

5

Informed Authority?

The Limited Use of Research Evaluation Systems for Managerial Control in Universities

Jochen Gläser, Stefan Lange, Grit Laudel, and Uwe Schimank

1. The New Information Needs of Universities

The university is a key authoritative agency in the public sciences that is facing increasing political demands to improve research performance. In many countries, national research evaluation systems (RES) have been institutionalized that hold universities accountable for their research by assessing its quality and feed back this information to universities through public comparative assessments of research quality and/or by differential funding based on quality assessments. In all cases, universities receive both a signal that a major authoritative agency—the state—expects high-quality research and information about the extent to which they fulfil this expectation. From the perspective of the universities, this amounts to increasing pressure to improve their research (Whitley and Gläser 2007).

At the same time, governance reforms in many countries have increased the autonomy of universities from the state and strengthened the position of their central administrators *vis-à-vis* academics conducting research. Comparative studies have shown that, although governance reforms occur at different rates in different countries, there is a distinctive trend towards stronger hierarchical management in universities, which is accompanied by a weakening of academic self-governance (Schimank 2005). In many countries this amounts to a paradigm shift in authority relationships, with authority concerning decisions on education and research

shifting from the state to universities (Lange and Schimank 2007; Paradeise *et al.* 2009).

As a result of these two developments, university managers experience both pressure to improve research performance and increasing formal opportunities to ‘manage’ research activities. However, effective management of research requires accurate information about research. This is not to say that no management decision can be made without such information. The sociology of organizations tells that rational management based on accurate information is by no means commonplace (Isenberg 1984). Universities are no exception here—after all, it was a university that provided the inspiration for the garbage-can model of organizational decision-making (Cohen *et al.* 1972).

Having said this, we must also note that acquiring and processing information is a crucial process for organizational control. In order to maintain and improve their operations, the managers of organizations constantly need information about both their internal activities and their environment. Universities are increasingly expected to improve their research, and their managers’ authority to do so is growing—but do they have the information needed?

This information is partly produced by the RES themselves. By assessing research quality, RES not only create incentives to improve research but also provide information that can be used by universities in that very process. This function of RES has so far been insufficiently appreciated, although it has surfaced in the distinction between summative and formative evaluations, in which the latter are described as being aimed at assisting units to improve their research performance (Geuna and Martin 2003: 278).¹ In political discussions of RES, their use for university managers has been mentioned in a critique of the British plans to replace peer review by an indicator-based assessment procedure. According to Bekhradnia (2008), the indicator-based procedure would deprive universities of the information they need to improve their research.

The informational yield of different RES for university management varies, as do the uses to which this information can be put. The aim of this chapter is to suggest how the different evaluation techniques used in RES affect their informational yield for the exercise of authority by

¹ Unfortunately, the distinction between formative and summative evaluations (those ‘making judgments about the performance of a unit by comparison with similar units’ (Geuna and Martin 2003: 278) suggests that summative evaluations neither have the purpose to support units’ improvement of their research nor do actually contribute to such an improvement. Our chapter will demonstrate that this suggestion is misleading.

university management. Our analysis is based on the comparison of five RES, which between them cover the major variations in the autonomy of universities, purposes of RES, evaluation procedures, and informational yield.

2. Analytical Framework

The Contingent Authority of University Management

As managers of organizations we might expect university managers to set goals, to create conditions for task performance, to monitor goal attainment, and to respond to the outcomes of this monitoring by changing conditions of work. However, in practice the authority of university management in most public science systems varies between ‘almost non-existent’ to ‘limited’, for three main reasons. First, even though research is one of the core activities of the research university, it is only loosely coupled with the university’s authoritative structure. Universities provide researchers with some material resources and an organizational frame that serves as interface between research and society. However, they are not the actor that produces knowledge even though they provide an important social context for research. Goal setting, choice of methods, collaboration, the acquisition of critical resources (information), integration of results, and quality control are all primarily governed by the scientific communities to which academics contribute. Most of these communities transgress organizational and, usually, national boundaries.

Owing to their specific decision structure, scientific communities can produce new knowledge regardless of the uncertainty inherent to that process, which includes uncertainty about the problem, the existence of a solution, the approach to the solution, the resources required for solving the problem, and the meaning of the solution (Gläser 2006; 2007). For the same reason, only senior academics working in the same field as their colleagues can exercise effective authority over research activities. The impossibility for university management in general of ‘managing’ problem definition and selection, evaluating research strategies and techniques, and assessing the intellectual value of results limits the extent to which universities can turn into strategic actors (Whitley 2008: 24–6)—by which they can move beyond the rhetorical construction of ‘actorhood’ that has been discussed by Weingart and Maasen (2007).

Beyond this principal limitation for university management, its authority and action capabilities depend on the way in which universities are formally institutionalized (Whitley 2008: 26–31). The authority of universities depends on their autonomy *vis-à-vis* the state and *vis-à-vis* their academics, and thus varies considerably between countries. Simplifying the distinction introduced by Whitley in this volume, the poles of the spectrum are the ‘hollow organization’ (the traditional German model) and the ‘employment organization’ (mostly established in the Anglophone countries). The former was in place in Germany, Austria, several Nordic countries, and Japan in its pure form until a decade or so ago. It is characterized by largely powerless universities, whose personnel matters, budgets, and resources are controlled by the state. The high autonomy of university professors, who were appointed and equipped by the state, prevented universities from interventions in matters of teaching and research. Although most countries featuring the traditional German university model have initiated reforms in order to increase the autonomy of universities, many of these limitations still exist. In contrast, the ‘employment organization’ model features much more formally autonomous universities which have considerable control of budgets, capital investment, personnel, and salaries, as well as their internal structure and the courses taught. In many such university systems, internal hierarchies are replacing more collegial forms of governance (Clark 1998).

A third limitation faced by all universities is a consequence of universities’ limited access to ‘their’ research. In order to ‘manage’ research effectively, universities would need information about the organization and direction of research. Obtaining such information is difficult because work processes are embedded in scientific communities rather than being governed by formal organizations, and are opaque to university management. Most information about the research process, the quality of conditions of work, and the quality of results is generated in the course of conducting research. However, this information is largely internal to the work process, idiosyncratic, and understandable only to other academics—in most cases only to those from the same field.

Thus, the increasing pressure for universities to produce high-quality and societal-relevant research highlights a major problem for them: they need to be able to affect research processes but cannot generate the information to do so. This problem is aggravated in some countries by the limited formal authority and capabilities of university management, but exists in all countries because of the specific way in which the production of scientific knowledge is embedded in universities.

Research Evaluation Systems as Sources of Information

The emerging RES not only create incentives for universities, and thus their demand for information about research, but are also an important source of such information. RES are created for three purposes: (1) to inform universities about the quality of research, (2) to provide incentives for improving research, and (in many but not all cases) (3) to improve research by redistributing resources from weak to strong performers. The information they generate can be used for university management decisions. However, for most RES this is a side effect rather than their main purpose.

RES differ widely in the information they use as input, assessment procedures, and their output of information about research quality (see Gläser 2008 for an overview). The most important distinction is that between RES based on peer review and indicator-based RES. The former employ an assessment of research that is conducted by colleagues working in the same field as the academics whose research is evaluated, while the latter use quantitative data about research and its outputs. The two kinds of RES are easy to tell apart because even though peer-review processes often also use quantitative information on research performance, indicator-based systems are characterized by the absence of peer judgements, as will become clear from our case studies.

RES that are based on peer review take the view that a researcher's peers are the only people who can judge the quality of research by analysing its content.² Assessment of content is achieved by requesting the submission of research outputs (mostly publications) which are read by the assessors. In most cases, contextual information including statistical information on staff, resources, external funding, and outputs is used as additional input to the peer-review process. This information is provided for 'units of assessment', which might be constituted by all research of a university in a certain field, an organizational sub-unit of the university such as an institute or research group, or an individual academic. Apart from reading submitted research outputs, peer review procedures may include interactions with the evaluated units such as interviews of scientists or site visits. The outcomes of peer review-based evaluations are commonly provided as formalized ratings of the 'units of assessment' in one or more dimensions,

² The first RES ever—those of the Netherlands (introduced in 1983) and the UK (1986)—utilized peer review as the assessment method. The British RES—the Research Assessment Exercise (RAE)—has been copied (with varying degrees of match) by Hong Kong (first review in 1993), New Zealand (2003), Italy (2004), and Australia (planned for 2008 but abandoned by the new government that was elected late in 2007).

which make the assessments comparable across universities and disciplines. The ratings are usually accompanied by a short text that explains the judgement.

In contrast, indicator-based RES rely on quantifiable properties of the input, process, or output of research as proxies for research quality. Countries currently using indicator-based RES include Australia, Germany (at the level of federal states), Ireland, Norway, and Belgium (the Flemish region). The most frequently used indicators are

- the amount of external funding, which is seen as an indicator of quality because winning external grants usually depends on the project passing a peer review;
- numbers of publications, which are sometimes weighted by the quality of publication types (such as ‘international, peer-reviewed journals’); and
- numbers of Ph.D. graduations.

These indicators obviously measure the amount rather than the quality of research. Their use can be explained by two of their properties: the opportunity to easily collect the information about them, and their applicability to most fields of research. The former property makes indicator-based RES a cheap alternative to the expensive peer reviews, while the latter is essential for applying homogeneous RES to universities covering all fields of learning. More sophisticated and more valid indicators such as those based on citations cannot be currently applied to all fields, and require expensive methods of data collection and data cleansing (van Raan 1996: 403, 405).

Variables

Our analysis compares the informational yield of each of the RES and the use universities make of this information. Since university management is by and large unable to prescribe inputs, procedures, or outcomes of the research process, its opportunities to influence research are limited to shaping some of its basic conditions and to managing some of the behaviour of its actors—the researchers. These opportunities arise mainly in three kinds of decision. *Internal resource distribution* allocates the university’s block funding to sub-units of the university (faculties, schools, and individual academics) in order to maintain the infrastructure for teaching and research tasks. A second major task of the university management is internal (*re*)structuring of the university. Much of this restructuring occurs because of changes in student demand (which still is the major basis of

funding for universities). However, RES and in particular the performance-based allocations of block grants create strong incentives for universities to change their structures of research as well. A third kind of decisions aimed at improving research performance concerns the *management of individual performance*. This includes hiring decisions, the use of probationary periods, recommendations to academics about their conduct of research, decisions on tenure and promotion, and other uses of incentives.³

While these decisions mostly utilize different information, they share the requirement for information about research performance and about favourable conditions for increasing performance. This kind of information is—to varying degrees—provided by RES. In order to compare the informational yield of RES, we use five dimensions: (1) richness, (2) timeliness, (3) validity, (4) legitimacy, and (5) comparability. These are, as will be evident below, partly derived from the organizational sociology literature on management information processing. The evaluation techniques can be ranked according to their yield in each dimension (Figure 5.1). With the exception of timeliness, the rankings must be based on relative rather than absolute scales because no objective ‘yardstick’ exists against which richness, validity, legitimacy, and comparability of information could be assessed.

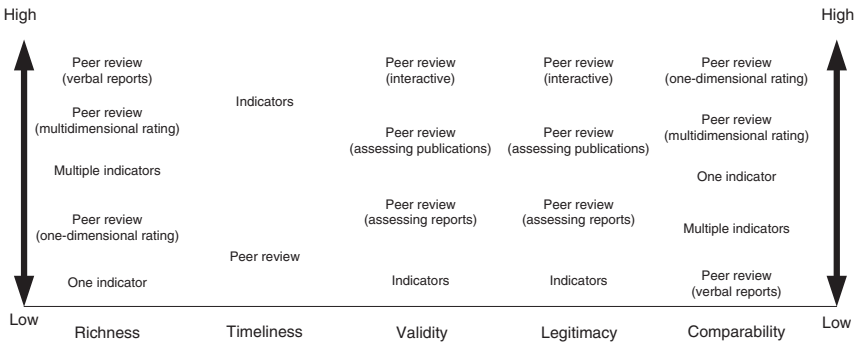


Figure 5.1. Ranking of information properties in the five dimensions

³ These decisions are often only analytically distinguishable. For example, the hiring of a senior academic is likely to simultaneously have resource, structural, and performance management aspects. As we have discussed in the previous section, the extent to which university management has the formal authority to make these decisions varies between countries.

RICHNESS

The richness of information provided by different communication channels is a central concern of the organizational sociology literature. While it is possible to distinguish between the amount of information and the richness of the channels through which it is provided, this is not necessary in our particular analysis. We define the richness of information as the number of different aspects about which information is provided, thereby synthesizing the concepts 'amount of information' and 'richness of communication channels' that are sometimes discussed separately in organizational sociology (Daft and Macintosh 1981: 210; Daft and Lengel 1984: 195–8; 1986).

The ranking of RES according to the richness of information they supply can draw on comparisons of communication media richness (Daft and Lengel 1984: 195–8). Peer reviews have the potential to provide very rich information. In particular, peer reviews—and only peer reviews—can provide information about a unit's potential for future high-quality research. Given the right input (such as research programmes) and procedure (especially site visits), assessors are able to draw inferences about the capabilities of researchers, the suitability of their equipment, and the fit between these conditions and the research programme. However, the design of peer reviews in RES often reduces the richness of information. We rank peer reviews that provide detailed verbal reports higher than those who provide ratings or rankings in several dimensions. A set of indicators that provides information about several aspects of research performance is less rich because the numerical values are lacking the link to research performance, which has to be established by *ex-post* interpretations. Among RES using peer-review-based ratings or sets of indicators, the richness of information decreases as the number of dimensions (indicators) declines.

TIMELINESS

The timeliness of information has received far less attention than the amount of information or media richness.⁴ Apparently, the literature implicitly assumes that management always can acquire the necessary information about its technologies, and has to cope with whatever is the timeliness of information provided in the organization's environment. The case of research organizations that obtain information about their

⁴ But see Choudhury and Sampler (1997) on the 'time specificity' of information.

technology by collaborating with their environment has not yet been taken into account.

In one sense, the timeliness of information can be considered to be an aspect of its validity—the older the information, the less likely it is to adequately depict current research quality. This aspect of timeliness depends on the evaluation cycles of RES, which vary considerably. The longer the period between two assessments, the more outdated the information used in current management decisions becomes.

However, timeliness has a dimension that merits separate treatment: namely, the limitation of timeliness that is produced by the input in evaluations. The material that is evaluated always represents past research. Peer reviews that rely on the examination of submitted publications assess the quality of research whose results were ‘written up’, reviewed, possibly revised, and thus finally published after months or years. Indicators such as numbers of publications or of Ph.D. completions have the same problem.⁵

VALIDITY

The validity of research evaluation techniques is an issue that is not well understood, and is highly contested and difficult to separate from issues of legitimacy. The major fault line in the discussion about evaluation techniques runs between an argument for peer review as the only way to evaluate the content of research directly and arguments pointing to peer review’s opacity and susceptibility to bias, favouring quantitative indicators as ‘objective’ methods that do not depend on personal judgements of few assessors. In order to be able to include validity as a dimension of informational yield at all, we disregard problems of bias (which reduces validity but is a random rather than systematic property) and opacity (which hinders the measurement of validity rather than reducing it). The validity of management information has not been discussed in the literature except for occasional references to information accuracy (for example, Saunders and Jones 1990: 38). However, it can be linked to information accuracy and one

⁵ If citation-based indicators were used, the situation would be even worse because for citations to occur some time must pass for a publication to be read, used in other research, and be cited in subsequent publications. The only exception is the indicator ‘external funding’, which relies on *ex-ante* evaluations and thus refers to research that is being undertaken when the data are collected. This relatively high timeliness is achieved by this indicator’s reliance on prospective peer reviews. Contrary to all other evaluation methods and indicators, the indicator ‘external funding’ reflects an evaluation of plans for the research rather than accomplished research.

of the central themes of management information processing: namely, the equivocality of information. Equivocality can be defined as 'the multiplicity of meaning conveyed by information about organizational activities' (Daft and Macintosh 1981: 211, see also Weick 1979). Both unknown validity and low validity of information about research performance increase the information's equivocality.

If we treat the evaluation embedded in RES as an interrogation of the research about its quality, the ranking of media richness (Daft and Lengel 1984: 195–8) can once more be used to assess the RES. RES based on peer review generally rank higher than those based on indicators, because the former use at least written verbal reports, while the latter just use numbers. Among the peer-review-based systems, those including personal communication between evaluators and the evaluated (interviews or site visits) rank highest because they use a medium of higher richness. The rather simple quantitative indicators currently used in RES partly reflect volume rather than quality (in particular Ph.D. completions and numbers of publications) and in any case reflect factors other than quality (for publications, see Butler 2004; for external funding, see Laudel 2006). Their validity is generally considered to be low (Jauch and Glueck 1975: 70–3; Phillimore 1989; Wood 1989).

LEGITIMACY

Legitimacy is related to the norms and culture of a social context, which is why the legitimacy of the information produced by an RES varies between the contexts of science policy, university management, and academics. For example, quantitative indicators often carry a higher legitimacy in science policy circles because they are considered to be transparent and unbiased (see above, on validity).

For our investigation of the informational yield of RES for university management we focus on the legitimization of management decisions in the scientific community. Here, the legitimacy of information about research quality is strongly tied to the dominant view of the validity of that information, which is that peer reviews produce the most valid information on research quality, and that measures that increase the validity of peer reviews also increase their legitimacy. Indicator-based information on research quality is generally considered a less legitimate basis for decisions than peer review, although some disciplines (especially the biomedical sciences) have begun to rely on citation-based indicators (which are not used in RES).

COMPARABILITY

Since many decisions in universities affect more than one field of research, information on research quality and conditions for research also needs to be comparable. Comparability is highest when peer reviews are designed to produce ratings or rankings. Quantitative indicators are less well suited for comparisons because of their dependence on unit size and field (for example, field-specific publication practices and dependence on external funding). Verbal reports on just one discipline, which are the outcome of some peer-review procedures, have the lowest comparability.

3. Informational Yield of RES and Internal Use of that Information by Universities

Five Case Studies

For our empirical analysis we have selected cases which exhibit different characteristics in terms of RES and autonomy of universities (Table 5.1). Both quantitative indicators and peer-review-based RES are represented in our sample. Australia has the oldest indicator-based system for funding university research, which was introduced in the second half of the 1990s. In Germany, many federal states have begun to introduce indicator-based systems over the last decade. We investigated universities in one of these states (which, in order to protect the privacy of interviewees, we cannot name). The countries featuring peer-review-based RES include those with the two oldest RES—the Netherlands and UK—and the German state of Lower Saxony, which has institutionalized a peer review-based RES.

The second dimension that is important to our analysis is the autonomy of universities, which influences the latter’s opportunities to act on the information they receive from RES. German universities still have a relatively low autonomy, while the autonomy of Australian, British, and Dutch

Table 5.1. Cases included in the analysis

	Autonomy of universities	
	Low	High
Evaluation technique		
peer review	Germany, Lower Saxony	UK, Netherlands
indicators	Germany, State X	Australia

universities is high.⁶ Table 5.1 demonstrates that all four cells of the cross-tabulation of evaluation technique and autonomy of universities are included. The UK and the Netherlands provide an interesting additional contrast because the Netherlands abandoned the link between evaluation and funding after the first round of evaluations, and have transferred responsibility for the evaluations to the universities.

Peer Review-Based RES and Universities with Low Autonomy: Lower Saxony⁷

In 1997, an 'Academic Advisory Council' (AAC) was established in the German federal state of Lower Saxony, and was tasked with the organization of the evaluation of all of Lower Saxony's university research. The evaluations are conducted as discipline-oriented peer reviews. Units of assessment are departments or institutes within universities and 'research units', which are self-defined by researchers according to local and disciplinary conditions. A 'research unit' can range from a team of scientists (such as in the natural sciences) to an individual chair (such as in the humanities). The usual evaluation procedure is as follows. A short framework paper is provided by the AAC to the evaluated disciplines within the universities to help them prepare a report on the last five years of research activity and future planning. Universities are then visited by the group of evaluators; approximately six professors from the evaluated discipline (but from other German federal states or from foreign universities). These evaluators talk to the university president, the respective dean, each professor of the discipline, some members of scientific staff, and some doctoral students, and discuss their findings. Based on these discussions, a draft report about the discipline and its relative performance at all Lower Saxony's universities is written by the evaluators and edited by the AAC's officer in charge. The evaluated units and individuals are then asked for their comments, which reach the AAC via the president of the university. On this basis, the final

⁶ There are significant variations within both groups. The autonomy of many German universities has increased since the time of our investigation (2005–6), albeit in an uneven pace that depends on the higher education legislation in the sixteen federal states. The government's control of universities is significantly stronger in the Netherlands than in the UK or Australia. However, the basic distinction between universities that have full control of their recruitment, internal structure, internal allocation of funds, and human resources management (universities in Australia, UK, Netherlands) and those that do not have this control (German universities) still holds.

⁷ This section is based on two empirical studies of the evaluation process and responses by universities (Schiene and Schimank 2007; Chapter 7). See Chapter 7 for a discussion of the impact of this evaluation process on authority relations.

report ('assessors' report), which contains evaluations and recommendations, is written and submitted by the evaluators to the AAC. The AAC discusses the report and its recommendations. The report is published with the exception of the evaluations of individuals, which are given to the individuals and their university presidents, with a complete copy sent to the Ministry. There are several follow-ups on the evaluation reports, including an intermediate report after three years.

The evaluation reports include detailed verbal judgements of the research performance and potential for future performance of a discipline in each of the universities of Lower Saxony. The recommendations are also quite specific, and deeply intrude into the structural and resource allocation decisions of universities. The major kinds of recommendation can be listed as follows:

- (1) establishment of new professorships, rededication of vacant professorships within the discipline, elimination of vacant professorships or their transfer to a different discipline;
- (2) participation of external peers in the recruitment commissions for vacant or new professorships;
- (3) additional scientific staff for professorships;
- (4) reduction of permanent scientific staff in favour of temporary employment contracts with younger scientists;
- (5) study programmes for postgraduates;
- (6) additional financial means from the government;
- (7) a more performance-oriented allocation of block grants within the university or faculty;
- (8) increased acquisition of project grants or research contracts;
- (9) infrastructural improvements of buildings, libraries, laboratories; and
- (10) intensification of internal and external coordination and cooperation.

The Ministry asked universities to implement these recommendations. The implementation is under way in most cases, although in some it had to be started against the will of the affected faculties or institutes, whose self-perception was completely different from the evaluation. Recommendations were also turned into strategic goals that became part of 'performance agreements' between universities and the Ministry. Even though the evaluated units disagreed with the assessments in some cases, the whole evaluation procedure was perceived as legitimate because it was conducted under the authority of the scientific communities. As a result, the position of the university leadership was strengthened by the double support from the

Ministry demanding the implementation of recommendations and the scientific community legitimizing them.

The paradoxical consequence of the whole evaluation process was that the autonomy, action capabilities, and authority of the university management were strengthened by a procedure that deeply intruded into university matters. The major reason for this paradox is the double limitation of the autonomy of the German university. German universities are not only dependent on the state, which still sets tight frameworks for most of the essential management decisions such as employment contracts and resource allocation, but are also unable to intervene in decisions of their university professors, who are appointed by the state as public servants and have a guaranteed personal budget. If German university professors decide that they do not want change, it is very difficult for a university leadership to achieve it—the more so when the professors of a faculty collectively decide that they do not want change.

In this situation, a demand for change by the Ministry that is legitimized by the scientific community significantly enhances the authority and capabilities of the university management. Assuming that the university leadership wanted to improve research and to build externally recognizable research profiles, and that it would have needed to conduct peer reviews to support these actions, the enforced recommendations of the evaluation procedure could solve most of the problems created by the university management's limited action capabilities.

While the legitimate, detailed, and intrusive recommendations of the peer review in Lower Saxony might seem to be a good solution under the circumstances, several problems need to be mentioned. First, the evaluations and the demand by the Ministry to implement the recommendations were accompanied by severe cuts in the block grants, which not only made the implementation of many recommended changes impossible but also undermined trust in the whole evaluation process. Secondly, the funding cuts occurred in a situation of still increasing teaching loads, which made the implementation of changes for the promotion of research even more difficult. Thirdly, the evaluation procedure effectively disaggregated the universities. Disciplines were evaluated at different times and by different groups of assessors, none of which took the whole university into account beyond the contributions to the process made by the university leadership. Therefore, it is at least open to question whether a more holistic look at the university would have led to a different consideration of context and local knowledge, and thus to different recommendations. As it was, each disciplinary panel by and large defended its discipline in each university. Only

one type of recommendation—to transfer a vacant professorship to a different discipline—does not strengthen the evaluated discipline, but another one. Being ‘altruistic’ from the point of view of the evaluators as disciplinary peers, such a recommendation is of course only given in ‘hopeless cases’; that is, very rarely.

*Peer Review-Based Devolved RES and Highly Autonomous Universities: The Netherlands*⁸

In the Netherlands, evaluations based on peer review are conducted according to a standard evaluation protocol. The previous standard evaluation protocol (valid to 2002) organized disciplinary evaluations at the national level; that is, all disciplinary units (programmes) in a discipline were evaluated at the same time.⁹ With the new standard evaluation protocol (from 2003), evaluations are more devolved. Under the current protocol, the universities themselves organize the evaluation of their research. A self-evaluation is prescribed after three years and an external peer review after six years. The procedures for evaluating the units remained the same. The units of assessment—‘programmes’ (research groups) or university institutes—submit self-evaluation reports and lists of publications to the peer-review committees. These committees form their assessment on the basis of the reports, of an examination of the submitted publications, of interviews with the programme leaders, and (in some cases, particularly in the laboratory sciences) of site visits. The programmes are evaluated in four dimensions (evaluation aspects): scientific quality, scientific productivity, scientific relevance, and long-term viability. These aspects are translated into specific evaluation criteria for each discipline by the peer-review committees. Each programme is rated on a five-point scale (excellent/good/satisfactory/unsatisfactory/bad) in each aspect. These ratings are briefly justified (with half a page of text).

The major change that was introduced with the new standard evaluation protocol from 2003 concerned the comparability of information from the RES. This comparability was already limited with the previous evaluation

⁸ This section is based on publications on the Dutch RES and university management by Westerheijden (1997); Jongbloed and Van der Meulen (2006); CPB and CHEPS (2001), VSNU *et al.* (2003); and Meulen (2007). We are grateful to Pleun van Arensbergen, Rathenau Institute Den Haag, for her support of the Dutch case study.

⁹ The definition of ‘programmes’ goes back to the first evaluation exercise that was initiated in 1979 and conducted in the early 1980s. In this first round of peer reviews, the university had to define ‘research programmes’ whose funding was conditional on an *ex-ante* peer review. Since then, the ‘programme’ has been the basic unit of the Dutch RES (Meulen 2007).

procedure. Since the evaluations were conducted for each discipline at a different time, the universities never had information of equal timeliness on all of their research. With the new standard evaluation protocol, evaluations are initiated and organized by universities for their research in a certain field. With the exception of a few cases where universities agreed to conduct a joint evaluation for all research in one discipline, most of the evaluations since 2003 produced information on just one unit of assessment, which could only be compared to other evaluation results from different points in time.

Nevertheless, university managers welcome the information provided by the RES as legitimizing the differential treatment of research groups and as a necessary input for this new approach. The actual use of the information has varied widely between universities. One university translated the results into a quantitative measure for the allocation of 10 per cent of its research budget. The formula developed by that university completely disregarded two of the four dimensions (scientific relevance and long-term viability). The other two become synthesized in one weighted measure (75 per cent quality, 25 per cent productivity). The university then ranked the units from all universities in each discipline, and determined the relative position of the University of Tilburg's unit in each discipline by dividing its absolute rank by the numbers of units. The resulting figure was used to compare units from different disciplines and to redistribute resources between them.

This straightforward and simplifying utilization of evaluation results is an exception. In other universities, evaluation results have indirect financial consequences because they are taken into account in budget negotiations between research groups and the faculty. Furthermore, evaluation results are considered when the directions of research and research strategies are developed. Thus, evaluation results are one of several inputs to negotiations about research conditions, such as research budgets or reduced teaching loads.

The flexibility of this approach is illustrated by the case of yet another university. Instead of simply rewarding good and punishing bad research groups, university management took the importance of a field to its discipline into account when decisions about the future of research groups were made. Therefore, high scores in the evaluations were no guarantee of further prosperity. When financial cuts had to be made, even highly evaluated groups were closed in the ensuing reorganization. On the other hand, a group that received a low score but was considered to be important for the

discipline by the faculty received extra funding as part of attempts to strengthen the group.

The general perception of researchers is that the results of evaluations do not have severe consequences. Good scores provide a certain protection from administrative intervention (reorganization) and a relative advantage in budget negotiations within the university. Bad scores may lessen this protection but have no automatic consequences either. These findings indicate that the university management at Dutch universities is quite active in terms of restructuring the universities, and that the evaluation scores partly help redefine the targets for these measures. It must be noted, however, that these findings are still preliminary and lack both detail and reliability. The responses by Dutch universities appear to be quite complex and merit further detailed investigation.

Centralized Peer Review and Autonomous Universities: United Kingdom¹⁰

The British RES was introduced in 1985 as a 'Research Selectivity Exercise' as a response to growing concerns that the quality of the British research base could not be maintained in its entirety, especially after severe cuts in university block grants in the early 1980s (Chapters 2 and 8). For each unit of assessment, universities submit up to four 'research outputs' (mostly publications) of every research active academic and contextual information including data on external funding, prizes and awards, and graduate students. The core of the evaluation procedure is an assessment of the submitted 'research outputs' by the members of the assessment panel. Examination includes reading a significant proportion of the submitted publications. Based upon the examination of publications and the analysis of contextual information, the panel arrives at a judgement about the quality of the research of a unit of assessment. In 2001, assessments took the form of a rating between 5* (international excellence in more than half of the publications, national excellence in the other publications) and 1 (no national excellence in any of the publications). In 2008, the format of the assessment results changed. A slightly changed rating—from 4* ('world-

¹⁰ This case study is based on the literature on the RAE including commissioned reviews and reports (Roberts 2003; evidence 2005, 2006), news reports (for example, Curtis 2002; Johnston and Farrar 2003), sociological analyses (for example, Morris 2002; Lucas 2006), reports from academics from various disciplines (for example, Dainty *et al.* 1999), and internal documents of British universities (for example, Queen Mary University 2003; Oxford University 2005; Northumbria University 2007). See also Chapters 2, 8 and 9.

Reorganizing Research Organizations

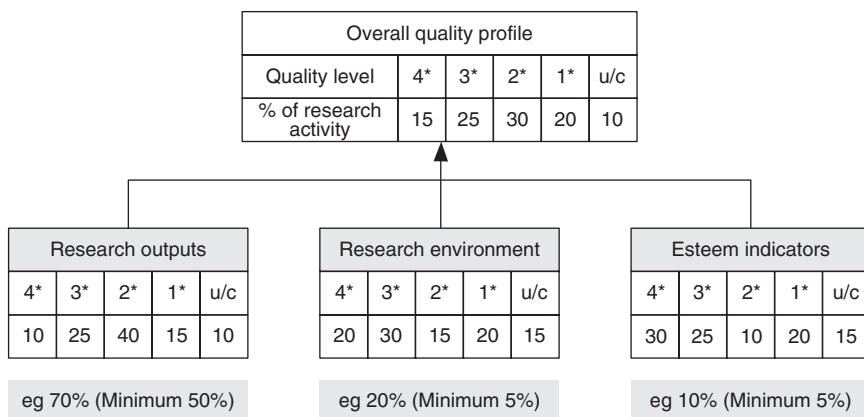


Figure 5.2. The construction of a unit's quality profile in the 2008 RAE (source: HEFCE, 2008: 99)

leading in terms of originality, significance and rigour') to 'unclassified'—was kept. However, instead of one amalgamated rating the panels issued 'quality profiles' that describe the distribution of a unit's research outputs, esteem indicators, and research environment across the quality categories (Figure 5.2). Thus, rather than receiving a '4*' or '4' for a unit of analysis, the university received a statement saying that 30 per cent of a unit's research is at the 4* level, 50 per cent at the 4 level, and so forth. This change in procedure avoided the 'cliff edge' effect produced by the enormous consequences of the small quality differences at the boundaries of rating levels. At the same time, universities received more complex information about the performance of their units of assessment.

Since its inception, the RAE's outcomes have informed the distribution of block grants for university research. Although on average the block grants for research constituted 7.5 per cent of all universities' income in the financial year 2001–02, they varied between 0 and more than 23 per cent.¹¹ The variation between universities is enormous. Some universities obtain half of their funding from the research block grants, while many others do not receive anything.

¹¹ It is difficult to obtain comparable data on the income of UK universities. Our estimates are based on data from the websites of the UK's Higher Education Statistics Agency (<http://www.hesa.ac.uk/index.php/content/view/full/807/251/>, accessed Oct. 2009) and from HEFCE (2004).

Given the actual or potential importance of the research block funding for universities it is not surprising that universities have responded strongly to the RAE.¹² Observers agree that the RAE has moved research into the focus of university management and turned research from an ‘unmanaged’ into a ‘managed’ activity (Lucas 2006: 73). This management of research performance has two targets, which can be analytically separated.

A first target of university management is to decide who is submitted in the RAE. Being in full control of the information submitted to evaluation panels, universities try to improve their RAE grades by shaping the material submitted to the evaluation. Careful internal evaluations and ‘dry runs’ of the RAE are conducted in order to decide on units of assessments to submit and to select the academics who should be submitted as research active. The trade-off faced by universities is that submitting a lower number of academics as ‘research active’ may improve the grade but, once a grade has been achieved, the amount of funding depends on the number of academics submitted. A second, much-discussed strategy of university management is head-hunting—the search for outstanding academics whose high-quality output is likely to improve a university’s grade when added to the submission. Universities also strategically change their structure by closing unsuccessful departments, reshaping and relabelling their research in order to submit it to a different unit of analysis, and combining stronger and weaker units in order to utilize a halo effect of the former.

These three strategies do not only target the evaluation process but are part of the more fundamental attempts of university management to strategically shape the content and quality of research. The strategies applied for that purpose include, foremost, the allocation of recurrent funding. There is a general tendency for universities to follow the results of the RAE—to allocate funding ‘as earned’. However, universities also create strategic funds that are used to strengthen weaker units and enable an improvement of their grades. An important aspect of the distribution of resources is the redistribution of time for research. Successful researchers whose output is important for achieving high grades are supported by a reduction of their teaching and administration tasks, which are assigned to

¹² There are indications for systematic differences between management approaches of research-intensive and other universities. For example, research-intensive universities submit most of their academics (more than 90 per cent) as research active, thus having no need for the strategic selection of academics to submit, and little opportunity to redistribute teaching and administrative loads. Unfortunately, there is not enough systematic research on university responses to the RAE for these variations to be reliably established, and their impact on research and teaching in universities to be assessed.

their colleagues who are not deemed good enough to achieve the aimed-for grade, and are therefore not categorized as 'research active'.

Finally, British universities apply procedures for managing the research performance of their academics. Research performance is monitored by using quantitative indicators, among which publications play an important role because they are the core of RAE submissions. Research performance is the subject of annual performance talks between academics and their supervisors, and plays an important role in decisions about appointing and promoting academics. In these decisions, research performance appears to have gained a higher weight than teaching performance (Parker 2008).

*Indicator-Based RES and Universities with High Autonomy: Australia*¹³

Australia has used an indicator-based system of allocating research block grants since the mid-1990s. As it currently stands, the system distributes about 7.9 per cent of the total income of universities according to external competitive funding (indicator weighted at 54.8 per cent), numbers of Masters and Ph.D. completions (29.1 per cent), numbers of publications (8.4 per cent) and current Masters and Ph.D. students load (7.7 per cent) (own calculations based on DEST 2007). The distribution is a competitive zero-sum game in which universities must participate, and the share of the research block grants in their income varies between 0 and more than 15 per cent. All universities have mirrored the external funding formulae in the internal distribution of research funding to faculties and in some cases from faculties to schools. However, while using the indicators from the external formula, some universities have given different weightings to the indicators. This was deemed necessary because within universities the indicators inform a distribution of resources between disciplines. Simply copying the external weightings of indicators would disadvantage the social sciences and humanities because of their systematically lower external funding. This is why this indicator's weight is reduced and the weight of either research student completions or of numbers of publications is increased.

These internal resource allocation systems are of little direct consequence for research projects because the money allocated is largely used to pay part

¹³ This case study is based on an empirical investigation of the Australian system of indicator-based research block funding of universities (Gläser and Laudel 2007), which is part of a comparative project including Australia and Germany (Gläser *et al.* 2008; see also Ch. 10).

of the academics' salaries and to maintain the basic infrastructure. In most cases there is no recurrent funding of research, regardless of the proportion of a university's income provided by research block grants. Instead, internal grant schemes have been set up by most universities. The grants are rather small and usually limited to one year. They have the function of recurrent funding insofar as they are intended to be used as preparation to acquire external funding, even though they are the sole source of funding for many researchers.

The major effects of the formula-based system concern the attempts of the universities to maximize their income by structural changes, targeted investments, and (to a lesser extent) individual performance management. Structural changes and targeted investments are mainly focused on the single most important indicator used in the funding formula: the amount of external funding. The universities attempt to create research units that are likely to acquire external funding. A characteristic way of doing this is to provide seed funding for 'centres' that is used to buy specialized research equipment or to reduce teaching loads of the leading academics involved. This funding is provided temporarily with the aim to create a centre that can exist on the basis of its external funding after some time (usually three to five years). Since this is rarely possible, it is quite common that the centres disappear after the university withdraws its funding. Apart from these approaches, which are purely quantitative in the sense that they are focused on external grant applications, there are only very few attempts to use detailed information on research quality in internal management decisions. Only two of the seven universities in our sample used peer review-based evaluations of their sub-units (schools or centres). Others applied the quantitative indicators in a rather superficial manner, or had no internal evaluation policies in place at all.

While research evaluation of sub-units played a minor role in most universities, procedures of individual performance management were in place in all of them. At the time of our investigation, the annual performance appraisals of academic staff were inconsequential because they affected neither incremental pay rises nor the distribution of work loads. More thorough performance appraisals were emerging during 2005 and 2006 because the federal government had made additional funding available for universities that implemented them.

The application of the quantitative indicators in performance evaluation schemes across all disciplines caused some problems because they are not equally applicable to all fields. In one university, the classification of academics who did not perform in two out of the three indicators external

grants, publications, and Ph.D. supervision as 'research inactive' caused a protest by the department of mathematics. Here, some internationally renowned mathematicians were suddenly classified as research inactive despite their many important publications, because they had no external funding and no Ph.D. students to supervise. We also detected some perverse effects of the indicator-based individual performance assessments, such as academics applying for research grants they did not need for their research or supervising Ph.D. students outside their core area of expertise.

The assessment of individual research performance has played a significant role in decisions on tenure (which took place in one university) and in decisions on promotions in all universities. The major approach to individual performance evaluation for promotions included the performance indicators used in the funding formula but was not restricted to them. Academics who applied for promotion had to submit information on external grants, supervision of Ph.D. students, and numbers of publications. These data were often looked at in context, and supplemented by other data that supported the case for promotion. Universities often use external peer reviews for decisions on promotion—at least where higher levels (associate professor and professor) were concerned.

Indicator-Based RES and Universities with Low Autonomy: Germany—State X¹⁴

The case of the German universities in federal state X—we call them university A and B—is both special and instructive because the universities in question had limited autonomy and therefore could not use information on research quality, regardless of its content and form. The limitations of autonomy were the same as already described in the case of Lower Saxony: namely, the subordination of universities to the state and the autonomy of university professors.

The universities received a part of their block grant according to a formula that includes indicators of research performance: namely, external funding (1.7 per cent of the block grant) and completed Ph.D.s (0.4 per cent). They thus had an information base for assessing the contributions by their faculties, institutes, and academics to their income. However, this information was not used by the university management for steering or managing

¹⁴ This case study is based on the investigation of German universities in the collaborative German–Australian project. For a detailed report on one of the German cases, see Lange (2007).

purposes because there were no legitimate grounds for doing so.¹⁵ Therefore, management activities which take into account the quality of research did so informally. 'The university leadership knows its good performers and treats them accordingly', as the head of administration of university B expressed it.

The *internal funding* of faculties by the university had not changed since the introduction of the funding formula. Funds were still allocated according to a system that was kept secret. The funding of faculties needed to honour the agreements between the state and each individual professor, which guarantee basic supplies consisting of posts for research and teaching associates and assistants, equipment, consumables, and travel.

One faculty in university A and the university leadership in university B had begun to implement performance-based funding schemes that redistributed a small proportion of the recurrent funding.¹⁶ The only instrument at university level was introduced by university B, which rewarded external grant acquisition according to sources of funding. Professors who were successful with their proposals for large competitive grants from the Deutsche Forschungsgemeinschaft (Germany's most important funding agency for university research) received a reward that amounts to 5 per cent of the external grant in the previous year. Successful proposals for graduate schools and funds from selected private foundations could be rewarded with up to 2.5 per cent, and grants from the Federal Ministry for Education and Research received a reward of 1 per cent on top of the grant budget.

In the faculty of social sciences at university A, some money was taken away from professors and turned into small awards for teaching or research initiatives. The rewards for research performance varied between 750 € for presenting a paper at an internationally recognized conference and 8,000 € for leading a successful bid for a large collaborative research grant. Most interesting in the German context is the attempt to steer the publication behaviour in the faculty by rewarding publication in an international top journal with 6,000 €, in a high-ranking journal with 4,000 €, in a good journal with 2,000 €, and in an applied journal with 1,000 €.

The major new instrument for performance-based funding in universities was not implemented in both universities until 2007. A change in higher

¹⁵ There are some exceptions. A natural science institute organized peer-review evaluations on its own to boost its international visibility and reputation, while neither the university nor the faculty had established any evaluation procedures.

¹⁶ In both universities the faculties for medical sciences led the way by using internal resource allocation regimes that rewarded both external funding and publication behaviour.

education legislation introduced in 1998 made it possible for the university management to make the basic supplies of a newly appointed university professor subject to a renewal every five years. The renewal shall depend on the professor's performance in teaching and research. However, only a few of the professors appointed after 1998 in universities A and B were aware of such a performance evaluation procedure or the performance criteria for this evaluation.

The universities also attempted to build their research profiles by creating research centres. The purpose of these *structural changes* was to create 'critical mass' or simply to increase the visibility of a certain research area. They were not aimed at increasing research performance, be it measured by the indicators or by any other means. The new centres were by and large administrative layers added to the traditional structure of the professoriate, and had little or no resources of their own. According to our interviews, they did not much influence the content of research. In the natural sciences they were deemed to be useful by professors for the purpose of pooling staff, laboratory equipment, and other resources. In the humanities they had not yet had any effect on research.

Apart from the few attempts at performance-based funding of professors mentioned above, no system of *individual performance management* had been developed within the university. The more recent introduction of performance-based salaries applies only to very recent appointments, and thus falls outside the scope of our investigation.

4. Comparison of the Cases

Informational Yields of RES Compared

'Informational yield' is only one of the properties that need to be taken into account in the design of a RES. The following discussion should therefore not be read as an assessment of RES. In particular, RES utilizing peer reviews are much more costly than indicator-based systems, and this property appears to affect political decisions on the design of RES.

The information outputs of the five RES are obviously different, with another significant variation between the 2001 and 2008 rounds of the British RAE. Our comparison of informational yields is based on the listed five properties of the information and their ranking. Figure 5.3 summarizes the comparison by providing 'informational footprints' for each of the five RES. We use the empirical information from the case studies to 'rank' the

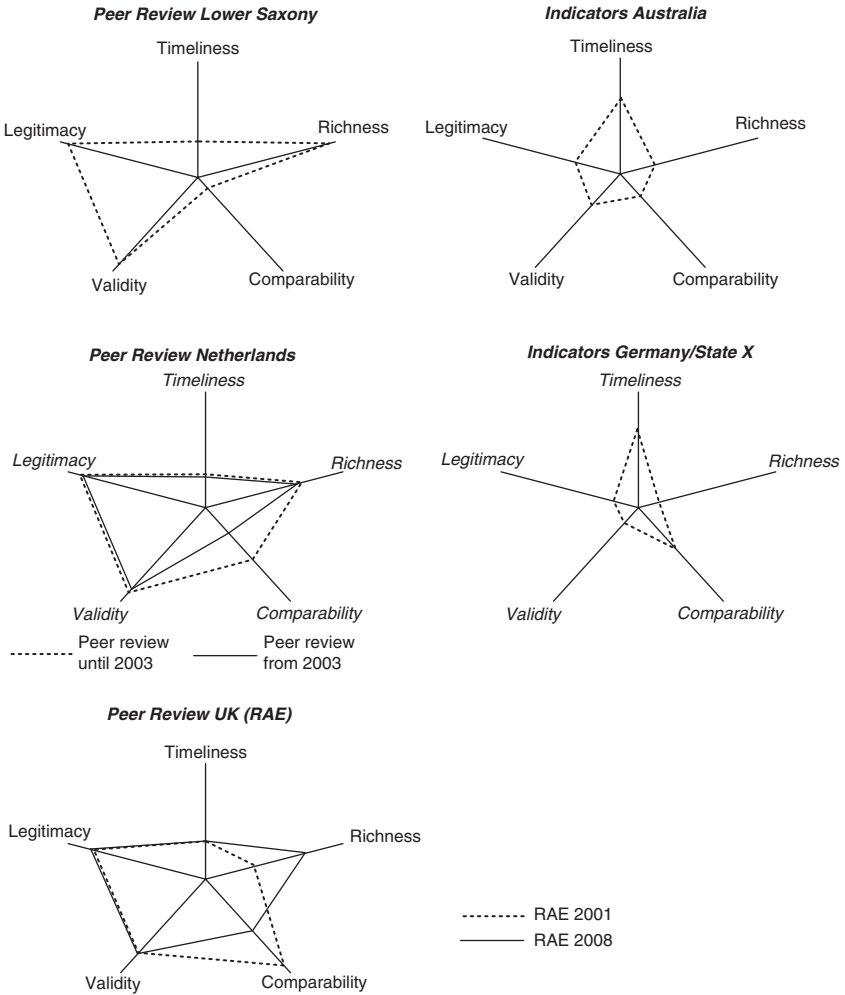


Figure 5.3. A comparison of five RES' informational yields for intra-university decisions

five systems in each of the five dimensions, and summarize the rankings in characteristic 'footprints' for each RES.

RICHNESS

The kind of rich information that is characteristic of peer reviews—detailed, multidimensional, and assessing past performance as well as potential—is only provided by the peer review conducted in Lower Saxony and its 'open-

ended' evaluation reports. Since the richness of information increases with the number of dimensions in which units are rated, the four-dimensional rating in the Netherlands is richer than any one-dimensional rating. Quantitative indicators offer little information at all, but the three indicators used in Australia provide richer information than the one-dimensional rating provided by the RAE until 2001. The quality profiles of the 2008 RAE are much richer than the outcomes of previous rounds. They consist of the proportions of research activity that fall into each quality category in three dimensions (quality of outputs, indicators of esteem, and research environment). As a result, it has become impossible to unambiguously rank the universities even in one unit of assessment (see below, comparability). The two indicators used in the German state X provide the least rich information.

TIMELINESS

The timeliness of information provided by peer review-based RES depends on the evaluation cycles. The peer review in Lower Saxony was conducted only once. Meanwhile, the state government decided not to repeat it—much to the regret of all universities' central administrators. The RAE started with a three-year cycle and then moved to distances of four, five, and now seven years between evaluations. The Dutch peer reviews are to be conducted once in six years. In all these cases the information lag already contained in the submitted material becomes aggravated, and universities are soon left with outdated information. The time lags are less severe in the case of indicator-based systems, because information collection is relatively cheap and already occurs annually in many universities.

VALIDITY

In terms of validity, peer-review-based systems are superior to the indicator-based systems currently in use (but see our remarks above on opacity and bias). The more information is gathered in a peer review, the more intense the analysis, and the more the assessors interact with the evaluated academics in order to validate information, the higher the validity.

The rather simple quantitative indicators used in Australia and in the German state X (which also is representative of the indicator-based systems in the other German states) do not achieve any satisfying level of validity. The information provided by these RES, while tolerated as the basis of resource allocation to universities, is therefore ill-suited for internal

decision-making about conditions for research (which does not preclude them from being widely used for precisely that purpose).

LEGITIMACY

The legitimacy of the interactive peer reviews in Lower Saxony can be considered to be the highest, while the legitimacy of the information provided by the British RAE is also very high even though it does not include interactive elements. Indicator-based information on research quality as it is currently produced by the RES with their simple indicators is not considered a legitimate basis for decisions. However, indicator-based information carries its own legitimacy when it stems from RES that are used for resource distribution. In the Australian universities the fact that the university receives funding according to the indicators legitimized decisions of the university management about their use for internal resource allocation, while no such argument could be observed in the German universities yet.

COMPARABILITY

The comparability of information was highest for the 2001 RAE, which produced a one-dimensional rating. The outcome of the Dutch RES is a rating in four dimensions, which makes comparisons difficult. The introduction of quality profiles with the 2008 RAE demonstrates the trade-off between the richness and the comparability of information. The introduction of a second dimension (the proportion of research that attained each quality level) makes comparisons dependent on a selection of information or on aggregation procedures (Travis 2009). The comparability of the indicator-based information is low because the numbers are field-specific, and the reference values that are necessary for comparisons do not exist.

The footprints in Figure 5.3 show how peer review-based and indicator-based systems produce quite different kinds of information. The outcomes of peer reviews are considered valid and legitimate. Their timeliness is problematic, mainly due to the large intervals at which they are conducted. Peer reviews have the potential to produce rich information about research, which leads to a trade-off between richness and comparability.

Indicator-based RES are not considered valid and carry little legitimacy. The numbers produced by these RES are neither comparable, nor do they contain rich information about research. Their only advantage is that some of the quantitative information can be produced quickly and cheaply,

which enables shorter evaluation cycles and increases the timeliness of information about research.

The Use of Information by Universities

Decision-making in universities is of course a political game in which the information provided by RES is just a resource. Nevertheless, it is possible to say something about the usability of this resource in that game. *Internal resource distribution* is best supported by RES that render their units of assessment comparable both within and between disciplines in one dimension. This is clearly expressed in the attempt of a Dutch university to turn the outcome of the Dutch peer review into a basis for internal resource allocation. The only way to achieve this was apparently to further collapse the multidimensional outcomes of the peer reviews into a single figure.

The importance of comparability for resource distribution that is indicated by this practice is due to an important difference between RES-based funding of universities and funding of sub-units within universities. The former is based on the implicit assumption that universities represent similar mixes of disciplines, while internal funding includes a *redistribution of resources between disciplines*. This aspect of universities' responses to RES has yet to be fully appreciated. The traditional use of peer reviews is limited to internal evaluations and redistributions of resources *within fields* exactly because research is evaluated by *peers*. When units of assessment are rendered comparable across disciplines, resources can be redistributed between disciplines on the basis of the comparable attribute (ratings in one dimension or a synthetic quantitative indicator). Since this attribute is some version of research quality, disciplines become differentially funded within universities on the basis of their quality as measured by the current RES. This makes universities major sites of a new process whose aggregate effects have yet to be investigated. From our cases we can draw the conclusion that quantitative indicators are ill-suited for redistributing resources between disciplines because of their dependency on size and field-specific research and publication practices. However, even if valid information from peer reviews is used, the redistribution of resources on the basis of only one criterion may have unanticipated aggregate effects.

Three of the RES we investigated (those of Lower Saxony, the UK, and Australia) are highly consequential for universities, and thus created strong incentives for universities to *change the structures of their research*. This task is best supported by traditional, dedicated peer reviews that take into account local conditions and potential and result in extensive verbal

recommendations. Even simple approaches such as the combination of 'weak' with 'strong' groups or the disguise of weak groups by relabelling them cannot be easily based on the comparable quality ratings or quantitative indicators provided by most RES, because this information does not enable any projection of what will result from such mergers. This is why Australian universities often used *ad hoc* peer reviews in order to assess the necessity of structural changes.

The most detailed information base for restructuring is provided by peer reviews as conducted in Lower Saxony. The detailed recommendations do not leave the university much choice beyond implementing or rejecting them. This kind of peer review thus restricts the opportunities for the university management to make its own decisions. There is a trade-off here between the richness of information and the room of manoeuvre for university management. The more formalized and parsimonious the outcomes of peer reviews are, the wider is the room of manoeuvre they provide for university management. If peer review outcomes are reduced to a one-dimensional rating, its function is reduced to legitimization, and it provides the widest range of options for university management. On the other hand, none of these options is any longer supported by information and legitimization from the peer review.

A third task for which information produced by RES provides an input is *individual performance management*. Since individual performance management is turning into a ubiquitous activity in universities, considerations of efficiency and practicability become very important. Australian universities tend to use the simple indicators of the funding formulae as simple indicators in yearly performance appraisals. The basic approach is benchmarking—one has to be as good as the average colleague from other universities. Another variant that is spreading is labelling. In order to count as 'research active', academics have to score on one or two of the three major indicators. Both practices rely on the use of quantitative indicators at the individual level, where they are largely invalid. Promotions are handled with greater care and include peer reviews for promotions to higher levels (professors). In other countries individual performance management appears to be less widespread. However, it will be interesting to see how individual performance will be measured in Germany, where it potentially could have severe consequences for both the resource base and the salary of a professor.

An important aspect of individual performance management is linked to academic identities (Henkel 2000, 2005). The RES significantly differ in their definition of academics' contributions to the income of the university. The peer-review-based system in the UK makes universities categorize

academics as either good researchers who are 'breadwinners' for their university, or bad researchers who are not. This categorization and the ensuing differential treatment of academics have been shown to create enormous stress and tensions in universities. The formula-based systems with their simple indicators are less disruptive insofar as they enable academics to contribute to their university's income with relatively small 'research' contributions. A refereed conference paper of 3,000 words earns the Australian university more than 2,000 AU\$, and the co-supervision of a Ph.D. student earns research money as well. This situation makes it possible for most academics to 'save their identities', but also deprives universities of an effective control instrument, which is why the dichotomy of 'research active' and 'research inactive' academics is becoming increasingly popular among Australian university managers.

5. Conclusions

The university's position in the system of authority relations concerning research is constrained by the specific nature of the research process. The authority of university decision-makers is limited because they cannot define research goals and because the relationships between conditions, aims, and outcomes of research processes are opaque to anybody except researchers in the same field. The authority of university decision-makers is also contingent, among other factors, on the information about the directions, quality, and efficiency of 'their' research they possess. This is why RES are not only important as stimuli for universities to improve their research, but also as sources of information that can be instrumental in the management of research quality within universities.

Our chapter has considered how information provided by RES can be, and is, used to manage the research process. We compared three peer-review-based and two indicator-based systems according to their informational yield and to the use to which the information is put by university management, and uncovered clear differences between the informational yields of varied RES for universities and between the ensuing usability of information for different purposes. The clear superiority of peer reviews was demonstrated by the range of changes it enabled in Lower Saxony, as well as by the distortions resulting from the internal use of quantitative indicators in Australia. Further indirect confirmation stems from the widespread *ad hoc* use of peer reviews in structural decisions, and decisions about promotions to professorial levels in Australia. The only advantage of quantitative

indicators (at least of the simple ones that are currently used in RES) is that they are cheap and timely.

In all our cases the information provided by RES is used by universities for managing their research. Thus, RES affect authority relations in science not only by stratifying the organizational field of higher education and by providing incentives for universities but also by strengthening (to varying extents) the authority and action capabilities of university management *vis-à-vis* its academics. However, there is an interesting trade-off between richness and action capabilities of university management. The richer the information, the more it might suggest a particular solution to a problem, which then becomes the only legitimate solution in the eyes of the academic community. The more advice management receives from peer reviews, the more difficult it becomes to act against that advice. If this information is ambiguous, the room of manoeuvre for university management increases again because it can choose which information to use.

In the case of peer-review-based RES, the internal use of the information by university management contributes to another change of authority relations. These RES do not only affect universities, but also the relationships between researchers and the scientific elite of their field. Since it is the elite who decide what is good research, RES employing peer review increase the dependence of researchers on the norms and assessments of their scientific elite. By using the same assessments internally, universities increase this dependency, thus relatively weakening the researcher's authority and relatively strengthening the elite's authority. This effect is strengthened by the fact that, except for the very rich information provided by the non-comparative RES in Lower Saxony, RES provide little support to university management when the latter needs to handle exceptions—cases of research that deviates from the mainstream and is of unknown quality.

Finally, comparing the five RES suggests some conclusions for the design of RES. Taking into account the internal use of external evaluations considerably increases the complexity of the impact of RES. It becomes obvious that the quest for the holy grail of the best RES is futile. The trade-offs between various aspects of the informational yield of an RES, on the one hand, and the usability of RES for other tasks, on the other hand, make any design decision a compromise, which in itself is a product of the authority relations in science.

References

- Bekhradnia, Bahram (2008) *Evaluating and Funding Research through the Proposed 'Research Excellence Framework'*: <http://www.hepi.ac.uk/466-1324/Evaluating-and-funding-research-through-the-proposed-Research-Excellence-Framework.html> (accessed Oct. 2009).
- Butler, Linda (2004) 'What Happens When Funding is Linked to Publication Counts?', in Henk F. Moed, Wolfgang Glänzel, and Ulrich Schmoch (eds.), *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (Dordrecht: Kluwer), 389–405.
- Choudhury, Vivek, and Sampler, Jeffrey L. (1997) 'Information Specificity and Environmental Scanning: An Economic Perspective', *Management Information Systems Quarterly*, 21: 25–53.
- Cohen, Michael D., March, James G., and Olsen, Johan P. (1972) 'A Garbage Can Model of Organizational Choice', *Administrative Science Quarterly*, 17: 1–25.
- Curtis, Polly (2002) 'Leicester's Communications Department to be Axed', *Guardian* (4 July).
- CPB (Netherlands Bureau for Economic Policy Analysis) and CHEPS (Centre for Higher Education Policy Studies, University of Twente) (2001) *Higher Education Reform: Getting the Incentives Right*, CPB Bijzondere publikaties, 29: <http://www.cpb.nl/nl/pub/cpbreesen/bijzonder/29/> (accessed Oct. 2009).
- Daft, Richard L., and Lengel, Robert H. (1984) 'Information Richness: A New Approach to Managerial Behavior and Organizational Design', in L. L. Cummings and B. M. Staw (eds.), *Research in Organizational Behavior*, vi (Homewood, Ill.: JAI Press), 191–233.
- and — (1986) 'Organizational Information Requirements, Media Richness, and Structural Design', *Management Science*, 32: 554–71.
- and Macintosh, Norman (1981) 'A Tentative Exploration into the Amount and Equivocality of Information Processed in Organisational Work-Units', *Administrative Science Quarterly*, 26: 207–24.
- Dainty, Roger, Williams, Syd, and Brown, Paul (1999) 'Current Trends and Changes to Research Strategy in UK Universities and Medical Schools', paper presented at the 50th Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Pittsburgh, PA.
- DEST (Department of Education, Science, and Training) (2007) *Higher Education Report 2005* (Canberra: Department of Education, Science, and Training): <http://www.dest.gov.au/NR/rdonlyres/5213D64C-C3BA-4D8A-B139-7906281D27AD/16972/HigherEdReport2005FINAL.pdf> (accessed Oct. 2009).
- Geuna, Aldo, and Martin, Ben R. (2003) 'University Research Evaluation and Funding: An International Comparison', *Minerva*, 41: 277–304.
- Gläser, Jochen (2006) *Wissenschaftliche Produktionsgemeinschaften: Die soziale Ordnung der Forschung* (Frankfurt a. M.: Campus).

- (2007) 'The Social Orders of Research Evaluation Systems', in Richard Whitley and Jochen Gläser (eds.), *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems* (Dordrecht: Springer), 245–66.
- (2008) *Evaluationsbasierte Managementsysteme für universitäre Forschungsleistungen* (Report to the German Ministry for Education and Science): http://www.bmbf.de/pub/evaluationsbasiertes_managementsystem_universitaere_forschungsleistung.pdf (accessed Oct. 2009).
- and Laudel, Grit (2007) 'Evaluation without Evaluators: The Impact of Funding Formulae on Australian University Research', in Richard Whitley and Jochen Gläser (eds.), *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems* (Dordrecht: Springer), 127–51.
- Lange, Stefan, Laudel, Grit, and Schimank, Uwe (2008) 'Evaluationsbasierte Forschungsfinanzierung und ihre Folgen', in Renate Mayntz, Friedhelm Neidhardt, Peter Weingart, and Ulrich Wengenroth (eds.), *Wissensproduktion und Wissenstransfer* (Bielefeld: transcript), 145–70.
- HEFCE (Higher Education Funding Council for England) (2004) *Higher Education in the United Kingdom* (Bristol: HEFCE): <http://www.hefce.ac.uk/Pubs/hefce/2004/HEinUK/HEinUK.pdf> (accessed Oct. 2009).
- (2008) *RAE 2008: The Outcome*: <http://www.rae.ac.uk/pubs/2008/01> (accessed Oct. 2009).
- Henkel, Mary (2000) *Academic Identities and Policy Change in Higher Education* (London: Jessica Kingsley).
- (2005) 'Academic Identity and Autonomy in a Changing Policy Environment', *Higher Education*, 49, 155–76.
- Isenberg, D. J. (1984) 'How Senior Managers Think', *Harvard Business Review*, 62: 81–90.
- Jauch, L. R., and W. F. Glueck (1975) 'Evaluation of University Professors' Research Performance', *Management Science*, 22: 66–75.
- Johnston, Chris, and Farrar, Steve (2003) '100 New Chairs Created in Bid to Lift RAE Scores', *Times Higher Education* (12 Dec.).
- Jongbloed, Ben, and Meulen, Barend van der (2006) *Investeren in Dynamiek: Endrapport Commissie Dynamisering*, 2nd edn. (Twente: Universiteit Twente, CHEPS).
- Lange, Stefan (2007) 'The Basic State of Research in Germany: Conditions of Knowledge Production Pre-Evaluation', in Richard Whitley and Jochen Gläser (eds.), *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems* (Dordrecht: Springer), 153–70.
- and Schimank, Uwe (2007) 'Zwischen Konvergenz und Pfadabhängigkeit: New Public Management in den Hochschulsystemen fünf ausgewählter OECD-Länder', in Katharina Holzinger, Helge Joergens, and Christoph Knill (eds.), *Transfer, Diffusion und Konvergenz von Politiken: Sonderheft der Politischen Vierteljahresschrift* (Wiesbaden: VS Verlag für Sozialwissenschaften), 522–48.
- Laudel, Grit (2006) 'The "quality myth": Promoting and Hinderling Conditions for Acquiring Research Funds', *Higher Education*, 52: 375–403.

Reorganizing Research Organizations

- Lucas, Lisa (2006) *The Research Game in Academic Life* (Maidenhead: SRHE/Open University Press).
- Meulen, Barend van der (2007) 'Interfering Governance and Emerging Centres of Control: University Research Evaluation in the Netherlands', in Richard Whitley and Jochen Gläser (eds.), *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems* (Dordrecht: Springer), 191–203.
- Morris, Norma (2002) 'The Developing Role of Departments', *Research Policy*, 31: 817–33.
- Northumbria University (2007) *2008 RAE Code of Practice: Northumbria University* (Newcastle upon Tyne: Northumbria University).
- Oxford University (2005) 'The University's Resource Allocation Method (RAM) (Updated for 2005–6)', *Oxford University Gazette* (20 Oct.).
- Paradeise, C., Reale, E., Bleiklie, I., and Ferlie, E. (eds.) (2009) *University Governance: Western European Comparative Perspectives* (London: Springer).
- Parker, Jonathan (2008) 'Comparing Research and Teaching in University Promotion Criteria', *Higher Education Quarterly*, 62: 237–51.
- Phillimore, A. J. (1989) 'University Research Performance Indicators in Practice: The University Grants Committee's Evaluation of British Universities, 1985–86', *Research Policy*, 18: 255–71.
- Queen Mary University of London (2003) *Research Strategy: 2003 to 2008*: <http://www.qmul.ac.uk/research/policies/docs/s-research0308.pdf> (accessed Oct. 2009).
- Roberts, Sir Gareth (2003) *Review of Research Assessment* (London: Higher Education Funding Council for England): <http://www.ra-review.ac.uk/reports/roberts.asp> (accessed Oct. 2009).
- Saunders, Carol, and Jones, Jack W. (1990) 'Temporal Sequences in Information Acquisition for Decision Making: A Focus on Source and Medium', *Academy of Management Review*, 15: 29–46.
- Schiene, Christoph, and Schimank, Uwe (2007) 'Research Evaluation as Organisational Development: The Work of the Academic Advisory Council in Lower Saxony (FRG)', in Richard Whitley and Jochen Gläser (eds.), *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems* (Dordrecht: Springer), 171–90.
- Schimank, Uwe (2005) "'New Public Management" and the Academic Profession: Reflections on the German Situation', *Minerva*, 43: 361–76.
- Travis, John (2009) 'Research Assessment: U.K. University Research Ranked; Funding Impacts to Follow', *Science* (2 Jan.): 323–4.
- Van Raan, A. F. J. (1996) 'Advanced Bibliometric Methods as Quantitative Core of Peer-Review Based Evaluation and Foresight Exercises', *Scientometrics*, 36: 397–420.
- VSNU (Vereniging van Nederlandse Universiteiten), NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek), and KNAW (Koninklijke Nederlandse Akademie van Wetenschappen) (2003) *Standard Evaluation Protocol for Public Research Organisations* (Utrecht: VSNU, NWO, and KNAW): <http://www.qanu.nl/comasy/uploadedfiles/sep2003-2009.pdf> (accessed Oct. 2009).

- Weick, Karl E. (1979) *The Social Psychology of Organizing* (Reading, Mass.: Addison-Wesley).
- Weingart, Peter, and Maasen, Sabine (2007) 'Elite through Rankings: The Emergence of the Enterprising University', in Richard Whitley and Jochen Gläser (eds.), *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems* (Dordrecht: Springer), 75–99.
- Westerheijden, Don F. (1997) 'A Solid Base for Decisions: Use of VSNU Research Evaluations in Dutch Universities', *Higher Education*, 33: 397–413.
- Whitley, Richard (2008) 'Constructing Universities as Strategic Actors: Limitations and Variations', in L. Engwall and D. Weaire (eds.), *The University in the Market* (London: Portland Press), 23–37.
- and Jochen Gläser (eds.) (2007) *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems* (Dordrecht: Springer).
- Wood, F. Q. (1989) 'Assessing Research Performance of University Academic Staff: Measures and Problems', *Higher Education Research and Development*, 8: 237–50.

Reconfiguring Knowledge Production

Changing Authority Relationships in
the Sciences and their Consequences
for Intellectual Innovation

Edited by
RICHARD WHITLEY,
JOCHEN GLÄSER,
and
LARS ENGWALL

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi

Kuala Lumpur Madrid Melbourne Mexico City Nairobi

New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece

Guatemala Hungary Italy Japan Poland Portugal Singapore

South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

© Oxford University Press, 2010

The moral rights of the author have been asserted
Database right Oxford University Press (maker)

First published 2010

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
MPG Group, Bodwin and King's Lynn

ISBN 978-0-19-959019-3

1 3 5 7 9 10 8 6 4 2