A solid base for decisions

Use of the VSNU research evaluations in Dutch universities

DON F. WESTERHEIJDEN

Center for Higher Education Policy Studies, University of Twente, Enschede, The Netherlands

Abstract. To gain insight into the use of the VSNU research quality evaluations (since 1993) in the practice of research and of institutional management in Dutch universities, interviews were held in eight cases evaluated in the first year of this procedure. The main conclusions are that use of these research evaluations is universal, both ‘instrumentally’ (in decisions directly based on the judgements) and ‘incrementally’ (in decision-making processes not directly linked to the evaluation). Underlying this is ‘conceptual use’: an important change in deans’ and rectors’ views of their role in managing research, which they now can *realise*, because the VSNU research evaluations give them, for the first time, solidly legitimate arguments on which to base strategic decisions. Next to use, other effects can be discerned within universities, pointing to a growing dependency of researchers on managers, necessitating amongst others ever more consciously strategic publication behaviour. Whether quality of research improves in this way, cannot be answered here, but certainly it is more difficult for academics not to engage in research.

Introduction: research and evaluation

Ever since science¹ became a social activity in the seventeenth century, evaluations of research outcomes have played a role in the control of what was accepted as valid knowledge and in the distribution of scarce resources: first of all reputation, but also awards and – increasingly – money. At first, this was practically completely a matter of inter-collegial or *peers*, control. Especially since the Second World War governments started to realise the essential role of scientific development for the national capacity, in military and economic terms. In recent decades this trend was strengthened by the tremendous growth of higher education systems, leading – in Europe more than in the United States – to ever higher governmental expenditures for the university system. But apparently a ceiling of public expenditure was hit in the 1980s in this respect in many European countries: governments no longer were prepared to expand their outlays for teaching and research in higher education unconditionally – or not at all (Van Vught 1995).

For this reason, and for other reasons such as the growing desire for transparency of government expenditure and effectiveness in all areas of policy

(cf. among many others Van Vught and Westerheijden 1994), more extensive systems of evaluation were developed in higher education systems in many countries, integrated within the 'steering strategy' of the government instead of being mainly self-regulative peer control. This is quite different from science policy in general, in which setting of priorities is seen as the problem to be solved, while apparently trusting that judging quality of research is unproblematic (e.g., Van der Meulen, Westerheijden, Rip and Van Vught 1991 for the Netherlands; Anderson 1994 on the UK; Teich 1994 on the USA). In Europe this move towards evaluation of all university basic research – not just a part of it in the form of grant request from research councils, as happens 'everywhere' – shows especially in the United Kingdom and in the Netherlands (see also Irvine, Martin and Isard 1990; Brennan 1994; Hammond and Devine 1994; Hanson 1994; Teich 1994). The British 'research assessment exercises' are directly tied to governmental research funding: a single, summative rating of a unit's past research performance determines how much research money the government will allot to it for the next period. The units in the lowest category will not receive any 'untied' money for research. Whether such concentration of research funds in a small number of higher education institutions really enhances research quality is debated (Johnston 1994). And other unintended consequences also show, most negatively worded as the growth of a 'compliance culture', i.e. trying to 'beat the system' by concentrating on ways to get a higher rating without caring about real quality enhancement, for example by hiring highly published researchers just before such a national assessment exercise.

In the Netherlands – the country I shall concentrate on in this article – the government tried to re-establish a self-regulative higher education system under state auspices, without such a direct link to funding.² The aim of this article is to see which effects the Dutch approach has to evaluation of research quality, especially from the point of view of how university managers can use the evaluation outcomes.

Evaluation of research quality in Dutch universities since the 1980s

One of the most ubiquitous forms of evaluation of research quality takes place in research councils (Rip 1994). These councils distribute grants for individual research projects, sometimes also for larger entities, but the individual research projects are the main paradigm of their work. In the Netherlands, this research council is the Netherlands Organisation for Scientific Research (NWO). The evaluations concern the quality of applications, which as a rule pertain to plans for individual research projects or programmes (extensively explained in: Van de Kaa 1994).

In the 1980s, the Dutch government wanted to ‘promote both quality and systematic discussion of priorities and the use of resources’ in research in Dutch universities (Spaapen et al. 1988: p. i) – accountability regarding government funding can be seen as an ulterior goal (Bijleveld and Goedegebuure 1989: p. 101). For this goal, the NWO evaluations were not sufficient, because they covered only a small part of university research, and because they looked at quality project by project, without explicitly taking into account the development of a scientific field, or of Dutch higher education.³ Most funding for research until then was given by the government unconditionally, as part of the basic grant to the universities, which primarily depended on the number of students at the university: the more students, the more research funds. A new policy was introduced in 1983 to change this situation, the so-called Conditional Funding policy.

To enable a discussion of priorities, the numerous research projects first of all had to be collected into a smaller number of larger ‘research programmes’ under the Conditional Funding policy. ‘Programme’ was not necessarily used in Lakatos’s ‘refined falsificationist’ meaning of the word (Lakatos 1970 [1974]), i.e. a coherent set of theories and their connected research problems, but covered any grouping of research projects under a common heading:

In some disciplines a long tradition of “programming”, of developing clearly structured and planned research projects and programmes, made adherence to the rules of the conditional financing relatively easy. Some critics claimed that these rules were based on how research in the natural and technical sciences is carried out, and showed little appreciation for the different ways of working of the humanities and social sciences. (Spaapen et al. 1988: pp. ii–iii)

The other goal, to promote the quality of research, was to be pursued through external evaluations of these research programmes, which, after an introductory period of five years, might have consequences for the universities’ research funding. These evaluations should use a three-point scale (‘excellent’ – ‘satisfactory’ – ‘unsatisfactory’), and should lead to research funding reallocation: good research was to be given priority in funding, taking it away from bad research programmes (it was, after all, a conditional *funding* policy).

But evaluators in fact refused to single out excellent research, and only very few unsatisfactory programmes were identified, leading to a very low degree of resource reallocation. The most important effect of the Conditional Funding procedure, from my present vantage point, has been the introduction of the research programme as the main level of thought about research within the universities.

As a further development of the Conditional Funding procedure, in the early 1990s the coverage of research evaluations was extended to – in principle – all research in the universities, except contract research. Since the Association of Universities in the Netherlands (VSNU) already co-ordinated the quality assessment of teaching in the universities, it was agreed that it would also co-ordinate this new procedure. The quality assessment of teaching was partly modelled on the Conditional Funding procedure, in that it too was based on evaluating all programmes of study in the country belonging to the same area of knowledge (loosely: discipline). This approach was now used again to design the procedure of research quality evaluations, which is applied to the level of research programmes.⁴

Briefly, the procedure that the VSNU implements since 1993–1994 is as follows (VSNU 1993; VSNU 1994):⁵ in consultation with the deans of the faculties involved in an area of knowledge, a chairperson is chosen for the *ad hoc* peer committee. The chair is usually Dutch, e.g. a recently emeritated full professor. To assure impartiality, the other team members (four to six, as a rule) mostly are foreigners, the whole procedure being conducted in English. The faculties provide the evaluators with information covering all non-contract research in the faculty during the last five years, giving for each research group information such as:

- a summary of programme aims,
- content of the programme and main results,
- input of research staff,
- five key publications of the programme,
- a list of all publications,
- other indications of quality and reputation.

Based on this written information, complemented with interviews with the programme leaders (who as a rule are full professors), and sometimes with site visits, the peer committee then judges each research programme on four dimensions:

- quality of output,
- productivity (quantity of output),
- relevance (to the discipline, and where applicable also to technology or to society), and
- long-term viability.

The judgements are made on a five-point scale, ranging from ‘unsatisfactory’ to ‘excellent’.⁶ Next to these four judgements (they are *not* integrated into a single score), the committee gives an explanation for its judgements of the research programme in a few textual statements (often not more than about half a page in length). In a separate part of its public report, the committee also judges the faculties, the state of the field as a whole, and each of the sub-

disciplinary areas against the international state of the art. But the part on the individual research programmes is read much more, within the universities, than the general part.

For completeness' sake, let me add that, as is the case with the quality assessment of teaching, there is no *direct* link between the VSNU research evaluations and governmental expenditure for higher education institutions.

In a separate policy initiative in the early 1990s, the government introduced the concept of 'research schools' in Dutch higher education. These can briefly be characterised as serving a dual purpose: a) concentration of top-level research in an area of knowledge, often through inter-university co-operation (a sizeable 'centre of excellence'), and b) post-graduate education for the next generation of researchers by the leading researchers in the national subfield. Not all researchers can take part in a research school: in principle, only the best are invited. The same goes for research trainees. Also, research schools take the programming of research a step further: their plans and programmes should be really coherent. This implies that strategic choices have to be made within higher education institutions to initiate a research school in a particular field of knowledge, to participate in one, or to recognise that the 'centres of excellence' in the country in a particular field are not within one's own university. Additional funds are available for research schools, and the Royal Academy of Arts and Sciences (KNAW) engages in a 'recognition' procedure for these research schools. As in an accreditation, recognition, which is valid for a five-year period, is far from a foregone conclusion of the procedure. It is seen as prestigious to have a research school recognised in which one's university is taking the lead, and universities value this highly.

Evaluation of evaluation: the research project

The most extensive research evaluation procedure in the Dutch universities is the one now operated by the VSNU since 1993. After three years of experience, the VSNU wished to *assess the effects of these evaluations of research quality in the faculties and universities concerned*. To have enough time for effects to become visible, only the committees involved in the first, experimental, round of evaluations (1993–1994) were selected for the present evaluation research.⁷ With the 'disturbing factors' that may blur relationships between the research judgements and decision-making in such a three-year interval, and without a clearer definition of the effects to be taken into account (it was not desired to narrow the range of possible effects in advance too much), flexibility was the guiding principle for this research project rather than a tight theoretical and hypothetical design. However, the primary focus

was on the ‘use’ of the evaluation process and of its outcomes by managers in faculties and at the central university level.

Accordingly, relevant documents were studied, and – as the core of the empirical work – open interviews were held with decision-makers at the faculty (in general the faculty dean or the vice-dean for research) and university levels (the rector magnificus) in two cases for each of the four evaluation committees concerned.⁸ Interviews with faculty and university representatives were held separately (with one exception), to enable different points of view to be presented unhindered. Apart from ensuring maximum spread over universities, the cases were chosen ‘blindly’, without knowledge of either the judgements of the peer committees, or the use and other effects of the evaluations. Respondents, especially at the university level, tended to answer the questions drawing upon all of their experiences with the VSNU research evaluations, making their responses more generally relevant than just for the committees officially involved in the research project.

Use and other effects

Passive and active use

Given the primary interest in ‘use’, but with a broader interest in ‘effects’, terms were not to limit the possibility of unexpected findings on beforehand. Therefore, it was decided to choose a broad categorisation of concepts. First, I distinguish passive use from active use, in line with Frederiks, Westerheijden and Weusthof (1994).

Active use is defined as making decisions in which the outcomes of the research evaluations are an important argument. Within this category, *instrumental use* is seen as decisions directly linked, in intent and in time, with the evaluation outcomes (judgements and – implied – recommendations), while *incremental use* denotes decisions made at a later time (operationally: more than a year after publication of the VSNU committee evaluation report), or not directly linked to the evaluation outcomes (Hoey 1995: pp. 38–39). As a rule, in incremental use the evaluation outcomes are only one among a larger set of arguments. Instrumental use is what is expected in a simple, linear decision-making model. In a more complex view of social systems, with iterative decision-making, with a multi-actor perspective and acknowledging the cultural aspects of behaviour (such as the construction of meaning of research evaluations), next to active use, passive use should also be expected.

Passive use is, first, at the basis of active use, in that it denotes gaining knowledge of evaluation outcomes, and perhaps discuss it in meetings, but not (yet) make any decisions in which the research evaluations play a role.

Next, passive use can be *persuasive use*, to convince other actors to make a certain decision.⁹ Finally, passive use can be *conceptual use*; this concerns the way actors think about their work and their organisation. The latter may be the most important type of use, because it may open up completely new frameworks for behaviour of actors in higher education. But it is almost certainly also the most difficult to discern, because actors may be unaware of a change in their mental frameworks.

Some other effects

Although the primary focus of the research project is on managerial use of the research quality evaluations, I realise that other effects result from the evaluation process as well as from the evaluation outcomes. While maintaining the focus on actors within the higher education institutions (effects on governmental actors, and systems-level effects remain out of my reach), a number of possible areas of effects are distinguished. The choice of areas is based on insights from higher education studies, sociology of science, and organisation theories (Goedegebuure and Westerheijden 1991: p. 515; Van der Meulen et al. 1991; Spaapen et al. 1988; Whitley 1984; Becher 1989; O'Neill and Meek 1994: p. 99, among others).

The main point here is that research in a university setting is a social process, in which researchers depend on colleagues and administrators for input (knowledge to build upon, money and facilities to work with, etc.), for throughput (especially in co-operative research projects), and for output (e.g., places to publish research outcomes). The feedback part of the process consists of peers' comparison of reputations, leading to awards, money and other inputs for a new round of research (Latour and Woolgar 1979: p. 201). The degree of dependence of researchers on others varies across disciplines (Whitley 1984; Becher 1989), but might be affected by changes in boundary conditions, some of which are aimed at by the policy initiatives studied here. Therefore, most of the areas mentioned below are connected to the researchers' dependence on other actors, and on the consequences for relationships between researchers and these other actors (mainly the managers of the higher education institutions). The areas distinguished are:

- a) co-operation among researchers within faculties;
- b) research praxis, especially publication behaviour;
- c) the relationships between and among researchers, programme leaders, faculty managers and university managers;
- d) quality policies at faculty and university levels;
- e) faculty and university management, especially internal resource allocation, personnel policy and coherence within research programmes;

- f) the position of research groups, faculties and universities *vis-à-vis* external actors (especially third party funding: research councils and commercial enterprises);
- g) finally – tentatively – whether improvement of quality of research can be discerned.

Findings I: Types of use

Passive use

Management of research quality in Dutch universities has considerably increased since the early 1980s. External incentives triggered this change: first the Conditional Funding procedure, then the VSNU research evaluations and the research schools. The Conditional Funding procedure led to a greater degree of programming of research, and although the evaluations of programme quality were not very successful *per se*, they helped prepare the ground for evaluation (cf. also Spaapen et al. 1988). Excellence now can be shown, which is new in the traditionally egalitarian Dutch culture.

In all cases that were investigated the VSNU evaluations lead to some types of use, both passive and active. Passive use is a necessary prerequisite for active use, because one cannot act upon something without knowing about it. *Passive, formal use* is part of the standard operating procedures of all universities investigated. This takes the form of incorporating the VSNU evaluation reports in the regular talks between the Faculty Board (dean) and the university's Governing Board. If the VSNU research evaluations are seen as an innovation, this testifies to their 'compatibility', one of the factors enhancing an innovation's chances for success (Rogers 1983: pp. 220–221). But *passive use* is not confined to administrators: all actors in the universities see evaluation reports as "hot stuff" (W1)¹⁰ to be read immediately: "People were anxiously waiting for the report . . . It was well-consumed, here" (P2) . . . and that showed. Some talked about it, others did not." (W1) The talking depending on whether one's own research group was evaluated positively (W1, P1, P2).

'Program review is an excellent mechanism for providing a new administration with unbiased information on the status of each department in the institution.' (Mets 1995: pp. 23–24); this applies equally to established administrators and to other levels of the university. All administrators crave for unbiased, externally validated information about the standing of their units. According to the respondents the VSNU research evaluations did not often bring completely new information to the light: "External judgements legitimise what you, as administrator, know" (P2). *This external legitimisation of*

reputations may well be the most important function the VSNU research evaluations fulfil. The degree of legitimisation depends on the *credibility* of the evaluations, which in its turn depends on several factors, the most important of these being – as appears from the interviews:

- a) the reputations of the evaluators and the composition of the evaluation committee,
- b) the fact that these evaluations cover practically all research of the evaluated,
- c) the differentiated judgements (on a four- or five-point scale), and
- d) the fact that they are public.

The credibility of the judgements also depends on the concurrence between these judgements and previously existing reputations: in one case, unexpected and probably unfounded judgements of 10% of a faculty's research undermined the credibility of the judgements of the other 90% (B1); in another case, on the contrary, the concurrence of most judgements with what was believed before helped to convince the administrators of the remaining judgements (W1). The verbal argumentation in the evaluation reports can help to bolster the judgements' credibility; again B1 was a contrary case.

Active use

Through legitimising differentiated degrees of quality of research groups, the VSNU research evaluations for the first time give university administrators the *actual* possibility to make decisions that differentiate between research groups – this is *the most important consequence of the legitimisation function*. As a result of the VSNU evaluation report and the talks between the Faculty Board and the university's Governing Board, the faculty is in all cases expected to draw up a plan for action. Here we enter the realm of *active use*.

Instrumental use, some form of which is found in all cases, can be either positive (rewards for 'excellent' judgements, often in the form of additional temporary researchers' posts), or negative (punishments for 'weak', sometimes even also for 'average' judgements, in the form of different degrees of restructuring of research groups or their research programmes).¹¹ Some universities are, until now, somewhat hesitant about giving out positive sanctions, because they wish to have an overview of all judgements before committing themselves – being too generous after the first verdict of 'excellent' can be seen as setting a precedent (a.o. B1, G1, W2). Some other universities are in a process of enlarging their central-level capacity for differentiated funding allocation (P1, P2).

The borderlines between very good, sufficient and weak groups were drawn differently in different universities,¹² based on institutional aspiration levels, interpretations of the meaning of the actual wording of the ratings, and with

some awareness of the unavoidable rests of subjectivity in committee decisions: a '4' could have been a '5', or the other way around, if other persons had been on the evaluation committee. All respondents agreed, however, that such a threefold distinction must be made, that at least these three categories were not too susceptible to subjectivity, and that these judgements should have positive and negative consequences for institutional decision-making.¹³

Incremental use seems to be even more common than instrumental use. Incremental use appears in all kinds of decisions that are somehow connected to research, resource allocation or personnel policy in the universities. The VSNU research evaluations imply "pressure on the faculties to tackle eventual problems" (B1). The largest decision-making processes I came across in which these evaluations play a role, concern restructurings of faculties, usually triggered by diminishing numbers of student enrolments, and guided by arguments of quality of research groups. The central role of quality in all these decisions should not overshadow other arguments that also play a role. As a rule, rectors and deans said they based their management decisions on strategy, profile and quality. Sometimes, the role departments play in the teaching of a faculty is a reason to protect research groups that perform less than satisfactory.

A decision of major importance for the profile and quality of a faculty occurs when a new professor can be appointed; in those, only partly foreseeable, occasions the VSNU research evaluation reports are often consulted, both for a view of the profile of the faculty and where possible to get some insight into the quality and productivity of candidates for the post.¹⁴

As a type of incremental use *ex ante* one could distinguish *anticipatory use*. Respondents recognised that knowing that their research would be evaluated has a "preventive effect" (P2), it "keeps things in order" (G1) and "prevents introversion" (W1). Anticipatory use includes ameliorative action before an evaluation committee comes, but may also imply the (inadvertent or strategic?) mixing of weaker research through stronger groups, making weak research less conspicuous, or even leaving some parts of research completely out of the evaluation procedure.¹⁵

Finally in this section, it should be mentioned that the evaluation results also become known outside of academe. The judgements produce a "halo effect" (B1) on professors' reputations in society, leading to, e.g., changes in earning capacity on the contract research and development market, even though the judgements are intended as purely scientific judgements.¹⁶

Conceptual use

In some cases, respondents mentioned that "For us, not much has changed. All our research was judged positively" (P2).¹⁷ Meanwhile, in this particular

faculty 'excellent' research groups were given preferential treatment in grant proposals, and leaders of research groups that were 'just' 'good' wonder how they can get a '5' next time – thought and behaviour have changed in this faculty, apparently without the respondent noticing. This is an example of conceptual use, and simultaneously of the difficulty to pinpoint it empirically.

The conceptual breakthrough, I gather from the interviews nevertheless, resulted from the Conditional Funding procedure in the 1980s; that policy changed the complete *discours* of research and its management. The Conditional Funding procedure also popularised the notion that research projects can be collected into programmes, and that such programmes could be used for management decisions. Also, it loosened the traditional administrative axiom of qualitative homogeneity a little, but still the differentiated reputations, which of course existed, were not sufficiently legitimised for administrative decision-making. In this respect the VSNU research evaluations were more effective than the Conditional Funding, as explained above (cf. also Spaapen et al. 1988).

Findings II: Effects on what?

To get a view of other effects of the VSNU research evaluations than their use in university management, I shall now elaborate the areas of possible effects mentioned above. As these areas only make up the secondary level of attention of this research project, I treat them more sketchily than use of the evaluation.

Co-operation among researchers

The degree of co-operation between researchers to a large extent depends on characteristics of the discipline (cf. Whitley 1984). Evaluation procedures cannot influence such compelling reasons very much. Yet, as mentioned in the previous section, the Conditional Funding procedure popularised the notion of research programmes, and introduced them even in disciplines without such traditions, especially in the social sciences and the humanities. At present, the most important structuring element in co-operation among researchers are the research schools. Accordingly, I conclude that the interdependence of researchers has grown during the last ten to fifteen years, but that the VSNU evaluations do not play an important role in this respect, because other policies affect this area more.

Publication behaviour

In view of publication habits, the VSNU research evaluations continue the trend started with the Conditional Funding. The pressure to publish, and to publish ever more 'strategically' in international (English-language) journals with high impact factors according to the ISI indices, increases in all types of disciplines (especially clear in W1, G2). Researchers' publication behaviour becomes more and more consciously strategic.

Relationships between and among university actors

The most important function of the VSNU research evaluations is to increase the legitimacy to make differentiated decisions – it is not a new tool for university managers, but a way to make existing tools effective (see also the next two subsections). In this way, the balance of power within universities is changed in favour of administrators, both at the faculty and the university levels. This brief conclusion indicates a very important change in academic processes; it signals the change from what traditionally was a fragmented university dominated by the chair-holders,¹⁸ to an ever more administratively integrated organisation – a tendency, in Kogan's words, towards 'managerialism'.

Quality policies at faculty and university levels

The VSNU research evaluations for the first time effectively enable the design and implementation of differentiated quality policy at faculty and university levels. Thus, they provide an impetus to the development of quality policies, as can be seen also in the previous and next sections.

The effort that it takes to prepare one's faculty for the external evaluation committee is sometimes mentioned as a burden, but less often than when the self-evaluations for assessments of teaching are concerned (cf. Frederiks et al. 1994). In combination, however, the VSNU research evaluations, evaluations by the Royal Academy of Arts and Sciences, by the National Organisation for Scientific Research, by other research councils, etc. (not to mention the quality assessment of teaching), are felt to produce an 'evaluation overload' that limits the acceptability of all these quality management activities for 'normal' academics. This may become a serious problem, as acceptance of quality evaluations by the academics is a crucial factor in their effectiveness concerning quality improvement. How to minimise the burden of evaluation while stimulating information-based management for improvement and simultaneously discharging the accountability demanded by government and society, is as yet an unsolved question.

Faculty and university management

The VSNU research evaluations in themselves are not new management tools, but they enable the use of other management tools. In most cases, until now no new tools were developed: the possibilities given by the regular talks taking place between the management of the university and of the faculty (a standard operation procedure in all Dutch universities) appear to be sufficient. The importance of such talks has increased, however, through the additional legitimised information that is now available about research performance. Accordingly, they tend to lead to more managerial activity than before in the faculties, mainly through ‘action plans’ that are regularly required of them by the university management to follow up on the negative (and positive) VSNU judgements.

In some cases, however, more modern management tools are introduced, such as management contracts or multi-year covenants between a university and its faculties.¹⁹

Direct consequences of evaluation outcomes for governmental funding of universities do not exist in the Netherlands, and are not desired by respondents. Research councils and the Royal Academy, it is reasoned by some respondents, can afford to link their project evaluations with finances, because for a large part of their research universities can always fall back on their basic government grants. But if the governmental grant itself would become insecure, an ‘English situation’ would arise: in the United Kingdom, the research assessment exercise may lead to a complete loss of research funds for a faculty. The complete loss of a secure stream of research funds would seriously undermine the Humboldtian unity of research and teaching, it is felt, and thus threaten the basic distinction between universities and other types of higher education institutions.

Until now, the quality assessment of teaching programmes and the evaluation of research, both co-ordinated by the VSNU, are separate procedures. Some faculty and university managers regret that in this way they may get incompatible signals about a faculty and they therefore advocate an integrated approach.

Position vis-à-vis external actors

Universities do not actively disseminate the evaluation reports outside of the institution. The existence of such reports in itself is a proof of transparency of the higher education sector (B1); whether this improves the social appreciation of the higher education sector remains an open question.

In this way, it appears from the present research project that universities do not actively try to influence individual external actors through the reports.

A “halo effect” nevertheless exists, affecting professors’ reputation in the eyes of external actors, which influences their earning capacity for contract research, consultancy, etc.

Improvement of quality?

The ‘proof of the pudding’ of research quality evaluation is whether improvement of quality can be discerned. The present research project cannot answer this question definitively; it would be hubris for any single researcher to try to do so. It is clear, however, that the continued existence of ‘weak’, non-publishing researchers cannot remain hidden any more from *le regard d’autrui* because of the public evaluation reports. It would not seem to be too speculative to expect that as a result of that, non-publishing researchers would be ‘weeded out’ in the long run. Whether more publishing researchers, hence more publications, also means *better* science, is quite a different matter. However, the drive towards better research seems to be a consequence of external evaluations. In the words of one respondent – with a catch at the end:

Research leaders are thinking about how they can score better next time. Because some were disappointed to be not ‘excellent’ but just ‘good’. Others were relieved by the same score (P2).

Conclusion

The VSNU research evaluations have gained a place in the management of universities in the Netherlands. They are not the first policy initiative in this area, because there was the Conditional Funding procedure that started in 1982, nor are they at present the only policy initiative structuring the universities’ research management, because there are also the research schools. The ‘niche’ that the VSNU research evaluations occupy is that they give the most comprehensive judgements of research within the universities; moreover, they do this in a way that is accepted as valid throughout academe. In this way, the VSNU research evaluations provide managers or administrators of Dutch universities with the solid base of legitimacy necessary for them to make decisions. As one respondent said it: “You almost cannot do without them any more” (G1).

Acknowledgements

This research project was subsidised by the Association of Universities in the Netherlands (VSNU). The author solely is responsible for this publication;

the views in this paper are not to be read as official views of neither the VSNU nor CHEPS. I wish to thank an anonymous reviewer of Higher Education for the comments on an earlier version of the manuscript.

Notes

¹ In this article I use 'science' in its broad, Continental meaning, to denote all types of scholarly research, encompassing the physical and life sciences as well as medicine, engineering, social sciences and humanities.

² The same difference of approaches between these two countries can be found in the quality assessments of teaching, cf. Goedegebuure, Maassen and Westerheijden 1990 or Westerheijden, Maassen and Brennan 1994.

³ The advisory mechanisms to forecast and influence developments of scientific fields in the Netherlands included permanent advisory boards to the Minister of Education and Sciences, such as the Advisory Council for Science Policy (RAWB), and temporary 'forecast committees' (see Martin and Irvine 1989; Van der Meulen et al. 1991).

⁴ Since in some cases these programmes are indistinguishable from the departments ('vak-groepen') performing them, I shall mostly use the neutral term 'research group' throughout this text.

⁵ The following is general enough to apply to all of the VSNU's regular research evaluations, although minor adaptations were made in 1994 after the first, pilot, year. In details, the committees may deviate from this 'general protocol' through 'discipline protocols'. Exceptions to this standard VSNU evaluation procedure are some 'orchids' (faculties without equivalents in the Netherlands) whose teaching and research have been evaluated simultaneously. Moreover, a large-scale experiment with such integrated evaluation takes place in 1997 for the programmes in language and literature.

⁶ However, in the first year, which is studied here, the procedure was still being experimented; e.g. four- and five-point scales were used, and the number and wording of the dimensions was not uniform across committees. One evaluation committee in humanities refused to use *explicit* judgement scales, but if one studied the report closely, one could read that "Saul hath slain his thousands, and David his ten thousands" (as one respondent quoted Samuel 1, 18: 7 in an interview).

⁷ The procedure itself was evaluated immediately after the first year, leading to relatively minor changes in the VSNU research evaluation procedure as it functioned from 1994 onwards (Werkgroep Kwaliteitszorg Onderzoek 1994). I assume that the *effects* of the evaluation are not affected by this adaptation of the *procedure*.

⁸ In three of the university-level interviews and in one faculty, respondents were members of the permanent staff of these decision-makers. The two cases of mechanical engineering were in technical universities, the six others (biology, psychology and history) involved six general universities.

⁹ As persuasive use empirically surfaces in active use, if it is addressed to other actors within the university, I shall not discuss persuasion separately, but see also the section 'Position *vis-à-vis* external actors'.

¹⁰ Quotes from interview transcripts will be put between double quotation marks followed by a code of the case in parentheses. Translations by the author – except for "hot stuff", which was said in English. Note that there were two respondents in each case; quotes and codes thus do not coincide with individual respondents.

¹¹ As dismissing individual researchers is very difficult in practice, negative and positive sanctions are mostly those, on a collective level, mentioned above.

¹² Only in one university was it completely left to the individual faculties to decide on their own standards of quality. In the others, either the Governing Board had its factual standards,

or decided in consultation with the faculty concerned which research groups needed positive or negative sanctions.

¹³ Respondents did not all agree about which dimension of judgements is crucial; quality of output (as in case P2), a group's viability (as in W2), or both (as in G2). None mentioned productivity or relevance.

¹⁴ Although the judgements pertain to research groups, names of the group leaders (as a rule full professors) are mentioned in the VSNU reports.

¹⁵ Clearly there are university-internal visions about strong and weak research before external evaluations take place, sometimes but not always based on a tradition of internal research evaluations.

¹⁶ The theoretical importance of reputation in diminishing the need for extensive information as a base for behaviour has been stressed before (Stigler 1961; Lieshout 1995; Westerheijden 1988).

¹⁷ As an aside, it is interesting to note that many respondents have formed a general, coherent view of research quality in their institution – usually one that was 'better than average' – even though the actual VSNU research evaluations are on the level of basic research groups or programmes, and practically always range from very positive to very negative between units within each faculty. Does this tendency arise out of psychological parsimony or necessity, or does it grow from persuasive purposes?

¹⁸ It may be good to remember here Waltz's (1979: p. 58) observation that: 'A subsystem dominant system is no system at all', with all its advantages and disadvantages.

¹⁹ Note that on a national scale, at least until 1996, budgets in the Netherlands are single-year budgets, without certainty for next years, contrary to the Swedish and French multi-year contracts between national ministry and universities.

References

- Anderson, J. (1994). 'Foresight analysis in research funding: A UK experiment', *Higher Education* 28, 39–57.
- Becher, T.A. (1989). *Academic Tribes and Territories: Intellectual inquiry and the cultures of disciplines*. Milton Keynes: SRHE and Open University.
- Bijleveld, R.J. and Goedegebuure, L.C.J. (1989). 'The paralysing effects of a balanced power system: an analysis of the introduction of the conditional funding system in university research', in Maassen, P.A.M. and Vught, F.A. van (eds.), *Dutch Higher Education in Transition: Policy-issues in Higher Education in the Netherlands*. Culemborg: Lemma.
- Brennan, M. (1994). 'Excellence and relevance – two sides of the same coin', *Higher Education* 28, 129–135.
- Frederiks, M.M.H., Westerheijden, D.F. and Weusthof, P.J.M. (1994). 'Effects of quality assessment in Dutch higher education', *European Journal of Education* 29, 181–200.
- Goedegebuure, L.C.J. and Westerheijden, D.F. (1991). 'Changing balances in Dutch higher education', *Higher Education* 21, 495–520.
- Goedegebuure, L.C.J., Maassen, P.A.M. and Westerheijden, D.F. (eds.) (1990). *Peer Review and Performance Indicators: Quality Assessment in British and Dutch Higher Education*. Utrecht: Lemma.
- Hammond, L.S. and Devine, S.D. (1994). 'The New Zealand experience of a comprehensive fully costed science funding system', *Higher Education* 28, 119–127.
- Hanson, R. (1994). 'Allocation and evaluation: The approach at the social sciences and humanities research council of Canada', *Higher Education* 28, 109–117.
- Hoey, J.J., IV (1995). 'Organizational factors in program review', *New Directions for Institutional Research* 86, 37–47.

- Irvine, J., Martin, B.R. and Isard, P.A. (1990). *Investing in the future: An international comparison of government funding of academic and related research*. Aldershot: Edward Elgar.
- Johnston, R. (1994). 'Effects of resource concentration on research performance', *Higher Education* 28, 25–37.
- Kaa, D.J. van de (1994). 'Picking the winners by consensus: Grant-giving practice in the Netherlands', *Higher Education* 28, 59–83.
- Lakatos, I. (1970 [1974]). 'Falsification and the methodology of scientific research programmes', in Lakatos, I. and Musgrave, A. (eds). *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.
- Latour, B. and Woolgar, S. (1979). *Laboratory Life*. Beverly Hills: Sage.
- Lieshout, R.H. (1995). *Between Anarchy and Hierarchy*. Aldershot: Edward Elgar.
- Martin, B.R. and Irvine, J. (1989). *Research Foresight: Priority-setting in Science*. London: Pinter.
- Mets, L.A. (1995). 'Programme review in academic departments', *New Directions for Institutional Research* 86, 19–36.
- Meulen, B. van der, Westerheijden, D.F., Rip, A. and van Vught, F.A. (1991). *Verkenningsscommissies tussen veld en overheid*. Zoetermeer: Ministerie van Onderwijs en Wetenschappen.
- O'Neill, A. and Meek, V.L. (1994). 'Academic professionalism and the self-regulation of performance', *Journal of Higher Education Management* 9, 97–113.
- Rip, A. (1994). 'The republic of science in the 1990s', *Higher Education* 28, 3–23.
- Rogers, E.M. (1983). *Diffusion of Innovations*, 3rd edn, New York: Free Press.
- Spaapen, J.B., van Suyt, C.A.M., Prins, A.A.M. and Blume, S.S. (1988). *Evaluatie van vijf jaar Voorwaardelijke Financiering: De moeizame relatie tussen beleid en onderzoek*. Zoetermeer: Ministerie van Onderwijs en Wetenschappen.
- Stigler, G.J. (1961). 'The economics of information', *Journal of Political Economy* LXIX, 213–225.
- Teich, A.H. (1994). 'Priority-setting and economic payoffs in basic research: An American perspective', *Higher Education* 28, 95–107.
- VSNU (1993). *Quality Assessment of Research, Protocol 1993*. Utrecht: VSNU.
- VSNU (1994). *Quality Assessment of Research, Protocol 1994*. Utrecht: VSNU.
- Vught, F.A. van and Westerheijden, D.F. (1994). 'Towards a general model of quality assessment in higher education', *Higher Education* 28, 355–371.
- Vught, F.A. van (1995). 'Dangerous curves: An international analysis of participation growth and higher education policy in seven countries', *Industry and Higher education* August 1995, 198–204.
- Waltz, K.N. (1979). *Theory of International Politics*. Reading, MA.: Addison-Wesley.
- Werkgroep Kwaliteitszorg Onderzoek (1994). *Onderzoekbeoordelingen 1993: Evaluatie van de proefronde*. Utrecht: VSNU.
- Westerheijden, D.F. (1988). *Schuiven in de Oosterschelde: Besluitvorming rond de Oosterschelde 1973–1976*. Enschede: Universiteit Twente, Faculteit Bestuurskunde.
- Westerheijden, D.F., Brennan, J. and Maassen, P.A.M. (eds.) (1994). *Changing Contexts of Quality Assessment: Recent Trends in West European Higher Education*. Utrecht: Lemma.
- Whitley, R. (1984). *The Intellectual and Social Organization of the Sciences*. Oxford: Clarendon Press.