

Performance Assessment of State-Financed Research

Raluca M. MIHALCIOIU¹

Abstract

During the last three decades assessing the performance of the public funded research has become a political issue in most countries. Governments now expect from science maximum efficiency and value for money allocated to science (value for money). Therefore, tools were developed and were implemented to evaluate research performance in universities.

This article analysis the research evaluation systems in first - mover counties, UK, Netherlands, Australia and Germany and aims at drawing conclusions about positive and negative effects that the implementation of this systems brings.

Keywords: evaluation, research funding, performance, higher education systems.

JEL classification: D61, H52, I22, I28

1. Introduction

"Whether you think you can, or you think you can't - you're right."
Henry Ford (1863 - 1947)

The financial crisis has seriously affected the budgets of many states, therefore the problem of performance evaluation research in universities has become of major interest. Implicit social contract with science that state finances unconditionally research worsened. Governments expect today maximum efficiency for money allocated for science and even establish framework conditions for research and how value can be created. For this purpose, many countries have introduced research performance evaluation systems in their universities.

Evaluation systems are based on studies and analysis of performance. They are used for the redistribution of university research basic funding and they usually focus on limited resources and financial incentives for universities to increase their research performance. By implementing evaluation systems, universities aim to inform policymakers about the efficiency of research and to develop an university management for research performance. With this new theory it opens a field for research of the governance of science.

Since late 90's Germany experiences implementing tools for assessment to increase the performance of the university research. Many federal states (Länder) have introduced funding systems of instruments for higher education funding based

¹ Raluca M. MIHALCIOIU, Spiru Haret University Bucharest, Romania
Email: ralucamarilena@googlemail.com

on performance indicators of research. Pilot studies were undertaken at Federal level by the Scientific Council in order to evaluate federal research and excellence initiatives, to highlight performance and to enter a differentiation among higher education institutions (Whitley & Gläser, 2007). These experiments launched numerous political debates on the success of performance improvement initiatives on public funded research based on assessments and the validity of the assessment procedure and negative side effects that may occur.

The purpose of this paper is to comparative analyzing the evaluation of research systems in countries that are considered pioneers in this field and it wants to make a contribution to the debate regarding the increase of research performance. This paper discusses the emergence and development of research evaluation systems in Germany, UK, Australia and Netherlands stressing the positive and negative effects of these systems.

2. University research evaluation systems in Germany

Since 1970 the Federal Council of Science has undertaken ad hoc individual assessments in research institutes, and has made the first systemic evaluation of the research institutes only after German reunification in the early 90s, but it had no significant influence on universities. With one exception of the Land Niedersachsen that applied the assessment procedure of the Federal Council of Scientific for research at universities, no another provincial government has conducted a systematic evaluation by peer review of research in higher education institutions. Much of the budgets of universities today face difficulties in differentiating between the resources devoted to research and teaching (König, 2007). Moreover, in most provinces budgeting takes place on the organizational scheme and "eternal" appointments of the professors.

This situation changed in the late 90s when many universities have gained limited autonomy with implementation of the new higher education framework law, followed by new laws at local level regarding higher education that provide a global budget, fixed-term appointments, salary bonuses according to performance and skills for newly hired professors and extended competences for deans. In parallel with the establishment of this new relationship between state, university and teaching staff have been introduced rather fragile systems of performance evaluation research, such as Target Agreements and performance oriented allocation of the resources. The Target Agreements between Government of Land and its universities provide stimulating innovations in research and developments in education and offer universities a quite higher degree of protection or safety on light deterioration of their financial status quo. Often by making "innovations" it was meant the establishment of training centers on specific areas on research, centers and schools for PhD. students. Target Agreements provide only in exceptional cases additional resources from the provinces (for land), they are not subject to rigorous control. Moreover non-Fulfillment of the agreement is not sanctioned. Therefore it is estimated that they are rather weak (Classen, 2009).

Performance-oriented allocation of the financial resources correlates in a little way the annual budget items with the fulfillment of the performance criteria in education and research, which is measured by quantitative performance indicators. Thus in the Land Nordrhein - Westfalen are granted funds for research and education in a percentage of 20% based on performance indicators and in Land Niedersachsen even 80%. This was observed in 13 of the 16 German states (land), where funding for universities are given a lesser extent on performance indicators. Most of them are allocated for human resources and are not influenced by the indicators.

As mentioned above, Germany has 16 different institutional routes (according to its 16 federal states-land), therefore great efforts are made to achieve a federal system of assessment in order to increase the research performance. Thus, the Scientific Council has undertaken two pilot studies at federal level for Research Rating in which the whole research in fields of chemistry and sociology was assessed by expert committees and appreciated compared according to several dimensions. Other pilot studies have taken place in electronics engineering and computer science. This Research Rating didn't affect the financing of universities. Quite different happens for the excellent initiatives at federal and local level, where important financial resources are granted for PhD schools, the scientific excellence groups and developing concepts for the future, which describe long-term development of research in universities (Frey, 2008). Excellent initiatives have been promoted by the federal government and can be considered a three-tier approach, in fact a compromise between the three groups of stakeholders:

- the Federal Government with his political need to promote selectively universities in international rankings of the elite higher education;
- Federal states (Länder), with their interest in a more equitable distribution of federal funds;
- the scientific community, which demonstrated that excellence in research doesn't take place at the university level (but of specific departments) and it can be established only by science itself.

3. Research Assessment Exercise - Assessing the university research in the UK

The implementation of the Research Selectivity Exercity in the late 1970 - renamed in 1989 as Research Assessment Exercise was the response of the concerns about the degeneration of the Research Infrastructure. The Merrison Report in 1979 records that the extent to which the infrastructure of university research had been eroded and argued that if further damage was to be prevented then resources for scientific research had to be allocated selectively. Following this situation, the British government under Thatcher diminished in the early '80s the public funding for universities. In the first research selectivity exercise was asked to provide information about attracting external funds for research and the number of students engaged in research (master and PhD Students). Moreover, the

evaluators commission asked for 5 publications/ department of university, which can be considered relevant for the research done in a specific field by the university (Barker, 2007). The evaluation was then progressively improved and professionalized and it is characterized by three main elements:

- the centralized comparative method which is the responsibility of public funding bodies (Higher Education Funding Council for Great Britain, Scotland and Wales);
- the application of Peer Reviews as an evaluation method;
- the use of evaluation results for the selective award of basic hardware needed for research in universities.

The Research Assessment Exercise classifies the level of research made within a department in quality profiles on a quality scale: (from 5= world-leading to 1= weak performance also at national level), without receiving a final grade.

The British universities allocate most of the funds received from Research Assessment Exercise on merit criteria: departments, which by their good performance contribute to the university budget are given some resources while others get nothing. The top management of the

University preserves part of the money collected by Research Assessment Exercise to secure the financial resources necessary to make strategic decisions (appointments, training, etc.). Strategic investments are a key feature of management research in British universities.

In some Universities (for example the University of Liverpool) have been undertaken restructuring of departments, fact that actually focused on strengthening of specialized units that would have better opportunities to the next Research Assessment Exercise.

Moreover, the head-hunting even of excellent researchers has become a common practice in order to strengthen the university own culture of research and bring better performance to the next Research Assessment Exercise.

Through the implementation of the Research Assessment Exercise was achieved an improvement of the overall quality of research in British universities. At least this is the opinion of most managers and university scientists. Research Assessment Exercise promotes a new culture in British universities, which provide high value aspects of the research regarding recruitment and employment of scientists.

However, some Research Assessment Exercise results put into question the growth of research performance by selective financing.

The table below shows a weak connection between selective funding and performance: research units that in 2008 have been certified by 35% to 50% research at world-leading level and excellent internationally haven't received financing from the Research Assessment Exercise budget due to poor grades in the range 2001-2008. This means that only limited can be considered the funding for high or low performance.

Table 1: Comparative classification by Research Assessment Exercise 2001 and 2008 for universities in informatics field

University	RAE ² 2001	The significance level from 2001 in funding formula for England 2003	Research Assessment Exercise 2008 Quality Profile (Percentage distribution of high quality research on five levels)				
			4*	3*	2*	1*	unclassified
YORK	5*	3357	25%	50%	20%	5%	0%
BIRMINGHAM	5	2793	30%	45%	20%	5%	0%
ESSEX	4	1,00	15%	50%	30%	5%	0%
LOUGHBOROUGH	3a	0	15%	40%	35%	10%	0%
GOLDSMITHS COLLEGE	3b	0	20%	40%	30%	10%	0%
BOURNEMOUTH	2	0	5%	30%	50%	15%	0%

Source: adapted from Barker & Gläser

The degrees show: 5*: >50% excellent international, 5: <50% excellent international, 4: excellent at national level, some evidence of international excellence, 3a: >2/3 excellent at national level, some evidence of international excellence, 3b: > 50% excellent at national level, 2: < 50% excellent at national level, 1: no level of excellence at national level. The quality profile shows: 4* international leader, excellent, 3* international excellent, 2* internationally recognized, 1* recognized at national level.

Research Assessment Exercise presents also negative effects that can be attributed on one hand to evaluation procedures used and on the other hand to the reaction of the universities. But Peer Review that is behind Research Assessment Exercise has the advantages over other procedures used in the quantitative content analysis because it allows analyzing research on quality assessment.

The focus of the university management on research performance, which extends to assessing individual performance and competence in the educational process include redistribution of the competences and turn education actually into a punishment for poor research. This tendency to split the process of education from research has not been investigated so far in its dynamics and consequences that entails.

4. Australia - Evaluation based on indicators

In the late 1980s the Australian Government has abolished his dual higher education system after reforms on national higher education. Due to this reforms all institutions of higher education were upgraded to universities during a process of modernization, fact that immediately followed the efforts for application of selective financing of universities. In White Paper of Higher Education, the federal government has declared its intention to develop a financial system that take into

² RAE - Research Assessment Exercise

account the performance of universities. In 1990 was implemented the Relative Funding Model, which differentiates between education and research bodies in financing an university. The research component included costs of research infrastructure but it didn't fully cover the whole research costs. Then, there were used indicators to measure performance, like the number of publications, the number of graduates per degree (Master, PhD.) and the number of students enrolled in research. In determining the number of publications it has began to operate with 22 categories, fact that met difficulties in practice. Therefore, it was then operated only with four categories of publications considered relevant for financial resources for research, namely articles appreciated magazine, papers accepted for famous conference, books and chapters in books. Currently, the proportion of research funding based on evaluation's results is 7.9% of university revenue.

The Australian academic environment reacts to a rating system that connects core funding for universities in large part by external sources (third parties). Indicators such as publications and the number of Ph.D. students play small part but are considered and used as criteria in decisions regarding promotions (Parker, 2008).

In Australia can be observed the desired positive effects of an evaluation system. Also, the redistribution of resources (based on used indicators) to the strongest researchers contribute to increased quality. Negative effects arise in general orientation on a few exclusive universities of research performance indicators and the limitation of funds, which compel all university research community to focus on the needs of external funders. These third parties (external sponsors) are often described as monopolistic, karg and biased. Australian researchers are connected to a single relevant source of financing, the Australian Research Council and National Health and Medical Research for medical research. Other sources, like Industry and the Government, do not play any important role. Research Councils are subordinated to the Ministry of Science and Ministry of Health (for medical research). Ministers finally agree on funding sources for projects and even have jurisdiction to refuse projects.

5. Netherlands - Standard Evaluation Protocol

The Dutch research performance assessment was introduced in the mid of 80s in order to provide the government with information on the quality of publicly funded research. The cause which led to a system of evaluation was the lack of transparency and accountability for research in universities. The Holland University research was supported to a usually large extent through the general block grant of the university (Blume, 1988) and global performance of the research was not controlled.

Initially, universities were responsible for evaluating their own research organization: they should undertake an internal assessment every 3-years and an external evaluation every 6-years. Then, the responsibility for evaluation was transferred to Vereniging van Universiteiten (Ministry for the United hollandish universities) and it was introduced the Standard Evaluation Protocol for Public

Research Organizations, which includes standards to increase performance evaluation and final assessment of the quality (Blume & Saapen, 1988). The Protocol provides an evaluation research in the four dimensions of quality, productivity, scientific relevance and ability of development. In these dimensions the assessment operates on a scale of 5 points from excellent to poor.

Also in the Netherlands the information regarding evaluation of the quality of research are the base for budget negotiations between different levels of management of the university: the top management, department, institute, and research group / research school. The results of these internal negotiations do not make the research quality evaluation in inputs or outputs of money, by which units that are assessed as weak are punished and the one evaluated as excellent awarded. Instead, research units assessed as poor could receive more money because they are important for college or department for other reasons (teaching is good, attractive to students, etc.) and therefore they should be strengthened. There are cases in which research units assessed as excellent should give up money if, for example the number of students drops significantly.

Most scientists see assessments as having a major impact on university environment. Good performance provides some protection against management interventions, such as for example the restructuring and also gives a strong position in budget negotiations (van der Meulen, 2007). Poor outcomes reduce this protection automatically but don't have necessary negative consequences in financial terms. Assessments of research in the Netherlands are therefore an important source of information for decision making for an active management.

6. Comparative analysis of research performance evaluation systems

The main reason behind the establishment of common evaluation systems in all countries analyzed above (Germany, UK, Australia and Netherlands) has been to provide on the one hand the university management with information and on the other hand information for government for a selective allocation of financial resources for research.

If the UK is characterized by ratings that make obvious the best research locations and the worst ones, the Netherlands focus on strengthening the attractiveness for students and research centers regardless of evaluations results, and external financing of research in Australia is an important indicator of performance measurement research, in Germany cannot outline a specific element of assessment the performance research. In the German university landscape there is rather a specific combination of elements met in leading countries (UK, the Netherlands) in assessing research performance. Unlike Australia, the German external financing system is pluralistic, rich, neutral, meaning there are a lot of funding agencies with public and private financing, with a large number of foundations that support research (Schwarz & Westerheijden, 2004).

Current research ratings tested by the Federal Council of Science resembles the procedures of research assessment systems in the Netherlands and the UK. Of course, one can hardly speak of an institutionalized system in Germany, given the pilot studies carried out so far.

Since there is no link between the mechanism of allocating financial resources for research and evaluations undertaken, as the Research Assessment Exercise in the UK, it can be considered that the German is closer to the Standard Evaluation Protocol from the Netherlands. Here, the results are less part of national higher education policy and more an organization problem, serving the university itself for the strategic orientation (Lange & Gläser, 2009). Besides that, decisions on the distribution of financial resources or structural changes are not dominated by the evaluations results. Therefore, excellent research units can be closed or forced to merge with each other if too few students draw or they are assessed as less relevant to the profile of the university and research units that are poorly evaluated and consolidated and financial encouraged in order to develop their research capacity (Gläser et al, 2009).

Table 2 Comparison between German, Dutch and Britisch research evaluation systems

	WR ³ -Rating Germany	Standard Evaluation Protocol Netherlands	RAE United Kingdom
unit to evaluate	research in one field of the university		
Participation	voluntary (but under a high informal pressure)	compulsory	voluntary
Information Required	Publications and contextual information, like applications for external funds, and number of students involved in research (PhD. students)		
Experts recruiting	mostly national Peers		
Method	Analysis of publications and contextual information		
Results	Rating from 5 („excellent“) to 1(„poor“) in three dimensions (research, young researchers, transfer of knowledge)	Rating from 5 („excellent“) to 1(„poor“) in four dimensions (quality, productivity, scientific relevance, ability to develop)	Quality profile: distribution of research in five levels of quality from "world leader" to "below the national standard"
Link to finance	no	no	Yes, strong (about 7,5% from university financial resources)

The table above shows that for the evaluation of performance research in Germany and Great Britain the participation of the universities in classification is voluntary, therefore not all universities are classified. Both rankings are based on Peer Review with the difference that in the UK they have significant consequences.

³ WR - Wissenschaftsrat - Rating = Classification of the Federal Council of Science, Germany
RAE - Research Assessment Exercise in the UK

In Germany and the Netherlands there is no link between budget for research and the evaluation's results.

7. Conclusion

Performance evaluation of public funded research is an important topic of debate in the current financial crisis. State budgets are increasingly poorer, public debt is hardly kept under control and the one affected mostly from this situation is the educational environment. Therefore, discussions are needed on the effectiveness of public money allocated for research in universities.

Several conclusions can be drawn from international comparisons. A first observation is that performance appraisal systems are systems that improves, develops and adapts to new conditions. For this learning process they need a permanent evaluation.

Peer Review assessment process used in quality of research evaluation systems is more effective than the simple quantitative indicators. Australian experience shows that systems based on indicators developed a strong goal displacement, where a good graduation in numerical value as indicators reflect replaces in improvement of quality in research.

Two important international experiences, which are used equally in the assessment systems and incentives to excellent affect selectivity limits and performance processes balance of universities. Research Assessment Exercise in Great Britain shows that selective research funding risk not to reward the niches of excellence and co-finance the mediocre research.

An important positive feature of performance improvement tools for research is that the attention of university management and scientists focuses on the performance of research. This feature is both highly problematic because it moves the priorities in favor of some evaluated and rewarded processes in a very complex performance process. In Great Britain and Australia it is obvious that strong evaluation systems lead to a relative neglect of some performance processes, that will not be rewarded as strong as the others, so as the non-oriented research to external financing in Australia and applied research in the UK and education in both countries. This side effect can not be eliminated through improved tools, because it affects the balance between performance processes evaluated and rewarded on the one hand and performance processes unrated and not rewarded on the other hand. It is necessary to use performance oriented governance performance instruments for other processes, which then should coordinate proportioned funding volumes properly.

It can be concluded that tools for improving research performance should ideally not only to influence the content of research but to improve the quality of research. Studies on Research Assessment Evaluation in Great Britain and the indicators based evaluation in Australia shows that applied research, risky research and non-conformist research - eventually all the research that do not attract external funding and have a reduced echo in magazines, endanger the University revenues and the careers of researchers if they are connected with the evaluation systems of financial resource allocation decisions. In the center of interests are "milk cows" that bring to

universities more money. On such a development react the researchers themselves with opportunistic behavior. In a cumulative effect the research undertaken in universities may slip to an irreversible spiral of diversity pool. A potential loss of diversity in itself is hardly measurable. This will achieve a basic dilemma of every research performance evaluation, which should be supervised in the same way by the research policy and by the scientific management: performance evaluation and assessment systems often lead to strong homogenization of the institutional research environment, which exposes research to a strong pressure of homogenization. Therefore it must not forget that universities have a social function of increasing complexity and not to reduce it.

References

1. Barker, K. (2007). The UK Research Assessment Exercise: the Evolution of a National Research Evaluation System, *Research Evaluation*, No. 16;
2. Blume, S.S. & Saapen J.R. (1988). External Assessment and Conditional Financing of Research in Dutch Universities, *Minerva*, Vol. 26;
3. Classen, C.D. (2009). Zielvereinbarungen - Ein neues Steuerungsinstrument der Hochschulpolitik aus juristischer Sicht, *Hochschulmanagement*, Vol. 4;
4. Frey, B.S. (2008). *Evaluitis - eine neue Krankheit*, in *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluation*, Wiesbaden, Verlag für Sozialwissenschaften
5. Gläser, J., Lange, S. & Laudel, G. (2009). Auswirkungen der evaluationsbasierten Forschungsförderung an Universitäten auf die Inhalte der Forschung. Australien und Deutschland im Vergleich, *Wissenschaftsrecht*, Vol. 42;
6. Higher Education Funding Council for England (2008) *Research Assessment Exercise*, accessed on www.rae.ac.uk
7. König, K. (2007). *Die Reform der staatlichen Hochschulsteuerung durch Kontraktmanagement und Ihre Bedeutung für Mitbestimmung und Chancengleichheit an den Hochschulen*, Wittenberg, Institut für Hochschulforschung
8. Lange, S. & Gläser, J., (2009) *Performanzsteigerung durch Selektivität*, Der moderne Staat, Zeitschrift für Public Policy, Recht und Management, nr. 2
9. Parker, J. (2008). Comparing Research and Teaching in University Promotion Criteria, in *Higher Education Quarterly* 62
10. Schwarz, S. & Westerheijden, D.F. (2004). *Accreditation and Evaluation in the European Higher Education Area*, Dordrecht
11. van der Meulen, B. (2007). *Interfering Governance and Emerging Centers of Control*, University Research Evaluation in the Netherlands
12. Whitley, R. & Gläser, J. (2007). *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, Dordrecht, Springer