berlin-brandenburgische
AKADEMIE DER WISSENSCHAFTEN

Barend van der Meulen

Evaluation and the governance of academic research : experiences and recent developments in the Netherlands

In: Berichte und Abhandlungen / Berlin-Brandenburgische Akademie der Wissenschaften (vormals Preußische Akademie der Wissenschaften) ; 9.2002, S. 267-281

Persistent Identifier: urn:nbn:de:kobv:b4-opus4-32448

Die vorliegende Datei wird Ihnen von der Berlin-Brandenburgischen Akademie der Wissenschaften unter einer Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (cc by-nc-sa 4.0) Licence zur Verfügung gestellt.

Barend van der Meulen

Evaluation and the governance of academic research

Experiences and recent developments in the Netherlands

Analysing the emergence of new evaluation practices for academic research in the eighties, one cannot but link these to developments in the relationship of universities and governments. After World War II academics became accustomed to an autonomy regime, in which governments were willing to provide funds for academic research exercising control neither on the academic performance nor on the returns of the investments. Quality control was left to the academic sector or, more precisely, quality control was seen as implicit to the dynamics of science and not something that had to be organized separately. Similarly, the value of basic science was not disputed and, although unpredictable, return on investment could surely be expected at the time of knowledge application. In the eighties, in many Western countries governments became less generous, more demanding on the return on investment, and implemented new evaluation practices to increase 'accountability'.[1]
In the Netherlands, new evaluation practices started already in the early eighties. Since then it has evolved from a loosely co-ordinated exercise organized by the government to a well-established practice for which the universities themselves are responsible.
This paper analyses the impact of these evaluation exercises on the governance of research. Governance is a broader concept than government, referring in general to processes of co-ordination and collective decision-making. Governance can occur without a government playing a dominant governing role. From a political perspective the concept of governance is becoming more important with the changing role of governments in Western society, the development of new interactions between government and society and the emphasis within political theory on institutions.[2] Nowadays, responsibilities traditionally delegated by society to the polity, like the

---

1 See for instance: OECD (ed.): Universities in Transition, Paris: OECD Publications, 1998.

2 March, J. G. & J. P.Olsen: Democratic governance, New York: The Free Press,1995.

development of collective goods, are secured within new institutions that are collectively maintained by agencies, civic groups, firms, regulation bodies, public organizations, and governmental bodies (rather than 'the government').[3] Of course such conception of governance is all but strange to the scientific system. The autonomy of science has for a long time been considered to be at odds with a delegation of any responsibilities for the functioning of the academic research system to the government, but the financing of the research.

From that perspective, it is ironic that 'evaluation of academic research' has become so much identified with accountability to government, and so little with other issues of the governance of science. Evaluation practices have developed up to a point that they have become part and parcel of academic life. Therefore it is important not only to assess the impact of evaluation practices in the context of the government-university relation, but also on the organization of research within universities, on the production of scientific knowledge and on the functioning and role of academic research in the wider society. Although in many countries systematic evaluation of academic research came about in times of budget cuts or other crises in the government-university relation, the increase of evaluation processes is also related to the important role universities have in knowledge societies. The development of new evaluation practices has been seen as characteristic for new modes of knowledge production, characterized by trans-disciplinarity, application-oriented, organization diversity and new forms of quality control.[4]

Within this paper not all these issues can be done full justice, but at least Dutch experiences make clear that such aspects of the governance of research should not be neglected. The first two sections of the paper summarize the development of university research evaluation and specifically address the changes in the government-university relationship as well as the evolution of evaluation methods. The third section describes in detail the evaluation process of the Association of Dutch Universities (VSNU). The last two sections assess the impact of this evaluation practice on the governance of research. The fourth section addresses the role of the evaluation outcomes in university decision making, the fifth addresses impacts on the mode of knowledge production.

---

[3] Kooiman, J. (ed.): Modern Governance. New Government-Society Interactions, London: Sage Publications, 1993.

[4] Gibbons, M. et al.: The New Production of Knowledge: The dynamics of science and research in contemporary societies, London: Sage Publications, 1994.

1 Evaluation of university research: key actors

Like in many research systems, until the eighties, in the Netherlands evaluation of research was limited to peer review of research proposals within the national research council. As a consequence most research was not evaluated at all. Since then evaluation has emerged in more contexts and evolved into an accepted practice, but without a dominant model or dominant actor.[5] Key actors are the Netherlands Research Council (NWO), the Royal Netherlands Academy of Arts and Sciences (KNAW) and the universities, collaborating in the Association of Dutch Universities (VSNU). The research council was and still is responsible for competitive evaluations, but with the diversification of funding modes, it had to develop also other forms of evaluation. In addition to the traditional peer review of proposals, NWO implemented new procedures for selecting research programmes and excellent researchers, for evaluation of its research institutes and for ex post evaluation of research programmes. Recently, the research council was responsible for the selection of six excellent graduate schools, *Toponderzoekscholen,* that were granted extra funding from the general university funds.

The Royal Academy is responsible for most of the ad hoc peer reviews, e. g. of national research programmes, sometimes of disciplines and, of course, for the evaluation of its own research institutes. Examples of ad hoc evaluations are the selection of three technological top institutes – collaborative institutes of universities, technological institutes and industry – in 1995 and the evaluation of medical research in 1998. The KNAW also hosts the accreditation committee for graduate schools. In the Netherlands, most university research and PhD education is organized within (inter-)university graduate schools. These graduate schools have to be accredited every five years. Criteria for accreditation include the structure of the PhD education, the management of the school, the quality of the research programme and the labour market for PhD students.

The third block of evaluation of academic and basic research is the evaluation of university research programmes by the Association of the Dutch Universities. This evaluation scheme ascertains that every university research programme is discipline-wise evaluated every four years in a rolling scheme. It was implemented in the eighties by the government to increase accountability, but since then has made various transformations. By now, it is organized by the Association of Universities aiming first of all at management of research by the universities. The next section describes the evolution of this practice.

---

[5] See also Rip, A. & B. J. R. van der Meulen: The Patchwork of the Dutch Evaluation System. In: Research Evaluation, 5 (1995) 1, S. 45–53.

2 The evolution of an evaluation practice

In the early eighties, the Ministry for Education and Sciences came up with ideas for evaluation of research in order to improve the planning of university research within the context of higher education policy and science policy. Already in 1982, the first evaluation round started. Since then three rounds of about five years have been completed and the fourth is halfway. It is not possible to give a full overview of each of these evaluation cycles. To understand the present functioning and position of the VSNU research evaluation, it is useful to highlight some experiences and developments in relation to evaluation methods and policy context.[6]

The evaluation procedure that set off in 1982 was linked to a new funding scheme for university research. This funding scheme forced the universities to organize most of their research (at least 80 %) in research programmes that had to be assessed ex ante by peer committees. Universities that could not fill up a considerable part of their research budget with approved programmes could lose funding. Positive assessment implied a budget protection of the programmes for five years. Within hindsight, in terms of policy planning the new system indeed led to a more transparent organization of university research. But as each university easily met its target within four years, there were no consequences in terms of budget allocation or priority setting. However, during these years the government announced and implemented its first cuts in the university budget. These cuts did not only minimise the trust that research programmes were protected indeed for five years, it also enforced the impression, that the government wanted to centralize the university research system.

The results of the ex-ante assessments, although in general positive, did not have much legitimacy within the universities and outcomes had little impact on decision making. With respect to evaluation methods, for all actors involved it was unclear how assessments had to be made and how they were made. All peer committees were asked to evaluate scientific quality and societal relevance, but without proper discussions or indications how these concepts had to be interpreted, the committees used different indications for judging the quality. Some of the committees assessed only the programme descriptions. Others focused on the scientists within the programmes or, being peers based their judgement on their general knowledge of the group's research performance. None of the committees was able to evaluate societal relevance systematically. Moreover, with an eye on the policy context, most evaluation committees were not very harsh in their judgements, while a few were very strict in their evaluations.[7]

---

[6] See also Ball, D. F. & A. Verkleij: University research as a business: a comparison of research assessment exercises in the United Kingdom and the Netherlands, 1999.

[7] Spaapen, J. B. et al.: Evaluatie van vijf jaar voorwaardelijke financiering, De moeizame relatie tussen beleid en onderzoek, Zoetermeer: Ministerie van Onderwijs en Wetenschappen/Staatsuitgeverij, 1988.

In 1985, the universities successfully negotiated a change in the evaluation procedures. In 1987, the second round of evaluation of conditional financed programmes set off. Instead of planning, the emphasis was now on accountability and instead of ex-ante assessments, the research programmes were evaluated ex post. Although within the Ministry the idea remained that the evaluation outcomes should have consequences for budget allocations to universities, no rules or procedures were developed. Responsibility for the evaluations shifted, depending on the discipline, to the research council, the Academy and the Royal Institute of Engineers. Although a strict evaluation procedure was still lacking, making ex-post evaluations turned out to be less problematic. Contrary to the first round, most evaluation outcomes were seen as reliable. Within the universities the outcomes became more important, and actors became aware of the possibility to use the evaluations as part of university research management.

The appreciation of the outcomes was also related to a growing acceptance of the use of output indicators to evaluate performance. However, what remained difficult was the evaluation of societal relevance. Especially in fields like engineering sciences, humanities and fields like the environmental sciences and policy studies researchers argued that an evaluation of scientific quality alone was too limited. They wanted the peer committees to take into account other performances, like the contribution to industry, to culture or societal problems. The role bibliometric indicators had acquired in the evaluation of scientific quality, made several actors look for indicators of societal quality.[8]

At the end of the second round, in 1992, universities and researchers had accepted the idea of accountability and even thought evaluations to be necessary. But they also felt that the feedback they received from peers was rather limited considering the evaluation effort. By now universities wanted to have the full responsibility for the evaluation, which they indeed got. In 1993, the responsibility for the evaluation of university research was transferred to the Association of Dutch Universities (VSNU). The VSNU developed an evaluation system that informed university and departments about the performances and the progress of the research programmes in detail. The government was informed about the quality in general terms only. Evaluations were based on a combination of self-evaluation and visitations by peer committees. This evaluation approach developed rather soon into a well-established practice. In 1998, the fourth round of evaluations started, with an even larger focus on the information needs of research management, by introducing within self evaluations mission statements and ask for mission related assessments for each research programme.

---

[8] Van der Meulen, B. J. R. & A. Rip: Evaluation of societal quality of public sector research in the Netherlands. In: Research Evaluation, 8 (2000) 1, S. 11–25.

From the overview we can conclude already two major developments. One is the change of policy context in which the evaluations took place. In fact, evaluations have moved from being a policy instrument of the government to improve higher education, research policy planning and implement accountability to an instrument of university research management. Maassen and Weusthof have called this the Dutch version of the 'evaluative state': the institutions are getting more control (again) over 'input' and 'process' of the research programmes (including it's evaluation), the government is concentrating its control on the output.[9] The second development is that evaluation of university research has evolved from a disputed practice without established methods to an accepted practice, with in many respects a clear methodology. The next section describes this evaluation practice in more detail.

3 The VSNU university research evaluation

The VSNU university research evaluation is organized as a rolling scheme of about four years. In these four years all university research programmes are evaluated. At the beginning of the four years the universities enact a protocol for the whole evaluation cycle (see Table 1). The protocol defines the disciplines and their year of evaluation, the responsibilities of every actor involved in the evaluation, the criteria for evaluation, the minimal information on which the evaluations have to be based and the procedure of the evaluation. The protocol establishes the framework for the disciplinary evaluations and assures the quality of the evaluation process. Within the framework, university departments from a discipline can agree to add components. The peer committees can be asked for instance to give assessments of the state of the art in the sub-disciplines as well, or the departments can agree to enlarge the information base and add performances that are of specific interest for the discipline. In the sciences the information of the departments is usually complemented by a bibliometric analysis.

The actual evaluation is preceded by a self-assessment of the departments, in the form of a report written in a format according to the guidelines of the protocol. The self-assessment reports consist of a description of the research programme, an overview of the performance of the last five years, future plans as well as a list of five key publications. Full publication lists are usually added as an appendix. If programmes have been evaluated before, the self-assessments should also make clear how previous recommendations have been taken up. Formally it is the university board that requests the VSNU to evaluate programmes and offers the self-assessment reports. In practice, the self-assessment reports often are directly sent to the peer

---

[9] Maassen, P. A. M. & P. J. M. Weusthof: Quality assessment in Dutch Higher Education, In: Maassen, P. A. M. & F. A. van Vught, Dutch Higher Education in Transition, Management and policy in higher education series: 11, Culemborg: Lemma, 1989, S. 129–150.

committee by the departments themselves. The actual evaluation is done by a peer committee, who makes its judgement on the base of the self-assessment reports complemented with interviews of programme leaders. In most disciplines, and certainly in those where research is laboratory based, site visits are made by the evaluation committee. In some disciplines and especially when a large number of research programmes has to be evaluated, instead of site visits, programmes leaders are asked to visit the committee for interviews.

Protocol of VSNU Research Evaluations

1. The protocol includes a classification of the disciplines as well as a rolling scheme for evaluation of the disciplines.
2. After consultation of the involved departments, the VSNU determines a time schedule for every evaluation.
3. The directly related standing disciplinary committee of the VSNU nominates two or more candidates as chairman of the evaluation committee as well as decides upon a profile of the expertise of the committee members.
4. The VSNU appoints a chairman, after consultation of the Royal Netherlands Academy of Arts and Sciences KNAW. The chairman, in consultation with the KNAW, puts together a committee of which the majority of the members are from abroad. The working language of the committee is English.
5. The directly related standing disciplinary committee of the VSNU specifies within the discipline-specific protocol the *terms of reference* for the committee.
6. Based upon the general protocol and the discipline-specific protocol the involved university departments make self-assessments of their performances of the last five years and describe their future plans. The unit of evaluation is a research programme. Of each programme, five key publications are put up as part of the self-assessment. In addition, a profile or mission statement of the department is requested.
7. The evaluation committee is requested to judge, for each programme, its quality, productivity, relevance and viability on a five-point scale. For each programme, a brief explanation of the scores is given, which might nuance the general judgement. In addition, the committee gives an assessment of the state of the art of the discipline and of each department.
8. The committee's judgements are based on documents, complemented by interviews with programme leaders and the department's management. Especially in the engineering and natural sciences the committee will make site visits.
9. The report will be finalised, stipulated and presented to the VSNU chairman, after the departments have got the possibility to react on the draft report.
10. The costs of the evaluation are covered by the universities involved in evaluation. The costs depend on the discipline and the discipline-specific protocol. The base costs are Dfl. 27.000 per university.

Table 1
Protocol of VSNU Research Evaluations (translated)

The peer committee consists of peers from abroad. Its chairman however is a (retired) Dutch researcher with knowledge of the field and the Dutch research system. The chairman has a specific role in the relationship of the committee with the disciplinary groups and with the VSNU. For the committee members, meetings are often concentrated around a few days: an introductory day to discuss such things as the process of evaluation, control information base and agree on tasks; some days with site visits or interviews with programme leaders and heads of department; a meeting for discussion of the findings and drawing conclusion. For the chairman the evaluation is more intensive. He has to meet with other actors involved in the evaluation and, assisted by a professional secretariat from the VSNU bureau, has to prepare meetings and to write the evaluation report.

Every research programme is evaluated in four aspects – quality, productivity, scientific relevance, viability – on a five point scale (excellent, good, satisfactory, unsatisfactory, poor). A comparison of the research evaluations of chemistry and of law makes clear that the protocol leaves space for disciplinary interpretations of these four aspects (Table 2). In the chemistry evaluation the four aspects were translated in specific indicators. For quality and productivity different output indicators were defined. Relevance and viability were not defined by strict (quantitative) indicators but the committee did define the information on which the judgement had to be based. Remarkably, the chemistry committee interpreted 'relevance' primarily as scientific relevance. Usually this aspect is considered to refer to societal relevance and committees are expected to judge the contribution of research programmes to socio-economic development and to issues and actors like environment, health, policy and industry. Maybe as a consequence, representatives from the engineering sciences protested when the results were published. They argued that for the research programmes in chemical engineering the committee insufficiently had taken into account the importance of these programmes for industry. Whether that is true or not, another analysis showed that by looking are patents and technological impact, the committee was at least more consistent in evaluating 'societal relevance' than the self assessment reports were.[10]

The evaluation committee for law research was less specific in the interpretation of the evaluation aspects. For quality, the committee looked at the key publication and trusted its own peer competence to evaluate their quality, or asked external peers to give their opinion. For the productivity it developed a formula by which programmes could be ranked into different categories. The other two aspects were actually not evaluated, only mentioned if programmes did very well in both respects. About 're-

[10] See van der Meulen, B. J. R. & A. Rip: Maatschappelijke kwaliteit van onderzoek tussen verantwoording en management: een inventarisatie van beoordelingspraktijken, (Societal quality of research between accountability and management: an inventory of evaluation practices), Rapport in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschappen, 1997.

Evalution aspect	Chemistry	Law
Scientific Quality	Quality of output International visibility	Quality of key publications
Scientific productivity	Number of PhD theses Number and kind of international publications Number of patents Number of invited lectures	Calculation of output according to a formula
Scientific relevance	Research topics and methods Expected impact on progress of chemistry and other sciences Expected impact on progress of technology	Relevance and viability are assessed only for those programmes that deal with particular relevant subjects and that are clearly progressing
Long term viability	Future research plans Human resources Research facilities	

Table 2
Criteria used in chemistry and law research evaluations

levance' the committee even argued that this did not belong to the 'mission' of an university: "Although it is often useful if members of the scientific staff share their knowledge with (non researching) jurists and the wider society, this kind of work is not really different from publications that many jurists themselves publish. With respect to the mission of the university such work has to be considered as a 'service to society' rather than as 'scientific research'".

The public report about the evaluation is, in general, not very detailed in its assessment of each programme. Scores are given with only a few sentences of clarification – which is however more than what was done in the early evaluation processes when programmes were only given a plus or a minus. Bibliometric profiles are presented at an aggregated level without specifying the names of the groups. More detailed findings and judgements of the committee are given in confidence to the university faculties, who may decide to publish them or not. Usually they are not, reflecting the autonomy the universities have obtained in this respect and their success in putting the government at a distance.

One of the interesting things of the evaluation process is the balance between standardised process and disciplinary input, laid down in the protocol. The basic procedures are established by the VSNU. Disciplinary Committees within the VSNU have a say in specifying the information that has to be included in the self-assessments, may come up with more precise formulation of the four aspects and may list addi-

tional issues for the committee to assess. The distinction between the four evaluation aspects has tempered a lot of the discussion about how to interpret research quality and what aspects had to be evaluated. Although in general the scores on the four aspects correlate significantly, it is obvious that for some programmes it has been useful to distinguish e.g. between quality and productivity or between past performance and long term viability. It is also interesting to mention that within the current round, the VSNU has taken liberty to introduce some experiments with user evaluations, with integration of teaching and research evaluation, and discussing possibilities to integrate or combine evaluations of research programmes and graduate schools.

4 Evaluation and the governance of research

Impacts of evaluation processes are difficult to assess. Surely, if evaluations go on for about twenty years it is always possible to tell anecdotally about measures taken that were closely linked to evaluation research. For the actors directly involved, such cases may be sufficient to prove the value of evaluations, or their danger. But in general these are not the impacts that are most interesting for understanding the role of evaluations, and if such cases go with numerous cases in which evaluations had no visible aspect, one might wonder why all this evaluation effort is needed. Looking at the Netherlands, one can indeed identify such cases, and especially within the first rounds of evaluations in the eighties. Budget restrictions and the increasing attention for research performance at that time gave legitimation to 'cutting the dead wood' – an unhappy phrase that incorrectly suggest – that the cutting went without pain.
More interesting than these anecdotes, are the more systematic impacts of the evaluations on decision making processes. Unlike the evaluations in the UK, in the Netherlands there is no direct relation between the evaluation outcomes and government allocation of funding. Basic funding for university is provided as a lump sum, which is calculated, for every university, with a formula that includes some teaching performance indicators and some historically determined elements. If not affecting directly the allocation of basic grants, do evaluation outcomes affect decision processes within the universities? This is of course a complicated issue as within universities, evaluation outcomes interfere with other policy information and processes. The evaluation of university research is not the only policy innovation that has entered the academic world. Other developments have pushed the universities to become more strategic in its research strategy, formulate priorities, identify centres of excellence etc. None of the universities has an explicit procedure by which evaluation results are translated in policies, except for some bonus funding as a reward for excellent evaluations. Nevertheless, there is evidence that there is an impact on decision processes, and in the last section I will reflect on the relation between these evaluations and some tendencies of academic knowledge production.

An evaluation of the third round (the first VSNU evaluation round) revealed that within the universities evaluation outcomes are used as a solid base for decision making.[11] It appeared to be common that reports are used incrementally in decisions on e. g. investments, new professorships and the organization of the faculty. Although it is hard to find major decisions based on evaluation outcomes only, actors surely consider the evaluation outcomes as crucial information for strategic decisions. Evaluations affect the local reputation of researchers and research groups. They give a more independent base to existing reputations of groups and in some cases correct these. Some evaluations have brought to light under-performances of politically strong groups and excellent performances of weaker groups. Consequentially, the evaluations have created their own Matthew effect.[12] The outcomes of the research evaluations and required reputations are used within other contexts and facilitate good results in other evaluations, provide access to research collaborations and key positions in the disciplinary field, and also stimulate the acquisition of competitive funding.

As important for the relation between the evaluation and the intra university decision processes, are the self-assessment reports. At the level of the university groups and the departments, the self-assessment reports often force these actors to acknowledge weaknesses. Expected criticism is anticipated and measures are taken to overcome the weaknesses. An extra stimulus for this effect is that in current research evaluations, departments are expected to indicate the consequences of previous evaluation outcomes.

5 Impact on knowledge production

If direct impacts on university decision making are difficult to assess, the more are impacts on the functioning and organization of academic research. Our understanding of the organizational, political and epistemic effects of evaluations are still based more on stories, extrapolations, interpreting reasoning than on systematic research of evaluations. One reason is that often the implementation of new evaluation practices creates a high stakes context. Another reason is that the new evaluation practices have emerged at a time, when not just the relation between government and university was changing, but the academic system as such seemed to be in transition at many levels and in many respects. As far as organization of research, the quality of

---

[11] Westerheijden, D. F.: A solid base for decisions: use of the VSNU Research Evaluations in Dutch Universities. In: Higher Education, 33 (1997) 4, S. 397–414.

[12] According to the biblical text "For to every one that has shall be given, and he shall have abundance: but from him that has not, shall be taken away even that which he has." (Matthew 25:29). See: Merton, R. K.: The Matthew effect in science, Science 159 (1968) 3810, S. 56–63.

research, universities role in society were affected by this transition – and it is quite certain they were – these effects can not simply be ascribed solely to the new evaluation practices.

Still, some conclusions can be drawn in this respect and it is interesting to put these in the context of ideas about the 'new production of knowledge.' Gibbons et al. have claimed that at the end of the 20th century the dominant mode of scientific knowledge production (which they call Mode 1 and for which physics seems to be the example) is replaced by a new mode of knowledge production, Mode 2[13]. Mode 2 knowledge production, for which the current life sciences seem to be exemplary, is characterised by five main attributes. First, knowledge is produced in a context of application, which does not rule out the possibility of fundamental or basic research, but indicates that research should be legitimatised by referring in advance to possible applications. Consequentially, knowledge production is more transdisciplinary, not only by transgressing the boundaries between disciplines but also the boundaries between empirical, theoretical and practical knowledge. The third attribute Gibbons et al. mention is the organizational diversity in terms of the sites of knowledge production, the flexible collaborations between different sites and the on-going reconfiguration of disciplines, expertise, areas in new specialities. The strong connection with societal contexts is also reflected in social accountability and reflexivity on the effects of knowledge production. Last, and of specific relevance for our discussion, is the emergence of new forms of quality control.

One can question whether Mode 2 is as new as Gibbons et al. claim and whether Mode 1 is as starving as they pretend in their essay-styled argument.[14] But a fact is that it has captured a lot of attention and recognition among science policy makers and researchers. Therefore it provides a useful framework for understanding broader impacts of the university evaluations in the Netherlands, especially on the organization of research and the position of the university in society.

With respect to the organization of research, it is clear that the introduction of research evaluations has induced changes, but into another direction than Mode 2 characteristics. A major consequence of the evaluations is the organization of research into research programmes of substantial mass and duration of 4–5 years. In the first round of research evaluation in the eighties, within especially the humanities

---

[13] Gibbons, M. et al.: The New Production of Knowledge: The dynamics of science and research in contemporary societies, London: Sage Publications, 1994.

[14] See e.g. Weingart, P.: Neue Formen der Wissensproduktion: Fakt, Fiktion und Mode, IWT Paper 15, 1997; David, P. A.: Science Reorganized? Post-Modern Visions of Research and the Curse of Success, Paper based on speech to International Symposium on Measuring the impact of R&D, Ottawa, 13–15 September 1995.

and social sciences, researchers had difficulty to come up with such research programmes. As a consequence, quite a lot of the programmes existed on paper only. Because of the limited consequences of the ex ante evaluations on decision making (as long as the programme got a 'plus'), this was not seen as a problem.
Since evaluations are done ex-post and since the idea of research programmes has been accepted almost throughout the academic world, coherence and progress of research programmes is seen as important and 'paper-programme' is a negative qualification. The actual effect on the conduct of research depends on the discipline. In laboratory-based research the need and advantage of well-functioning instruments have always been an incentive for programmatic work. In other research areas, the organization of research into programmes at least provides a structured space for interaction within research groups and sets boundaries to individual research interests. In addition it is clear that the duty of self-assessment forces research groups to think in terms of progress of research programmes, productivity and is a pressure to take up new themes and come up with viable programmes.
A related consequence of the evaluations is a strengthening of disciplinary identities. The development of clear definitions of evaluation criteria has not only strengthened the legitimacy of the evaluations, but also invoked discussions within the disciplinary bodies of the VSNU about specific disciplinary performances and identification of core journals. For disciplinary groups this has increased the certainty about the expected performances. For interdisciplinary groups and research at the edge of dominant paradigms, the formulation of such disciplinary based criteria has created the problem, whether they should change publication strategies (with a risk to loose reputation in their own interdisciplinary field) or remain loyal to their interdisciplinary mission and accept that an 'excellent' judgement will be difficult to get. The problem for interdisciplinary groups to do well in the evaluations has been acknowledged since the introduction of the discipline-wise organized evaluations, but never tackled satisfactory to all actors.
A similar problem is observed if we look at how the research evaluations affect the role of the university in society. As such, the introduction of the research evaluations is an indication of the need of universities to be accountable. According to initial aims, research efforts have become more transparent for policy and society and research programmes are good linking points for relations with industry and other research institutes. On the other hand, the evolution of the research evaluations as an instrument of university research management has made the evaluations more internal oriented. The evaluation committee for law research was, as we saw, quite outspoken in this respect. But there are other indications as well. From the public evaluation reports one cannot but conclude that in general, university research in the Netherlands is at a high level. However, two years ago, in discussions on the research funding it was still suggested – like in the early eighties – that researchers did still not perform well enough, and therefore funding had to become more competitive. In addition, we have observed that evaluations emphasise scientific quality

and productivity. Researchers and peers still seem to lack 'competency' or an 'evaluative repertoire' to assess systematically the 'relevancy' of research programmes for industry and society.
But it is fair to say that in the new evaluation round criticism on this point has been taken seriously by the VSNU. A recent evaluation of agricultural research included an evaluation report in which relevancy was assessed systematically by certain indicators and interviews with users.[15] In general, a new principle of the present evaluations is that evaluations have to be more oriented at the specific missions that university groups have set themselves. Groups can define broader mission than a scientific only, but of course have to come up with related performances.[16]

Conclusions

In this paper I have described the evolution of the VSNU procedures for the evaluation of university research. In the Netherlands, the evaluation practice has obtained a stable position within the academic world and as such it is interesting for actors abroad to look at the different elements of the evaluation practice and learn from the experiences. In addition to the description I have tried to put the evaluation practice into context and assess its role in the governance of research. First of all by analysing how it has changed from an instrument for governmental higher education and research policy to an instrument for university management. But especially how as an instrument of university management it affects decision making within universities and whether it had an impact on the modes of knowledge production within universities. The evaluation outcomes happen to be a firm base for decision-making and indirectly affect organization of the university, and allocation of resources. Looking at the characteristics of research activities, we can conclude that the evaluation practice seems to have strengthened traditional aspects of scientific knowledge production – mono-disciplinarity, science oriented, university based. The implication is not that the developments Gibbons et al. observe do not occur in the Netherlands. The emergence of the evaluation practice in itself can be seen as an example of what Gibbons et al. called a new mode of production. What we have tried to do is isolate the effects of the evaluation processes from other developments. From that perspective, we see that the new practice of quality control does

[15] See Wamelink, F. J. M. & J. B. Spaapen: De evaluatie van universitair onderzoek. Methodiek voor het incorporeren van maatschappelijke waarde van onderzoek. Den Haag: Nationale Raad voor Lnadbouwkundig Onderzoek en Commissie Overleg Sectorraden, 1999.

[16] Verkleij, A.: Because every Situation is different... A Contribution to the discussion of the pros and cons of large scale research evaluations, Paper presented to the Confederation of European Union of Rectors' Conferences, Brüssel, 6 February 1998.

not facilitate other attributes of this form of knowledge production. Isolating such effects from other developments and pressures is somewhat artificially, but at least it made clear that when evaluation procedures are designed we should not discuss methods, and accountability, but also possible impacts on the organization and functioning of university research in society.