unclear</td>
<td>Suggestive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It is only a weak predictor of future performance</td>
<td>Yes</td>
<td>Agreement</td>
</tr>
<tr>
<td>Peer review is unreliable</td>
<td>Ratings vary considerably between reviewers</td>
<td>Yes</td>
<td>Agreement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>It struggles to achieve an acceptable level of consistency</td>
<td>Unclear</td>
<td>Conflicting</td>
<td></td>
</tr>
<tr>
<td>Peer review is unfair</td>
<td>It is gender-biased</td>
<td>Unclear</td>
<td>Conflicting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>It is age-biased</td>
<td>Unclear</td>
<td>Conflicting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>It is biased by cognitive particularism</td>
<td>Unclear</td>
<td>Conflicting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>It is open to cronyism</td>
<td>Yes</td>
<td>Agreement</td>
<td></td>
</tr>
<tr>
<td>Peer review is not accountable</td>
<td>Review anonymity reduces transparency</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Peer review is not timely</td>
<td>It slows down the grant award process detrimentally</td>
<td>Unclear</td>
<td>Suggestive</td>
<td></td>
</tr>
<tr>
<td>Peer review does not have the confidence of key stakeholders</td>
<td>It is not the preferred method of resource allocation</td>
<td>No</td>
<td>Agreement</td>
<td></td>
</tr>
<tr>
<td>What is the burden of peer review on the research system?</td>
<td>Peer review is an overly burdensome way of distributing research funding</td>
<td>Burden of peer review is increasing</td>
<td>Yes</td>
<td>Agreement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Burden of the peer review system is high and falls primarily on the applicants</td>
<td>Yes</td>
<td>Agreement</td>
</tr>
</tbody>
</table>

Source: Taken directly from Guthrie et al., 2018; page 3, table 3, ‘Summary of evidence from the literature regarding the effectiveness and burden of peer review’.
Based on three different statistical analyses, Pier et al. (2017) have examined how the participants in one of the National Institutes of Health’s (NIH) assessment panels evaluated the proposals they were given, and whether there was agreement in the reviewers’ assessments of the same proposal. Regarding the assessment of the quality of the applications, the results showed no significant agreement among the reviewers in their assessments, neither in their qualitative nor quantitative evaluations. Nor did the authors find any agreement in the reviewers’ ‘translation’ of the qualitative strengths and weaknesses of an application into a numerical score. They concluded that the assessment of the applications was more dependent on the reviewer who was assigned to assess the proposal than the research project being proposed in the application. Previous studies from other grantees support these findings (e.g., Cole et al., 1981).

So-called ‘ex-ante’ and ‘ex-post’ analyses support the conclusion that it is difficult for the peer review system to distinguish between a good application and an outstanding one. For example, a qualitative analysis of grants from the European Research Council (ERC) shows no correlation between ex-ante and ex-post ranking of successfully awarded ERC projects (see figure 1).

If one examines a number of the classic academic measures of impact, such as the number of publications and citations, we find the same tendency. Figure 2 shows the results of a study of the impact of NIH grants. It shows that within the field of cardiac research, those research projects which had the lowest or intermediate scores in the assessments performed just as well or even slightly better than the highest ranked grant projects in terms of number of publications per grant and citations per million dollars (Mervis, 2014).
No significant correlation seems to exist between the rankings in the ex-ante evaluation and the ex-post evaluation.

Source: European Research Council Executive Agency (ERCEA). Direct reproduction of figure ‘Qualitative Evaluation of ERC Completed Projects, Analysis (ex-ante versus ex-post)’.
Equal impacts. The publication record from proposals with the best scores was no better than for those scoring in the middle and lowest tiers among heart research funded by NIH.

* in the first 2 years after publication

Source: Authors’ own reproduction of the figure in Mervis (2014), page 597.
Is it a problem that the assessment system cannot distinguish the good application from the outstanding application within a pool of high-quality applications? Unfortunately, several studies show that there exists a systematic bias towards those who already have much, receive more (the Matthew effect - see, for example, Norn, 2019). Scientific output is not a linear function of the amount of funds awarded to individual researchers. When financing per grantee increases beyond a certain point, productivity decreases. Such a decline in marginal return also seems to apply to grant funding to US institutions (Wahls, 2019). Analyses of data for a representative cross-section of research institutions show that, on average, prestigious institutions had a 65% higher success rate in obtaining funding for a research application and received 50% greater amount of grants than those from less prestigious institutions; this was despite the fact that the less prestigious institutions actually produced 65% more publications and had a 35% higher citation rate per USD awarded. These findings suggest a systematic bias and a social prestige mechanism in the awarding of research grants based on peer review.
WHAT CAN BE IMPROVED IN THE PEER REVIEW SYSTEM?

When the success rate is 30% or higher, the reliability of the assessment panels’ decisions is high (statement by Richard Nakamura, former director of the Center for Scientific Review, NIH, at a meeting of the RoRi Welcome Trust, 2019; see also Li and Agha, 2015). On the other hand, if the success rate is low, the peer reviews cannot distinguish between the good application and those that are truly outstanding and innovative (Fang et al., 2016).

The current rating system can be improved in a number of ways (see, e.g., Guthrie et al., 2018). For example, reviewers with different cognitive skills could be involved. Specialists in translational or innovative research could be involved when the focus is on highly innovative research (Langfeldt, 2001). This approach has been used in grant-giving schemes such as the NIH’s high-risk Pioneer Award Program (Gewin, 2012). The reviewers could also be asked to state how confident they are of their evaluation, and the assessment panel could then incorporate this aspect into their final recommendation. This idea is currently being tested in assessments conducted by The Villum Foundation and the Lundbeck Foundation.

Another approach is to select researchers according to their merit, regardless of what kind of research they plan to conduct. The researchers then have the freedom to pursue new ideas and work in a more flexible way when opportunities arise (e.g., the MacArthur Fellows Program and the Howard Hughes Medical Investigator Program of the Howard Hughes Medical Institute). Of course, such a program is best suited for researchers who have a well-documented research career at the time of application.

Another strategy is to ensure better guidance for those researchers who have been selected to conduct the peer review assessments (Sattler et al., 2015). Both beginning and experienced reviewers become more reliable in their qualitative and quantitative assessments after having received guidance. The results indicate that both experienced and beginning re-
viewers do not have sufficient understanding of the criteria for how the applications should be assessed.

For research programs with very low success rates such as the programs of the Independent Research Fund Denmark (DFF), however, the above-mentioned measures are hardly sufficient. An increasing number of researchers are proposing a hybrid model that blends peer review and lottery. Roumbanis et al. (2019) believe that it is impossible to predict the truly innovative projects using a classic peer review system because these projects so often do not progress as described in the proposal and are difficult to capture in an assessment system based upon established application templates, where the importance of number of publications, impact factors, citations, etc. is weighted so highly.

Fang and Casadewall (2016) propose awarding project grants using a modified lottery system, whereby peer review is used in a classical meritocratic way so as to identify the most promising group of applications (see figure 3). If the applications in this pool exceed the funds available, a lottery-style drawing takes place until the funds are exhausted. A combination of peer review and a subsequent lottery would have several advantages over the current assessment system, including a likely reduction of bias and improvement of diversity in terms of research topics, seniority, race, gender, social prestige mechanisms and the like.
FIGURE 3: MODEL FOR A MODIFIED LOTTERY FOR THE AWARDING OF PROJECT GRANTS.

Model for a modified lottery for awarding of grants. First, peer review assessment is used to select those applications determined to be meritorious or non-meritorious. Non-meritorious applications receive detailed critique and can be re-submitted. The meritorious applications are placed in a consolidated pool, from which a computer randomly selects which applications will be awarded a grant. The lottery drawing is carried out until the total grant amount is distributed.

A combined peer review and lottery appears to be a timely method, but politically, it is a challenging idea. The method is now being tested in New Zealand, Switzerland, and Germany. The expectation is that the diversity of research topics will increase and that bias in terms of seniority, race, gender, and social prestige mechanisms will be reduced. The assessment panels will continue to be the cornerstone in the assessment of the applications. The novel aspect is that the panels will no longer have to rank the individual applications. Instead, they will be grouped into a meritorious group, a non-meritorious group and possibly a group that everyone in the panel agrees is so unique that it should therefore receive a grant prior to the lottery drawing. For the applicants, such a system will hardly change the way in which they prepare their application, except for the time saved from not having to have a technically perfect application in order to have a chance of support.

A growing number - but still few - of research funding bodies are experimenting with new ways of identifying innovative research. One of the challenges is that it will take several years to determine whether new ways of distributing research funding make any difference. In Denmark, the Villum Foundation, the Lundbeck Foundation, and the Novo Nordisk Foundation, which award grants within natural sciences and biomedicine, are testing new assessment procedures that hopefully will be an improvement over the current system in fulfilling the goals of these research programs. The idea of these three Danish foundations is to give researchers with an innovative research idea an opportunity to test it, regardless of which stage they find themselves in their careers or the length of their CV. For more information on these programs, see the list of references.

There is a need for open, transparent testing and evaluation of the way we fund research. It requires the entire scientific community be more receptive to such studies.

We recommend that those foundations and institutions that fund research continuously re-assess their selection processes, including the use of peer review, in order to increase innovation and diversity and to reduce the bias
and burden on applicants and reviewers and on the foundations/institutions themselves. Specifically, we propose that one or two public grant giving foundations experiment with awarding grant funds using a lottery system among a pool of qualified applications.
ASSESSMENT OF RESEARCH QUALITY IN ACADEMIC HIRING
How do we select the best researchers? What criteria should be used in the recruitment of scientific staff? From a ministerial point of view, they have wisely chosen to deal with these issues at the general level. According to the Danish University Act, the university is ‘tasked with performing research and offering research-based education at the highest international level in the disciplines it covers’ (Universitetsloven §2). The equal status given to research and teaching is also emphasized in the Ministry of Education and Research’s new executive order on university positions. Nevertheless, the equality between research and teaching is probably more formal than real, as the same executive order states that research forms the primary basis for research-based teaching (BEK no. 1443 of 11/12/2019). For the same reason, the research is usually given priority in the hiring process, since the character of the research can be said to indirectly influence both the quality and the kind of teaching that the university is able to offer. Research-based teaching is considered by the Ministry as the primary means of applying and disseminating research-based knowledge to the surrounding community (ibid.).

The mantra of most universities is presumably that you can teach an excellent researcher how to teach, while it is far from given that you can teach an excellent teacher to do research. Nevertheless, it sometimes happens that the competent, albeit in comparison to other candidates less outstanding researcher, will be the preferred candidate if the individual has outstanding teaching-related qualities within the ‘bread-and-butter’ disciplines which are especially important for the department where the person in question is to be hired. This reflects the current budget model for the Danish university funding, by which the universities’ primary source of income - at least for those faculties with less external funding such as Law, Theology, Social Science, and Humanities - derives from payments per student rather than massive amounts of external research funds.

An otherwise a well-qualified academic candidate may also fall short because of insufficient linguistic competencies if, for example, high-level Danish proficiency is regarded as a prerequisite for being able to fill a position at the university in a meaningful way. If the candidate, e.g. a post-doc,
is being hired only to undertake a research assignment linked to an externally funded project, the scientific level will naturally be given highest priority, whereas teaching experience will be a less important qualification.

Even within those academic positions where research quality must be expected to play a crucial role, such as in assessments of (full) professor positions, considerable uncertainty can arise. How do we measure the quality and originality of the research? And even more precariously: How do we measure it in a formal comparison between several candidates with different research profiles?

In filling scientific appointments at the Danish universities today, a researcher is assessed primarily according to the quality of their research contribution, typically in terms of international publications, although there are, of course, great differences between research areas (cf. the Royal Danish Academy of Sciences and Letters White Paper 2019). This development has been positive in many respects. It is no longer possible, as previously, to pursue one’s work on, for example, another country’s history or literature if one writes exclusively in Danish and thus cannot be read or assessed by the foremost foreign researchers in the field. This development helps to ensure that Danish universities possess a high level of research capacity in these areas. A more controversial development is the increased focus on research indicators, which endeavour to say something about the researcher’s productivity, measured in terms of ‘impact’ in the form of citations by others. Citations, impact factor and H-index now form the basis for the ‘bibliometric’ measurements that are increasingly used as a factor in making scientific appointments at the universities.
THE HIRING PROCEDURE AT THE DANISH UNIVERSITIES

There are marked differences between the hiring procedures at the Danish universities. However, when a position is to be filled, two committees are normally formed: a hiring committee and an assessment committee. The assessment committee chairman also sits on the hiring committee as an ordinary member. But here the similarities end, and for the same reason, it would be a near impossible task to describe all the various hiring procedures used across the Danish universities. However, the Science Policy Committee considers it important to problematize one particular approach that has become established in some institutions, namely that the chair of the hiring committee is the dean, assisted by the head of department and the director of studies without significant representation from the scientific staff. Common to the three representatives from the management is that they are rarely experts in the field of expertise for the specific position. In addition, the chairman of the assessment committee, who, as mentioned, also sits on the hiring committee, typically has specialized knowledge in one of the areas that is being evaluated, in that the position requirements can be quite broad. It is thus only the written documentation, i.e., the formal peer review carried out by the other members of the assessment committee, which constitutes the actual scientific foundation during the hiring process. The selection of the ‘right’ candidate is thus carried out by a committee that is largely under the influence of the department’s management team at the expense of scientific core competence.

It is important to point out that, in some institutions, a more appropriate procedure is used, whereby the dean is not a member of the hiring committee and it is instead the head of department who is chair. In this approach, there are more people with specific professional expertise on the hiring committee. Hiring processes that follow this procedure also use a thorough interview process where those highly qualified applicants selected by the hiring committee visit the department for 1-2 days and meet the majority of the scientific staff.
It is important for Danish research that hiring committees always possess a high level of professional expertise. For example, the challenge of advertising a position with a broad profile consists of comparing two candidates within two not totally uniform research fields, where both applicants are qualified and have also used outstanding, but also widely different, publication channels such as journal articles and monographs. Here the bibliometric research indicator (BFI) system falls short, for how do you compare apples and oranges?

One could argue that the candidate in the first case, through an active publishing strategy in highly ranked journals, has secured for him/herself a strong position within the research field with a large number of citations.

Conversely, the monograph is often associated with considerable degree of robustness (a quality that can only be recognized over time) and also a special prestige, as the monograph typically provides an exhaustive overview of an entire area, problem or field of research, has a longer chronological perspective and elucidates central issues from several perspectives using a much larger set of empirical data. A monograph will often be a natural jumping-off point for more specialized articles on sub-topics. However, if the two candidates’ qualifications are calculated solely on the basis of bibliometric research indicators, the monograph author will draw the shortest straw in as much as articles generate relatively more ‘points’ in this measurement system.

Within the humanities, a monograph also has the quality that it can often appeal to a wider academic audience. As we know from abroad, there is an increasing tendency to also assess the quality of research in terms of its importance to ‘the whole society’, although it may be quite unclear what this means. Hence, a report from the Independent Research Fund Denmark points out that instead of focusing on publishing channels, number of patents and the research’s economic impact, focus should be placed on the contribution of the research to civil society, policy development, education, patient organisations, etc. (Budtz Pedersen, 2017). This tendency can also be traced in the Ministry’s aforementioned Executive Order on
the structure of academic positions, which states that ‘the university must contribute to promoting growth, welfare and the development of society. As a central knowledge- and culture-bearing institution, the university must collaborate and exchange knowledge and skills with the surrounding community. The university, as an integrated part of its work, exchanges knowledge and competences on a reciprocal basis with a wide circle of actors, organisations, authorities, public and private companies and other institutions of higher education, etc.’ (BEK No. 1443 of 11/12/2019). However, one is tempted to ask, is it at all possible to measure this kind of impact? There are certainly many people working within the social and human sciences who feel that their research has such an effect, but can this kind of societal impact be calculated in any meaningful way in connection with considering an applicant for an academic position?

This account provides insight into some of the problems that arise in assessing a researcher’s scientific activity in connection with the university hiring process. As shown, several of the criteria included in the assessment are legitimate in their own way, but not particularly scientific. The same can also be said of the selection procedure itself, which, among other things, entails that only a fraction of the total field of applicants is subjected to a thorough scientific assessment, while the majority are weeded out during the preliminary screening. In this context, it is important to point out that this initial screening is not carried out by the assessment committee, but by the hiring committee, which, as mentioned earlier, does not always have the necessary scientific expertise, but is instead often dominated by the management team. Typically, this initial screening does not involve a close reading of the submitted scientific publications or a written summary assessment of the candidate. Which dean or head of department would have the time to undertake such a task?

Therefore, it is reasonable to assume that a hiring committee, in connection with such an initial screening, will endeavour to identify that research which immediately matches a number of formal quality criteria (e.g. impact and number of citations), and which is thus immediately recognizable as indicative of ‘good research’. Hence, it can easily occur that original research
contributions never even reach the assessment committee. Again, there are various practices across the Danish university landscape, and there are institutions where a larger proportion of the scientific staff is responsible for shortlisting, with the head of department as chair of the committee.

QUALITY RATHER THAN QUANTITY

Bibliometric measurements provide no guarantee of good research, as was demonstrated in last year’s White Paper on the basis of a larger analysis of the bibliometric research indicator (BFI) system (See the Royal Academy of Sciences and Letters White Paper 2019; Poul Erik Mouritzen, Niels Opstrup and Pernille Bak Pedersen, 2018). On the contrary, such measurements can be self-reinforcing. Instead of promoting creativity, they favour an established quantitative and formalistic understanding of what constitutes good research. Within the humanities, the inherent bibliometric rationale seems to be that when researcher X has achieved so and so many points and has published in journals Y and Z, he or she must be considered a good researcher. Do we have any guarantee that this is indeed the case? Hardly.

There are institutions where the hiring process can now be carried out in just a few weeks, and the considerable time pressure helps to reinforce the quantitative and formalistic aspect. Thus, we should not underestimate the importance of the fact that there are far more hiring appointments in the Danish university system today than just ten years ago. Looking at trends in the number of full-time equivalent positions for scientific staff at the universities, this number has increased by just over a third: from 12,715 full-time equivalent positions in 2008 to 17,355 in 2018 (Danske Universiteters statistical agency, table C).

In 2019, the Science Policy Committee interviewed a group of younger researchers, all of who expressed their frustration that their research careers apparently depended on the quantity and not the quality of their research, and on their ability to obtain external research grants. In the field of medicine, younger researchers experience a one-sided research environment,
where focus in some places is on researchers with a specialty in a concrete skill, i.e. being able to cure a disease, whereas the development of new methods that can contribute basic insights to the field has lower priority (Royal Danish Academy of Sciences and Letters White Paper 2019). Given the large grants allocated to the medical faculties during these years, and the many research positions that come with these grants, it is natural to ask ourselves whether we as a society are getting enough quality for the money invested.

As should be evident, the selection criteria for hiring scientific staff are seldom unambiguous, despite what various executive orders and guidelines formally prescribe. Rather, they are situationally determined. For the same reason, there will often be discussions and murmuring in the corridors when a position is to be filled. Indeed, it can never be completely objective, formal criteria, but rather a question of an overall assessment, which is affected by a number of judgements and interests. Pursuit of own interest can also be about which strategic areas should be prioritized in terms of new appointments. It is not an unknown phenomenon that powerful researchers - by virtue of their prestige and ability to procure external funding - can influence the hiring landscape of the university in a certain direction, even though there might be important professional arguments for rewarding other scientific areas.

Overall, therefore, the question revolves around the extent to which hiring scientific staff in Danish research institutions should be a process guided by scientific considerations. It is clear that the management team reasonably best ensures a number of institutional considerations, but it is hardly desirable, in view of the situation in several Danish institutions today, that the management, based on the University Act of 2003, can effectively decide who are the most suitable candidates for a scientific position.
3

STRATEGIC INITIATIVES AND INNOVATION
In the two previous chapters, we have examined the assessment of research quality in applications for, respectively, research grants and academic positions. This third chapter deals with quality assessments in connection with large-scale or completely new, innovative initiatives. The chapter has two parts: the first part focuses on initiatives to develop a given area with additional research funding. The second part discusses funding within existing programs aimed at identifying radically new projects that can break new ground.

There is a fluid boundary between the second part of this chapter and the discussion in Chapter 1 of the category of ‘outstanding’ grant applications. However, we treat these research grants separately in this chapter because it raises particular challenges, specifically that of searching out what is radically new. In addition, this entire field of ground breaking research is so important that we have no guilty conscience covering it one-and-a-half times.

**STRATEGIC INITIATIVES**

In the absence of selecting specific focus areas, research typically develops very slowly and is ‘path dependent’. That is, we get more of what we already have, with normal, gradual improvements. Internal balances in one field, all things being equal, will be maintained in the next generation of researchers, even if professional developments in the discipline might entail a shift in the priorities among the various sub-fields. Sometimes, however, it is necessary to identify areas that need more funding than previously. This may be because it is considered useful for society (‘strategic research’) or because it is thought that this particular area will develop in a more dynamic way during the coming years, and that it is the Danish national interest to move faster in this particular area than we would normally by simply adapting to the normal pace of research progress.

In large and well-run disciplines, this strategic priority process can occur internally in that the management and staff jointly identify those areas that should be accorded higher priority, for example by employing more
staff in the prioritized area. However, most departments do not have such financial leeway, nor do they have the internal procedures to undertake what may often be controversial choices among colleagues’ specialties for the identification of which areas will be important in the future (and by implication, which will be less important).

Who can identify important new areas and do it in good time? Politicians have sought to do this both through the former Strategic Research Council and presently through the Innovation Fund Denmark, as well as in the form of separate ad hoc grants made directly from individual ministries. Things often go wrong. When an area has achieved the visibility required for political support, it will already have obtained a level of maturity that the Danish research in the best case will be given funds to keep up with developments elsewhere. If we consider only applied research, this strategy can be useful enough, in that we obtain Danish research communities that can participate in the implementation of an existing development (e.g., climate adaptation, artificial intelligence, etc.). However, if the goal is to identify future research needs so early that an extra input is made at the time when this field is taking shape, then the process must involve researchers, and not just the normal research system, because here, too, the future field will be invisible. It requires very specific procedures, to which we shall return below.

The other problem with priorities set by the political-administrative system is that there will often arise compromises where, instead of making real choices, broad fields such as ‘climate’ are selected, after which the subsequent implementation of the initiative very easily results in an emphasis on the existing areas, not the new ones.

This is an almost impossible dilemma because it is about discovering what does not yet exist. The Danish research policy system has previously shown creativity in developing appropriate systems. For example, in the preparation of the ‘Research 2015’ report (‘Forsk2015’, Ministry of Science, Technology and Development, 2008), which in 2008 selected 21 special focus areas a broad consultation process took place in which all stakeholders,
from interest organisations to researchers, could submit proposals, which were then considered by an expert group. In the ‘Research 2025’ report (‘Forsk2025’, Ministry of Education and Research, 2017), this phase had been discontinued. In assessing this experience, it was pointed out that there were so many proposals for ‘Research 2015’ that officials and expert groups had complete freedom to compose the catalogue of ideas that they still wanted to have made, although some surprises emerged at individual points in the process (Ministry of Science, Technology and Development, 2009; Andersen and Rasmussen, 2012). The broad consultation process was probably more useful for generating public interest and understanding of the significance and character of strategic research than for selecting genuine research opportunities.

In identifying future research fields, a system should be developed that includes a search process that includes the research communities and an expert committee that can then consider the proposals submitted. Here one should distinguish the strategic (that is, the socially relevant) part from the purely research-based identification of dynamic areas. The strategic element requires that politicians and interest groups play a central role, because the strategic issue concerns the direction in which society wants to develop. Here, too, however, researchers should be included, because only they can assess where the research could make a difference (cf. the ‘Research2015’ report). Ultimately, however, this strategic part is necessarily politically based. It is less about research quality, and therefore a detailed treatment would need a different kind of White Paper than the present one.

On the other hand, a key element of the research quality issue is to be able to identify early on which sub-areas of the individual disciplines or interdisciplinary research fields will make a difference in a scientific sense. Where will developments proceed particularly quickly, so that these areas will obtain extra strong impact on the general development of the field? For this purpose, a special ‘beyond the horizon’ committee could be formed. It should not have determinant powers and should not be linked directly to the awarding of funds. In order to reduce the risk of bias based towards the participants’ self-interest, it should be a pure mapping exercise
that produces a catalogue of ideas that can serve as inspiration for Danish research. The committee should be staffed with researchers known for their breadth, openness and preferably knowledge of the history and sociology of science, and they should then initiate a search process with good time for receiving input from all research communities.

IDENTIFICATION OF RADICALLY INNOVATIVE RESEARCH

The first half of the discussion above was about selecting specific, high-priority topics or areas where special funds should be awarded. Another possibility for generating the innovative and future-defining research is to have special programs that assess applications, submitted widely across subjects, based on this innovation criterion. In Denmark, this has been the task particularly of the Danish National Research Foundation. Not all grants from the Danish National Research Foundation must necessarily be radically innovative. Many may be in ‘well-known’ fields, on the basis of which a project can credibly develop toward the highest international level. But a particularly demanding part of the National Research Foundation’s work is to identify those projects that lie outside existing paradigms.

This task is inherently difficult. Even in the peer review process, it is often the case that the most innovative research is not recognized (cf. Chapter 1). Almost by definition, that which is radically new and different is not popular, let alone possible - before it becomes so. Both in the Independent Research Fund, with its councils, and in most private foundations, the proportion of successful applications that make it through the ‘eye of the needle’ is so low that a project proposal will be rejected if it cannot fully explain itself, demonstrate results and present a clear timetable for results within a few years. The Danish National Research Foundation has this possibility, due its longer time horizon and extra focus on unique or unusual research.

Many Nobel laureates, such as the Danish Jens Christian Skou and other researchers who have made significant breakthroughs in research, describe
how today they would have little chance if they had to convince donors (including colleagues) about their project at a very early stage. It is thought provoking - but perhaps not so strange - if there had been a widely held view that this road was navigable, there would already be dense traffic.

A common solution to this dilemma is the ‘excellence’ logic, which focuses on individual researchers. In this logic, the idea is that perhaps others do not fully understand the idea behind a bold project, but one knows that this particular applicant has previously accomplished the unusual and especially valuable, so perhaps they can be trusted to do so again. This line of thinking comprises one of several criteria in some existing programs (e.g. European Research Council). However, there have also been programs that were based uniformly on this kind of logic (Münch, 2007; see also Chapter 1). Broadly speaking, about 20 years ago, there occurred a shift from the funding of research networks (based on the belief that it was crucial to connect different research communities) to focusing on the centre of a project. Networks were too often support pillar for the different participants, enabling them to continue doing what they were already doing, and researchers linked their efforts only in a loosely connected way. Therefore, the funding focus moved to the quality of a research community (as a physical, concrete place) and on the leader of a project. This tendency has produced many good results, but it is not without drawbacks. By definition, the approach is conservative and practically a formalization of the Matthew effect: ‘to everyone who has will more be given’. In addition, it has created a pressure to always put the most prominent researchers as director of projects, even when it is inappropriate in terms of consuming time from their own research, and in terms of the possibility that the crucial researcher in the project may be someone who is ‘one layer down’, and who would be best suited as the actual project director.

From a research quality perspective, however, the crucial question is whether the existing assessment systems are capable of identifying the highest quality in innovative projects. The challenge is connected to the fact that by its very nature, one is often compelled to move interdisciplinarily in order to find new fields that do not correspond to the existing
disciplinary boundaries. But our standards for ‘quality’ are often rooted in the concepts and methods of precisely these individual disciplines. A thought provoking study of ‘quality’ was conducted by sociologist Michèle Lamont of Sorbonne and Harvard. In the 2000s, she studied the work of a number of interdisciplinary foundations and councils (especially within the humanities and social sciences). Lamont found that there could not be formal agreement on definitions of quality across subjects, but nonetheless - informally - the reviewers developed a real practice whereby they acknowledged each other’s respective assessments and the common procedure as defining quality (Lamont, 2010).

Nevertheless, new problems arise when such selections occur over a broader spectrum, across all disciplines. It is difficult to ensure that all subjects or just faculties will be assessed on their own terms. No systematic assessment has been made of this, and such an assessment would also be difficult to carry out. Many scholars, especially in the humanities, and the ‘softer’ fields of the social sciences, feel that when selections occur across all disciplines, there may arise a bias in favour of projects that have the best chances if they include either interdisciplinary collaboration with something more natural-science-oriented, such as cognition research, or if they include a very solid data archive structure. There is a real risk that potentially innovative research will not benefit from this process. It is therefore important that the research funding organisations, if their selections take place across many disciplines, pay particular attention to ensuring that their selection processes take account of academic diversity.

Many foundations use a process whereby the decision-making body itself has a broad interdisciplinary composition, and the assessment of applications is carried out by qualified specialist colleagues. The risk is that the first organ is ‘too high up’ and the second ‘too far down’. That is, the board has too limited expert knowledge to assess the individual applicant, while the scientific peer reviewers are too close, and therefore already ‘know’ in advance what is possible and impossible in their fields, and will therefore reject what the not yet seen, i.e., the potentially ground-breaking project.
The two parts of this chapter complement each other; the one cannot replace the other. The first concerns identifying fields in a completely abstract way, regardless of who is designated to do so. The second discusses how a submitted application is assessed on the basis of its potential quality as truly innovative. Both parts should be strengthened if Danish research is to become better at carrying out genuinely innovative research. This criterion will never apply to the majority of projects and therefore concerns only a relatively small portion of the total research grants. Nevertheless, this small proportion of grants can have potentially great significance.
REFERENCES


Danske Universitets statistiske beredskab, tabel C. Accessed 24.02.20: https://dkuni.dk/tal-og-fakta/beredskab/


Universitetsloven, § 2 i LBK nr. 261 af 18/03/2015


FURTHER READING


Articles on the peer review system:
https://spars.aibs.org/bibliography.html

Read more about the experimental program mentioned in Chapter 1 in:

Villum and Velux Foundations:
https://veluxfoundations.dk/da/content/villumexperiment-hvad-betyder-anonym-ansoegningforskningen

Lundbeck Foundation:

Novo Nordisk Foundation’s NERD program:
https://novonordiskfonden.dk/da/grants/newexploratory-research-and-discovery-nerd/

NOTES

1 See, inter alia, *Fremtidssikring af forskningskvalitet*, the report on the future quality assurance of research quality by the Expert Committee for Results-based Distribution of Basic Funding for Research set up by the Ministry of Education and Research (Uddannelses- og Forskningsministeriet, 2019); *Impact redskaber og metoder til måling af forskningens gennemslagskraft* report published by Independent Research Fund Denmark (DFF); D. Adam D. 2019, Science Funders Gamble on Grant Lotteries, *Nature*, vol. 29, 575, 19 November.