German universities on their way to performance-based management of research portfolios

Jochen Gläser

The rise of «new public management» (NPM) in higher education systems of many OECD countries is based on, and fuels, research evaluation as a political and management tool. Owing to methodological problems of identifying changes in research and causally attributing them to the use of evaluations, there is a lack of evidence for positive or negative effects of evaluations alike. In this article, I argue that we nevertheless have enough empirical material to estimate the strongest possible outcome of evaluation-based research policy and management. Research evaluations back strong political expectations concerning research performance but produce inadequate information, which further reduces the already very limited action capabilities of university management. Under these conditions, which are unlikely to disappear, the institutionalisation of new public management is likely to lead to a performance-based management of research portfolios by universities.

Key words: new public management, higher education governance, research evaluation, university autonomy, excellence initiative

Introduction

The rise of «new public management» (NPM) in higher education systems of many OECD countries is based on, and fuels, research evaluation as a political and management tool. Central tenets of NPM – that competition increases efficiency, that resources need to be allocated to the best-performing units, and that hierarchical management is necessary to make organisations fit for competition – require research performance to be measured. NPM advocates believe that universities need incentives in order to improve their performance and that management decisions about research require information about research performance. Consequently, whenever science and higher education policies move from a habitual distribution of resources and *laissez-faire* governance of research to policies of efficiency-oriented resource distribution and managerial governance, it needs evaluations in order to inform or legitimise decisions.

The use of research evaluation as a tool is thus linked to a paradigm shift in higher education. Evaluation becomes an instrument for institutionalising new public management by means of which universities are transformed into market-oriented, performance-based organisations. In this process, evaluation primarily serves to ideologically justify the changes in resource distribution, organisational structure, and management procedures, and to allow governments to legitimise their policies. This is not to say that evaluation is not an important tool in this process. On the contrary, it is a valuable instrument that is used to inform and legitimise policies. However, the evaluation process is not free from distortions, as it is subject to political and managerial pressures. In this sense, evaluation becomes a means of institutionalising new public management, which is likely to lead to a performance-based management of research portfolios by universities.

---

1. I am indebted to Uwe Schimank for helpful comments.
2. I should point out that this transition also makes obvious that the previous habitual distribution of resources and *laissez-faire* governance were not explicitly justified. They appear to have existed only because nobody thought intervention in the science system necessary or possible, probably due to the conviction that external actors lacked the necessary knowledge.
education governance. This explains the debate about evaluations and their premises. The introduction and ubiquitous use of research evaluations are based on tacit assumptions that the evaluations are valid, can be effectively used in political and management decisions, and thus produce positive effects. Critics of research evaluations point out a lack of positive and an abundance of negative effects, while advocates of research evaluations use the reverse argument. None of these arguments has a strong basis in fact. The lack of evidence is caused by methodological problems of identifying changes in research and causally attributing them to the use of evaluations. Quantitative studies of effects of performance-based funding schemes often use rather simple indicators of research performance and do not pay sufficient attention to confounding variables. Qualitative studies are often limited to collecting data about self-reported behavioural change. When they manage to identify changes in research content, these findings are difficult to link back to the macro-level of knowledge dynamics in national scientific communities (Gläser and Laudel 2016, 129-134).

Thus, in spite of the far-reaching hopes and fears concerning research evaluations, most arguments remain theoretical, and empirically establishing any effects on research may fail due to the complexity of the situation in which evaluations are applied. In this article, I argue that we nevertheless have enough empirical material to estimate the strongest possible outcome of evaluation-based research policy and management. The aim of my article is to demonstrate that research evaluations back strong political expectations concerning research performance but produce inadequate information, which further reduces the already very limited action capabilities of university management. Under these conditions, which are unlikely to disappear, the institutionalisation of new public management is likely to lead to a performance-based management of research portfolios by universities. The emerging quasi-markets for research performance will maintain state-of-the-art research in some universities and fields but are liable to quasi-market failures such as a loss of epistemic diversity and possible other unintended effects. I illustrate this seemingly inevitable evolution of university governance towards research portfolio management by discussing research evaluation systems – national systems for evaluating research performance – in German higher education. The German higher education system and its evaluations can be utilized for this discussion even though Germany must be considered a latecomer to new public management (Schimank and Lange 2009). One might even argue that some features of German higher education such as its constitutionally guaranteed «freedom of research» and its federal structure prevent NPM from becoming as strong as it is in other countries. Nevertheless, the analysis of the German higher educa-
tion reforms offers interesting insights in the possibility of change and potential effects. The late and slow introduction of NPM can be utilized to analyse functions and limitations of research evaluation systems and their use.

I begin by providing an overview of research evaluation systems in Germany (1) and then look at the effects of the most important evaluations (2). The (non-)effects highlight three major conditions under which university management works and point to «research portfolio management» as the likely outcome of increasing pressure on and limited action capabilities of universities (3). As a conclusion, I discuss possible consequences of «research portfolio management» (4).

1. The evaluation environment of German universities

The politically most interesting question about research evaluations is whether they have any effects on research performance at all, and what these effects are. From a theoretical perspective, we need to know how properties of research evaluations and the circumstances under which they operate are linked to particular effects. Answering these questions requires comparative studies of research evaluations, which in turn suggests categorising them. The search for consequential properties of research evaluations has produced several typologies, most of which are based on the variation of purposes (Geuna and Martin 2003, 278; Molas-Gallart 2012, 589). Other typologies attempt to include differences in institutionalization, evaluation modes, and in the strength of an evaluation’s impact on the evaluated unit (Gläser 2007; Whitley 2007).

Evaluations of German university research can be differentiated in two dimensions, namely the kind of information provided by an evaluation (comparative versus customised) and the consequences the evaluation has for the evaluated units (weak versus strong, Gläser 2007). Cross tabulating the two dimensions produces four types of evaluations. Weak and strong comparative evaluations provide information about research performance by applying a common yardstick to all units of assessment within or across higher education systems. The yardstick can be applied by using indicators or peer review. In either case, information content is reduced to numbers, ratings or rankings because comparability cannot otherwise be achieved. By contrast, weak and strong customised evaluations provide specific

3. In addition to the literature on German evaluation systems, my argument is based on three empirical investigations, namely a comparative study of Australian and German performance-based funding systems (Gläser and Lauden 2007; Lange 2007; Gläser et al. 2008, 2010b), a study of responses by German universities to evaluations of their research (Von Stuckrad and Gläser 2012; Gläser and Von Stuckrad 2013), and an ongoing investigation of the impact of new public management reforms on the research-teaching nexus. All three projects were/are supported by the German Federal Ministry of Education and Research.
information about strength and weaknesses of a unit’s research and recommendations for its improvement. They are always based on peer review because the necessary detailed information and recommendations cannot otherwise be produced. The strength of political, reputational or financial consequences of evaluations varies independently of the kind of information provided. Comparative evaluations have strong financial consequences if they affect large proportions of university block funding, as is the case with the UK’s Research Excellence Framework or the Australian funding formula. The consequences of customised evaluations depend on the political or managerial decisions they inform and on the authority of those making the decisions. However, given the cost of peer review exercises, it is unlikely that weak customised evaluations will be conducted.

If we apply these distinctions to the evaluation environment of German universities, we find three of the four cells populated (Table 1). Weak comparative evaluations include performance based block funding of universities, which has been implemented by nearly all German federal states since the late 1990s. These systems are comparative and based on quantitative indicators. The most important performance indicator is external funding, followed by PhD completions. Other indicators (number of habilitation theses, number of publications) are rarely used.

Table 1  Evaluations of German university research

<table>
<thead>
<tr>
<th>Information provided</th>
<th>Consequences for evaluated units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparative</td>
<td>Weak</td>
</tr>
<tr>
<td></td>
<td>• Performance-based funding of universities</td>
</tr>
<tr>
<td></td>
<td>• National and international rankings</td>
</tr>
<tr>
<td></td>
<td>• Ratings conducted by the German Council of Science and Humanities</td>
</tr>
<tr>
<td></td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td>• Excellence Initiative</td>
</tr>
<tr>
<td>Customised</td>
<td>Weak</td>
</tr>
<tr>
<td></td>
<td>• Evaluations by the German Council of Science and Humanities</td>
</tr>
<tr>
<td></td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td>• Evaluations by the German Council of Science and Humanities</td>
</tr>
<tr>
<td></td>
<td>• Evaluations by the Academic Advisory Council Lower Saxony</td>
</tr>
</tbody>
</table>

Source: based on Gläser 2007, 250-256

4. The distinction between comparative and customised evaluations resembles the summative/formative distinction introduced by Geuna and Martin (2003) but differs from it by excluding assumptions about the purpose of an evaluation. Thereby, it acknowledges the possibility that comparative evaluations also provide guidance to universities on how to increase the «performance» of their research, e.g. by goal displacement (see below, 4).
The consequences of performance-based block funding are weak. In 2010, the share of performance-based funding of research in the federal states’ higher education funding varied between 0.57% in Bavaria and 12.5% in Berlin (the only federal state in which the share is above 10%, Dohmen 2015, 97). Possible gains and losses from performance-based funding are further curbed and counteracted by political decisions. For example, the current contract between the federal state of Berlin and its universities limits losses in the overall performance-based funding (for teaching and research) due to under-performance to 5% of the 2014 budget and gains to an annual growth of 1% (Hochschulvertrag 2014a, 2014b). In other federal states, ministries decided not to implement the performance-based redistribution completely or to compensate universities for losses (Jaeger 2008, 40–41).

International and national rankings are a second kind of weak comparative evaluation systems. In addition to international rankings such as the Shanghai ranking and the Times Higher Education World University Ranking, several national rankings have emerged in Germany. The most important national rankings are a regular overview of external funding produced by the Deutsche Forschungsgemeinschaft (DFG) and a ranking produced by the German Centre for Higher Education Development (CHE). The DFG regularly publishes an overview of external funding of universities from the DFG and other funding agencies. Although the DFG states that this information is not intended to be a ranking, it can be and is interpreted as such. The CHE publishes a national ranking of university research in seventeen disciplines based on a survey of universities and on databases. Indicators of research performance include external funding, publications, patents, and PhD completions. In addition, the survey collects data on the reputation of faculties.

These rankings do not have any direct consequences for universities. Although they are increasingly observed by higher education ministries and universities alike, changes in ranks have no financial or political consequences for universities. The only widespread ministerial expectation concerning rankings was the expectation to participate in CHE rankings, i.e. to provide the necessary information.

A third weak comparative evaluation of German university research is the «research rating» that has been conducted by the German Council of Science and Humanities as a pilot study. The rating is based on a peer review of all university research in a discipline. The research conducted at a university in this discipline is rated in the dimensions research quality, impact/effectiveness, efficiency, support of young researchers, utilisation of research findings, and providing findings to the general public. The discipline panels decide on the interpretation of these dimensions and on which data to use in their assessments. For example, the chem-
chemistry panel decided to include citation analyses, while the sociology panel decided not to use citation data but to read publications from each unit of assessment. So far, ratings have been conducted in the disciplines chemistry, sociology, electrical engineering and information technology, and Anglo-American studies. In 2013, the Council of Science and Humanities published a recommendation to continue the ratings in further disciplines (Wissenschaftsrat 2013). The Council of Science and Humanities ratings just provide information for universities and faculties, and do not have consequences for universities at all, financial or otherwise.

The only strong comparative evaluation in Germany is the Excellence Initiative (EI). The EI has been/is being conducted in two phases (2006 to 2011 and 2012 to 2017), in which German universities were/are supported with about €4.6 billion in additional funds. In contrast to performance-based block funding, the EI is an ad hoc initiative aimed at promoting selected universities, which receive funding in order to achieve the highest international level of research performance. It consists of three funding streams, two of which (Graduate Schools and Clusters of Excellence) are administered by the DFG, while the third one (Universities’ Institutional Strategies) is administered by the Council of Science and Humanities. Proposals are assessed by international peer review panels.

The strong consequences of the EI include political pressure, reputational gains and to some extent finances. In the beginning, higher education ministries had strong expectations concerning participation and success of their universities in the EI, which creates considerable pressure for universities. Participation and success were included in performance agreements between ministries and universities. Financial consequences include a large amount of additional external funding for each Cluster of Excellence and Graduate School, and funding for the implementation of Universities’ Institutional Strategies that is under the discretion of university senior management. In addition, several federal states created their own «excellence initiatives», in which they funded the preparation of proposals for the next round of the EI or proposals that were «near misses» in the main exercise (Simon et al. 2010). Finally, there are considerable reputational gains in addition to the funding, particularly for universities whose Institutional Strategies are approved. The EI increased both the national prestige and the international vis-

---

5. For English information on the procedures and results of the ratings in sociology and chemistry, see http://www.wissenschaftsrat.de/en/fields-of-activity/research-rating.html.
6. Ad-hoc initiatives, which have sprung up all over the world, provide additional resources for a fixed term. Both properties are advantageous for the state in its role as funder of university research. The distribution of additional resources does not affect the day-to-day operations of universities the same way as changes in their traditional budgets and thus can be much more selective. The fixed term gives the state tight control of its financial commitments (Gläser and Weingart 2010).
German universities on their way to performance-based management of research portfolios

...ability of the funded Clusters of Excellence, and particularly of the «Excellence Universities».

In Germany, customised evaluations play an important role. The German Council of Science and Humanities regularly conducts such evaluations of research institutes funded by federal ministries and of research institutes that apply for membership in the Leibniz Association, an organisation of state-funded non-university research institutes. Customised evaluations are also conducted ad hoc, e.g. at that request of higher education ministries. For example, a federal state’s higher education ministry may request an assessment of university medical faculties in that state as input in a decision about restructuring them. These evaluations are conducted by peer review panels formed by the Council of Science and Humanities. In Lower Saxony, an «Academic Advisory Council» was formed that organised a peer review of all university research in this state. Universities were asked to prepare a report on the last five years of research activity and future planning in a discipline. Universities were then visited by a group of evaluators consisting of approximately six professors from the evaluated discipline, who worked at universities in other German federal states or abroad. These evaluators talked to the university president, the respective dean, each professor of the discipline, some members of scientific staff, and some doctoral students. Based on these discussions, a draft report about the discipline and its performance at all Lower Saxony’s universities was written by the evaluators. The evaluated units and individuals were then asked for their comments. On this basis, the report was revised and submitted by the evaluators to the Academic Advisory Council. The detailed analyses of strength and weaknesses led to equally detailed recommendations about future directions of research strategies and recruitment (Schiene and Schimank 2007; Meier and Schimank 2010). Evaluations organised by the German Council of Science and Humanities follow a similar procedure. Customised evaluations can have strong consequences because the political actors commissioning them also have the power to see recommendations implemented. Thus, the strength of effects varies because it ultimately depends on the political interests of those who commissioned the evaluation.

2. The impact of evaluations on German university research

Effects of changes in the governance of science are notoriously difficult to identify and causally attribute. This problem is partly due to the difficulties involved...
in measuring (changes of) knowledge content. Behavioural changes of universities and academics are usually easier to observe and causally attribute. In particular, since changed behaviour of actors is a necessary condition for any effect of governance, the absence of behaviour change is sufficient to establish the absence of effects. This is why many empirical investigations limit their analyses of effects to behavioural changes and do not «follow through» to changes in research content (Gläser and Laudel 2016, 129-134).

2.1 Effects of weak comparative evaluations

Not surprisingly, evaluations that don’t have strong consequences for universities do not trigger responses by universities, let alone changes in research. The only substantial change in universities that accompanied the introduction of these performance-based funding schemes was the implementation of similar systems within universities. This implementation was not motivated by the universities’ attempt to reconstruct internally the incentives applied to them. Instead, universities complied with expectations of higher education ministries that they introduce such schemes (e.g. Biester and Flink 2015, 400). The sums distributed by these schemes are relatively small (overall 3-4% of a university’s budget) because personnel costs are exempt from performance-based distribution (Jaeger 2008, 38). The sums professors can win or lose are usually well below €10,000 per year, a sum that is small compared to the amount of external funding required to conduct research in many fields. The only discipline for which this is different is medical research, where much larger sums are distributed (Krempkow and Landrock 2013, 101). In medical faculties, performance-based funding creates stronger incentives for academics. Unfortunately, the actual impact of performance-based funding on quality or directions of medical research has not yet been investigated.

Rankings are largely ignored by universities and researchers alike. Researchers do not consider them valid representations of research performance. University senior management does not perceive strong expectations of ministries concerning these rankings beyond the expectation to participate in the CHE rankings. The ratings conducted by the Council of Science and Humanities did not lead to actions at universities, either. So far, they were only conducted for few disciplines and found little attention by higher education ministries. In some cases, university senior management did not even know that these evaluations took place.

8. There is also a performance-based pay scheme that ties part of professors’ salaries to performance indicators for teaching and research. This system has been introduced stepwise from 2004 onwards – it applies only to newly appointed professors. Pay schemes themselves and the extent to which they are implemented vary between universities. The actual effects on research content or performance are yet unknown (Biester and Flink 2015; Ringelhan et al. 2015).
2.2 Effects of a strong comparative evaluation system: the Excellence Initiative

Although the EI can be considered as having strong effects, it needs to be put into perspective. Its overall financial impact on the research and development expenses of the German higher education system amounts to 3.2% (Möller et al. 2016 forthcoming). Given the severe under-funding of German higher education, this is a rather small amount. In its discussion of the context of the EI, the independent international Panel of Experts compares selected German and foreign universities and provides the following thought experiment. If the RWTH Aachen (one of the «Excellence Universities») had received all the annual funding available in the EI, it would have had an annual budget similar to that of the University of Michigan, which has roughly the same number of students, or the ETH Zurich, which has less than half the number of students (IEKE 2016, 15).

The discussion about effects of the EI also faces the problem that insufficient time has passed for many potential effects to become visible and stable. The political decision about a continuation of the EI after its second round could have been based only on an assessment of the first round. In actual fact, it was made without any support by systematic empirical evidence about positive or negative effects. The basic decision to continue the EI was made before the Independent International Panel of Experts conducted its evaluation. But even the evaluation report stated repeatedly that it was too early to assess effects of the EI (pages 9, 18, 23) or that no effects of the EI were visible (pages 18, 20, 25). The Panel’s report nevertheless identified six «problem areas» of the German university and research system and attempted to identify the contribution of the EI to progress in these areas:

1. Differentiation of universities: the EI was intended to support both a horizontal differentiation of universities (the development of university-specific areas of research strength) and a vertical differentiation (the development of «élite universities»). The panel found the EI to have not (yet) contributed to horizontal or vertical differentiation. However, the EI did contribute to overcoming the illusion that all universities are equal by making visible differences in research performance (ibid., 19).

9. Small amounts can of course make a big difference if they can be invested strategically, as is the case for funding of Universities’ Institutional Strategies.
10. The expectation of the EI to facilitate both horizontal and vertical differentiation includes an interesting contradiction. If universities are expected to specialize in research by developing specific areas of research strength, they will also have areas in which they are weaker. Vertical differentiation, on the other hand, assumes at least some universities to be strong in all areas, i.e. not particularly specialized.
2. Governance of universities: the EI triggered changes in intra-university governance. Being made responsible for applications in the EI, university senior management gained new powers, particularly with regard to applications in all three lines of funding and with regard to the implementation of Institutional Strategies (ibid., 22). At the same time, the Clusters of Excellence formed strong units with their own budget and management, which created tensions with existing organisational structures of universities (ibid., 21).

3. Student numbers and quality of teaching: the report clearly states that the EI did not have the purpose to improve teaching, and contributed only little to that task (ibid., 24-25).

4. Promoting young researchers: although the EI created many positions for young researchers (PhD students and postdocs), the career prospects of young researchers did not improve because their chances to reach permanent positions did not improve (ibid., 28-29).

5. Collaboration with non-university public research institutes: the German science system features a large sector of public research institutes, which include institutes of the Max-Planck-Society, the Helmholtz Association, the Leibniz Association and the Fraunhofer-Society. These institutes receive more funding per full-time equivalent researcher and have much better recurrent funding, which kept increasing over the last decades (Hornbostel and Möller 2015). They also benefited from the EI because their researchers participated in Excellence Clusters. This led to an increase in collaboration between universities and research institutes (IEKE 2016, 30-31; Möller et al. 2016 forthcoming).

6. Internationalisation: the EI advanced the internationalisation of German research because it increased its international visibility and led to a significant recruitment of researchers from abroad (IEKE 2016, 33-34).

Effects of the EI on increased research quality were further studied by a bibliometric analysis of publications from the first round of the EI (Möller et al. 2016 forthcoming). The analysis identifies only a small effect of the EI on the share of publications in the top 10% most highly cited publications. Ultimately, this is not surprising given the limited overall financial impact on the research and development expenses of German higher education (see above, 3.1). Regardless of actual knowledge of effects, proponents lauded the injection of German university research with bitterly needed money and saw «excellent research» and «excellence universities» emerging. Observers also saw a «mobilisation effect» (Neidhardt 2010, 59) and an increase in the authority of university senior management over research. Applications for Graduate Schools and Clusters of Excellence had to be submitted by universities, which made senior manage-
ment select and in some cases even initiate applications. Applications in the third line of funding – for «Universities' Institutional Strategies» – were led by university senior management, which also had full discretion over the money received. Critics of the EI (most prominently among them Richard Münch, see Münch 2007) made four main points. First, the EI is seen as a «pseudo-competition» in which members of an establishment distribute resources among themselves. Second, the overall trend to funding larger units and networks and the use of external funding as an indicator of research quality disadvantage small universities as well as the social sciences and humanities. Third, the EI contributes to a growing imbalance between research and teaching. Fourth, the EI has a negative impact on academic careers, which is due to a disproportional increase in fixed-term positions. Critics have similar problems in providing empirical evidence as the proponents of the EI. In particular, it is not entirely clear to what extent the problem lies with the EI itself or with problematic contextual conditions in the German higher education system.

2.3 Effects of strong customised evaluations

Customised evaluations lead to detailed reports on the research performance of evaluated units and recommendations of measures to improve performance. When backed by the political actors commissioning the evaluations, the assessments and recommendations have strong consequences because they are followed by universities. Recommendations can even be implemented against the intentions of the evaluated researchers. This is highly unusual in the German context because university professors are highly autonomous and their authority over research fields and institutional structures of their faculties is difficult to overcome. Customised evaluations change this status quo through two temporary shifts of authority relations in German universities. First, university professors lose authority over the assessment of their research. The peer review of customised evaluations breaks the «assessment monopoly» of university professors concerning their own research because the assessors are considered at least equally qualified to judge research quality by all parties involved in the evaluation, including the professors themselves. As a consequence, university professors lose epistemic authority over their research to the external scientific élites conducting the evaluation. Thus, customised evaluations redistribute epistemic authority over research from university professors to their national and international peers, and university

11. There is also a social movement that collects signatures against the continuation of the EI, see https://www.openpetition.de/petition/online/fuer-gute-forschung-und-lehre-argumente-gegen-die-exzellenzinitiative.
management gains a relative advantage by this «authority drain» of its professors. Second, university management gains authority through the governance process in which customised evaluations are integrated. When higher education ministries commission such evaluations, they do so with the strong expectation that university senior management implements the recommendations. Ministries thus transfer authority to implement the necessary changes to university senior management. The «authority drain» of university professors is accompanied by an «authority boost» for university senior management, which is achieved and legitimised by the necessary compliance with political expectations.

These changed authority relations in universities are only a temporary effect of customised evaluations. They can, however, contribute to the implementation of lasting structural changes against the interests of the university’s professoriate, including:

- the establishment of new professorships, rededication of vacant professorships within the discipline, elimination of vacant professorships or their transfer to a different discipline;
- the participation of external peers in the recruitment commissions for vacant or new professorships;
- the allocation of additional scientific staff for particular professorships; and
- the allocation of additional resources for specific fields from the government (Gläser et al. 2010a, 161).

The change of authority relations is only relevant if recommendations are implemented against the interests of professors. Customised evaluations may also coincide with the interests of the evaluated professors or provide opportunity structures for professors to act as «institutional entrepreneurs» and initiate new thematic directions. Finally, if evaluations are positive, no change is required and no conflicts will occur.

3. Three key conditions of evaluation-based management

The impact of evaluations on German higher education appears to be rather limited. I already discussed one reason for this limited effectiveness. Evaluations are embedded in complex governance systems, whose many instruments overlay evaluation-based governance with competing influences. Against the background of this generic condition, which limits effects of any governance instrument, we can now proceed to a discussion of specific conditions for the operation of evaluation-based governance.

The increasing use of research evaluations as governance tools occurs in the con-
text of rising expectations concerning the management of research performance by universities and of the limited ability of universities to shape their research at all. Evaluations do little to improve universities’ action capability because the utility of the information they provide is very limited. These three conditions create a trend towards a «performance-based management of research portfolios» by universities.

3.1 Strong expectations

Higher education reforms include the double move of raising expectations that universities increase their teaching and research performance and increasing universities’ capabilities to do so. States transfer authority over budgets and personnel to universities, and legislate structural changes at universities that strengthen hierarchical and weaken collegiate governance (Paradeise et al. 2009). These structural changes are embedded in the discursive reconstruction of universities as collective actors who exercise control over their own core technologies and can interact strategically with actors in their environment (Krücken and Meier 2006; Meier 2009). The rhetoric behind these changes often draws on organisational models from the private corporate sector, and suggests that universities should become more similar to companies and «entrepreneurial» (Clark 1998; Marginson and Considine 2000). Universities are expected to use their seemingly increased autonomy to improve their performance in teaching, research, and other activities society expects.

Although structural changes and accompanying political rhetoric suggest that the autonomy of universities increased, this suggestion is difficult to accept in the light of empirical evidence. The state continues to send signals about expected behaviour, which universities follow closely (Capano 2011; Laudel and Weyer 2014). In Germany, this behaviour of universities includes the participation in evaluation exercises and implementation of recommendations discussed in the previous section, and the adaptation of public relations work to observed interests of state ministries (Marcinkowski et al. 2013).

This seemingly unchanged dependence of universities on the state can be explained with the help of an analytic distinction between strategic autonomy (the autonomy of selecting goals) and operational autonomy (the autonomy to select approaches to goal attainment). Higher education reforms increase the operational autonomy of universities but rarely touch their strategic autonomy. An increase of a university’s operational autonomy does not matter if the university does not have the autonomy to formulate goals (Gläser and Schimank 2014). The strategic autonomy of universities is even diminishing because the state
increasingly measures performance of universities and incorporates public policy goals in its higher education policies (Whitley 2010). Since public funding has been reduced in absolute terms or relative to tasks, universities do not have the opportunity to ignore state expectations that are directly tied to funding. But even if there are no direct financial pressures or incentives, political expectations are still met because the state remains the most important source of funding and legitimacy for universities (Whitley and Gläser 2014).

As a result of these processes, we find universities facing strong state expectations concerning their performance, insufficient strategic autonomy to ignore such expectations, but sufficient operational autonomy to respond to them. The introduction of performance evaluations and performance-based funding institutionalised strong expectations that universities improve their performance as measured by these systems. In addition, universities face strong expectations to follow other state policy priorities, e.g. concerning the development of specific research profiles.

3.2 Limited action capabilities

Although universities do not have much choice but to respond to the expectations concerning their research profiles and performance, the question remains if they can respond by shaping their research. The literature on higher education governance indicates that the extent to which university managers in any country have been able to do that is principally limited.

This is for two major reasons. First, scientific research is inherently highly uncertain and the significance of research results is often contested. Researchers themselves often do not know how their work succeeds or fails, and typically are unsure what its exact outcomes will be. Indeed, in many sciences, specifying the nature of the problem and potential ways of dealing with it are weakly codified activities that cannot readily be reduced to routine problem solving procedures and cannot easily be assessed as worthwhile or competent by outsiders. Equally, the meaning and significance of research results are often difficult to decide even by researchers, and are often subject to later reinterpretation and re-evaluation (Whitley 2008). This is why scientific research constitutes an ‘unclear technology’ (Musselin 2007) for universities as organisations.

Second, universities share authority over research goals and performance evaluation with scientific communities. Researchers develop their research goals and approaches in the context of their scientific communities, i.e. by using their knowledge and considering their preferences. Their findings are contextualised and evaluated by these communities. Whatever else is expected from them by other actors,
they must create contributions to their scientific community’s knowledge in order to remain members of this community and thus to be able to maintain their identity as researchers (Gläser 2006, 2010). As our discussion of evaluations demonstrated, universities typically defer to the authority of national and international scientific élites in determining research priorities and the standards by which academics are judged. While the extent of such international reputational control of research goals and evaluation standards varies between fields and over time (Whitley 2000 [1984]), universities have to accept the judgements of competent researchers throughout the world concerning the nature of important problems and the significance of results when attempting to shape research profiles or performance.

The conduct of research as part of a transorganisational communal effort, and the uncertainty and fluidity of scientific knowledge render some of the tools for controlling professional work ineffective. The formalisation and standardisation of skills (Freidson 1984, 13-18) is limited as scientists constantly develop them in conducting their research. Supervision by colleagues inside the university is also largely ineffectual due to different specialisations in most cases, and occupational control by clients (Child and Fulk 1982, 167-175; Freidson 1984, 7-8; Simpson 1985, 425-426) is difficult because in many cases the only «clients» are other members of the scientific communities. Performance measurements based on citation counts and similar indicators can be interpreted as an attempt to exploit scientific communities in this particular role.

The transition from recurrent to project-based funding that can be observed in all countries further reduces managerial control in universities. Peer review of competing applications for research grants makes researchers even more dependent on their communities, which thereby assume even more authority over research vis-à-vis universities. At the same time, researchers who acquire grants often control them independently, which gives them an additional base of authority and further reduces the authority of managers who control the shrinking recurrent funding. While, then, managers may be able to influence the selection of scientific staff and rely on peer judgements of the merits of their work in making promotion decisions and allocating discretionary resources, their ability to control the choice of work goals, the allocation of time and skills and the evaluation of task performance is inherently limited and subsidiary to continuing peer assessments and research activities. Essentially, their construction of organisation-specific competences based on the collective commitment of academics to organisational objectives and the development of organisation-specific knowledge remains highly uncertain, indirect and reactive to others’ judgements (Whitley and Gläser 2014). In addition to these general conditions, German universities face two additional limitations to their action capabilities. First, a fundamental limitation to chang-
focus

Jochen Gläser

In the governance of German higher education arises from the strong position of university professors, which is fortified by the constitutionally guaranteed freedom of teaching and research. Second, even within the framework permitted by the constitutionally guaranteed freedom of research, actual changes in university governance lag behind those made possible by reformed higher education laws. German universities are still characterised by traditional collegial decision practices and cultures.

At the current stage of higher education reforms, it remains unclear how effects of the constitutionally guaranteed freedom of research on university governance differ from practical constraints produced by the properties of research discussed above. However, limitations due to current decision cultures are likely to disappear with generational changes in universities and their management.

3.3 Information utility

Within the limits described in the previous section, evaluations change authority relations in universities by providing university management with information and legitimisation for decisions on resource distribution, promotion and recruitment. Most importantly, they provide information about research performance that is independent from those that are to be managed. Access to such information makes management decisions on research that contradict interests of researchers easier, which in turn contributes to the hierarchisation of university management. This utility for management depends on properties of the information provided, which vary considerably between research evaluation systems (Gläser et al. 2010a, 154-159). Information about research performance best supports management decisions if it is valid, i.e. reflects actual research performance; timely, i.e. reflects current performance; and rich, i.e. provides a detailed picture of the various aspects of performance. If information about research performance is used in decisions about the distribution of resources between subunits of the university, the information must also be comparable across these units. Finally, the information must be legitimate in order to contribute to the legitimisation of management decisions. The three major types of evaluation systems perform differently in these dimensions (Figure 1). The only advantage of indicator-based comparative evaluation systems is the information’s timeliness. The simple indicators commonly used in

12. Indeed, a reform of the higher education law of the state of Hamburg was invalidated by the German Constitutional Court (20.7.2010; 1 BvR 748/06) because it included a shift of authority from professors to deans that was considered unconstitutional. The Constitutional Court stated that universities can only develop such organizational structures and procedures that enable the application of their competencies by the carriers of the freedom of research, i.e. professors.
such systems (numbers of publications, external funding, PhD completions) can be linked to recent research performance. However, the information richness of these indicators is very low. Furthermore, the comparability of information suggested by the numbers is fictitious because the numbers are not comparable across or even within disciplines. For example, publication practices vary strongly due to communication channels used (e.g. books versus journal articles versus conference contributions) and varying typical durations of research processes (see also van Eck et al. 2013 on the incomparability of citation counts for basic and clinical research in medical fields). Researchers in many fields consider the legitimacy and validity of indicator-based evaluations as low because they are only weakly linked to the actual research.

If comparative evaluation systems utilise peer reviews, as is the case with the UK’s Research Excellence Framework and the research rating by the German Council of Science and Humanities, timeliness of information is lower because peer reviews are conducted at lengthy intervals and the timeliness of information decreases between the exercises. In order to achieve comparability, assessors have to reduce their complex judgments to grades or ranks, which means that the information richness is reduced to that of indicator-based systems. Comparability
focus

Jochen Gläser

can be achieved within disciplines, i.e. within the assessors’ scope of competence, but becomes problematic as soon as comparisons across disciplines are undertaken. Validity and legitimacy are considered high by researchers because peer review means having competent colleagues looking into the research. Customised evaluations are based on peer review and are thus considered as highly valid and legitimate by researchers but also have the problem of decreasing timeliness of information. They provide the richest information, which, however, is not comparable at all because assessment and recommendations are tailored to the assessed units.

This comparison of informational yields shows that no information system can provide university management with sufficient information. Only indicator-based systems and recent peer reviews provide timely information, and information can be either rich or comparable. Thus, the limitations to managerial capabilities of university management are aggravated by an inevitable lack of necessary information.

3.4 The inevitable response: management of research portfolios

In the three preceding sections, I argued that universities face increasingly strong expectations concerning research performance but are severely hindered in their attempts to meet these expectations by their limited organisational capabilities to shape research and the limited informational yield of evaluations. This combination of pressure and constraints reduces the options of university management to the use of three basic measures, namely changing the university’s structure, redistributing resources, and directing recruitment and promotion. None of these measures directly influences research content or performance. However, the university can increase the amount of research it believes to meet political expectations concerning content or performance, and reduce the amount of research that it considers to be at odds with these expectations or to be unimportant.

Thus, the combination of expectations and limitations initiates a practice of «managing research portfolios» at universities. Universities can acquire fields of research by recruiting academics from these fields and providing them with resources for research. They can strengthen fields already present by providing additional positions, reducing teaching loads, or providing resources. They can also let fields shrink or make them disappear by not replacing academics who leave the university, forcing academics to leave, or taking away resources. In other words, universities can invest in researchers and fields with a (proven) potential for meeting external expectations and withdraw investments from the others.13

13. As Laudel and Weyer (2014) show, universities can also withdraw investment purely for internal financial reasons.
Research portfolios can be managed according to a wide set of external expectations, the most prominent of which currently are expectations that universities

- build distinctive research profiles;
- conduct research in fields that are considered politically important; and
- improve their research performance.

Evaluations may play a role for all three purposes but are most central to the third. They describe how the university meets current expectations of performance and enable the identification of «strengths and weaknesses», i.e. fields in which the university performs well respectively less well. They also signal how performance is currently measured, thereby providing the means to assess the potential impact of «acquisitions» (fields to be built or researchers to be recruited) and «cuts».

«Portfolio management» can be clearly seen in countries with advanced new public management modes practices of higher education governance. Lucas (2006, 100-104) described how the department she observed decided to acquire biotechnology for its research profile because all biology departments with high scores in the research assessment exercise had biotechnology groups. The research assessment exercise has also created a «transfer market» for highly performing researchers, who would be recruited by universities in the hope of improving the rating of departments (Harley et al. 2004: 335). Laudel and Weyer (2014) describe several cases of portfolio management by Dutch universities in response to political expectations and financial problems. Australian universities were reported to respond to scarcity by fixed-term investments in «research centres» with the expectation that these centres were able to become «self-sufficient», i.e. to continue mostly or exclusively on grant funding (Gläser and Laudel 2007, 139).

In Germany, these strong cases of research portfolio management are very rare, if not impossible. A university’s research profile is still controlled by its professors, who compete for resources to strengthen their research areas but cannot be made redundant or completely stripped of resources. This is why fields may grow quickly but can be reduced only very slowly, and disappear only after all professors retired. Thus, university management has relatively few resources at its disposal and cannot «clear» its profile from unwanted fields quickly.

Nevertheless, there are clear signs that portfolio management is beginning and that universities’ capabilities to manage research portfolios are increasing. It can be observed in response to strong evaluation systems. The recommendations of customised evaluations to strengthen or abandon fields are often implemented (Meier and Schimank 2010). «Profile building», which is generally expected from German universities, is an explicit exercise of research portfolio management,
although commonly expected to happen only once. In the context of the EI, successful bids for Graduate Schools and Excellence Clusters led to the strengthening of fields. In some cases fields were also strengthened in anticipation of bids for the next round.

My discussion of German universities’ responses to evaluations also identified conditions under which the management of research portfolios is likely to turn into a permanent practice. These include

- increasing expectation and pressures on German universities, possibly by strong performance-based funding schemes;
- the full exploitation of decision rights by university management; and
- the institutionalisation of the redistribution of authority that can currently be observed in the context of customised evaluations.

An additional facilitating condition would be an increased discretion of university senior management over resources. None of these conditions is particularly unlikely to occur. They may emerge even within the current limitations set by a fragmented higher education system and the constitutionally guaranteed freedom of research and teaching.

4. Conclusions

Higher education reforms under the new public management paradigm rhetorically construct universities as corporate actors, increase their operational autonomy, and expect them to shape their research according to thematic and performance expectations of higher education policy. Since universities’ capabilities to shape their research are limited for principal reasons, performance-based management of research portfolios appears to be an inevitable point of convergence of management in predominantly state-funded universities. German universities appear to be in the first stages of the move towards this practice, and it remains to be seen whether they will go all the way.

As a conclusion to this article, I would like to discuss some possible consequences of portfolio management for knowledge production in the sciences, social sciences and humanities. A first consequence is that within and across universities, resources will be more strongly concentrated on fields that meet external expectations, and on academics whose work meets such expectations. These fields and academics will have secure support even in times of relative or absolute scarcity.

Second, performance-based management of research portfolios is likely to miss its target due to two intervening processes. Indicator-based performance evalua-
German universities on their way to performance-based management of research portfolios

tions can lead to goal displacement (Merton 1940), a mechanism that replaces the goal whose attainment indicators are supposed to measure by the goal of performing well on the measures. In the case of research evaluations, goal displacement means that the evaluated universities focus on increasing performance in indicators rather than increasing research performance. Furthermore, once university management has invested in certain research fields or researchers, it loses control of the content of research conducted by these researchers or fields. The researchers recruited by the university decide autonomously on research problems and approaches, and neither the directions of this research nor its performance can be predicted with certainty.

Third, since performance-based funding and other competitive evaluation systems create quasi-markets for research performance, they are also liable to quasi-market failure (Gläser 2007). In the case of competition for funding and political legitimacy, the quasi-market failure that is visible and has already been a cause for concern is the loss of epistemic diversity. In their construction of research profiles, universities are likely to invest in large and highly visible fields, fields that perform well in the indicators currently used for measuring research performance, or fields that are marked by science policy as being of current public interest and are therefore likely to attract external funding. This was already noted in Germany, where the German Association of University Rectors voiced its concern that profile-building activities that follow the same logic at all German universities may make small subjects disappear (HRK 2007). Laudel and Weyer (2014) observed that Dutch universities’ uniform responses to political signals contributed to the disappearance of one field and the stagnation of another.

Fourth, performance-based management of research portfolios might introduce short-termism in the funding of science if evaluations and political priorities shift too rapidly. A research base that supports advances in knowledge production is quite complex and includes state-of-the-art equipment, theoretical and methodological knowledge including tacit knowledge, positions at universities and attractive career paths, and national as well as international collaborative relationships. Such a research base can be dismantled much faster than it can be re-built. If performance criteria and political expectations shift at short intervals, and universities’ management of research portfolios follows them, the science system might lose

14. A particularly interesting and potentially consequential case of goal displacement is the internal mirroring of formulae and indicators applied in national performance-based funding. Australian universities used the indicators applied to them in performance-based funding schemes internally for distributing funding to faculties or schools in order to provide incentives for high performance in these indicators. The indicators designed for distributing resources between universities were thus used by universities to distribute funding between disciplines. Since publications and external funding depend on field-specific research practices (Laudel 2006; Laudel and Gläser 2006), the attempts to maximise performance on indicators disadvantaged certain disciplines (Gläser et al. 2010b).
fields that turn out to be essential for progress a few years later. In this process, academic careers and the attractiveness of science as a profession might suffer. Although many of these possible effects of performance-based management of research portfolios look detrimental, we should not forget that due to its decentralised and in many cases international nature, the science system is amazingly robust and adaptable. On the international level, adaptability means that scientific communities seemingly «shift around» research problems because attempts at problem solving emerge under those local institutional conditions that provide the best conditions for success\(^\text{15}\). Locally, we see this adaptation in scientific communities that manage to conduct long-term projects on successive three-year grants, as «bootlegging» of funding or as window-dressing in response to thematic expectations. A political conclusion that can be drawn from this article is that the role of organisations representing the interests of scientific communities is changing. As quasi-market failures indicate, the use of evaluations as policy and management tools may lead to aggregate effects that change the conditions of research for whole national scientific communities, and may threaten the development of fields. This is why interests of scientific communities must be more actively represented in decisions on institutional conditions for research. Professional associations in the sciences, social sciences and humanities appear to be best placed to represent macro-level interests. It seems important that they recognise this, and become politically more active.

References


---

\(^{15}\) This requires international institutional diversity, i.e. countries providing different institutional conditions. NPM as a worldwide trend has the potential to reduce this diversity.


Meier, F. and Schimank, U. (2010), «Mission Now Possible: Profile Building and Leader-


Musselin, Ch. (2007), «Are universities specific organisations?», in Krücken, G., Castor, Ch., Kosmützky, A. and Torka, M. (eds.), Towards a Multiversity?: Universities between Global Trends and National Traditions (Science Studies), Bielefeld, Transcript Verlag.


Wissenschaftsrat (2013), Empfehlungen zur Zukunft des Forschungsratings, Köln, Wissenschaftsrat.