12 Peers, Performance and Power
*Quality assessment in the Netherlands*

Don F. Westerheijden

12.1 Introduction

Quality assessment in higher education is linked inextricably with both peer review and performance indicators. Yet quality assessment systems in various countries stress the two methods to very different degrees. In this chapter the operation of the Dutch system of quality assessment will be reviewed, seeing it as a system of quality assessment in which peers play a predominant role and intersubjective information in the form of performance indicators is much less important. Prior to the description of a number of quality assessment processes taking place in the Netherlands in section 4, I shall go into the concepts of quality and the goals of the higher education system (section 2), and the interest which actors may have in quality assessment (section 3). A sketch of an explanation, derived from insights about strategic necessities besetting higher education institutions in the processes of quality assessment, will be presented in section 5.

12.2 What is quality?

Definitions of quality are not often given, and, if given, are often vague, because quality is a concept that can be applied to almost any kind of 'product'. In current definitions of quality, the proportion of goal attainment is mentioned as the characteristic element. An equally current definition, that for all practical purposes is equivalent stresses the degree to which the 'product' does what it is intended to do. Both definitions link quality with goals (see also: Cave *et al.*, 1988: 29; De Weert, 1990: 59 ff). For that reason, quality has been called a relative concept: it can be defined operationally only in relation to a set of goals. In the present case, these goals are the goals of higher education and research. Quality also has been called a multi-dimensional concept. In terms of goals this may mean either
that quality is measured relative to the multiple goals which higher education has to satisfy simultaneously, or that it can be measured relative to different sets of goals - sets of goals held by different actors. Conforming to this use of the concept of quality as a relative and multi-dimensional concept, I adopt a multi-actor perspective, placing quality assessment in the environment of the social and political processes that bring these actors together (see also: Van Vught & beatenheijden, 1989).

The question of collective goals of a system is always something of a problem. If it is assumed - as I do here - that collectivities can have goals, the question remains: which collectivities? When trying to explain the mechanisms of a system, it is not useful to consider the system as a unitary whole which has goals as a whole. Rather, one should consider the goals of the constituent elements, or actors. The higher education system is a complex multi-level system because it encompasses many actors with very different interests and standpoints, operating on different levels, such as: individual scientists and teachers, the national subfields of scientists of a discipline, administrative staff of faculties and institutions, faculties and institutions as units, intermediate organisations of higher education institutions (the Association of Co-operating Universities in the Netherlands [VSNU] and the HBO-Council, the association of HBO-institutions), (potential) students, employers, the professional fields and the funding bodies - in the Netherlands: the government (see also: Pollitt, 1990: 63).

The higher education institutions themselves are characterised by a matrix structure consisting of an academic and an administrative dimension. Within the administrative dimension conflicting interests exist between the central level of the university and the decentralised level (the faculties). At the central level corporate policies are conceived, the budget is obtained from the Minister (which accounts for about 90% of the total budget of a university) and it is divided among the faculties. The Minister’s money comes as a lump sum, providing the central level of the university with great leeway in the process of dividing the budget among the faculties.

The disciplinary dimension enters the university predominantly at the level of the department or the research groups (Spaanen et al., 1988: 14, 18). The national subfields of scientists from a particular discipline can put demands on members of departments or research groups that conflict with those of the administrative unit. The relative power of the demands and goals of the disciplinary dimension vary with the mutual dependence among researchers in a discipline; in paradigmatically homogeneous disciplines the power is high, in less coherent disciplines it is low (Whitley, 1974). In the first case disciplinary goals may be more important than the institution’s goals, in the latter case this is not to be expected.

Traditionally, the Dutch university organisation closely fitted the model of the ‘chair-based system’ (Clark, 1983). This implied that the balance of power within the universities was tilted towards the chair holders, the professors. Goals of higher levels of the institutions counted less than the professors’ goals. Since about 1970 the balance of power has shifted towards the middle levels of the institution, i.e., the research groups and the faculties (Spaanen et al., 1988: 13-14). In recent years the central level of the institution has gained more power as a consequence of the tighter financial conditions. Accordingly, administrative goals have acquired more weight in decision making in higher education institutions. Significant differences exist, however, in the local situations: in some institutions the administrative staff ‘always’ has had more power vis-à-vis the academic staff than in others.

HBO-institutions have a comparable administrative structure, with a central governing board, a middle level (such as economics, engineering, or health) and a lower level of study programmes. The middle level, which includes one or several comparable study programmes, bears the name of ‘sector’. The HBO-institutions developed this multi-sectoral form only recently, during a large restructuring operation in 1983-1987, ‘Scale-enlargement, Task allocation and Concentration’ (STC) (see: Goedegebuure, 1989; and chapter 1).

12.2.1 Goals of the higher education system

The Jarrett Report states that in the United Kingdom ‘objectives and aims in universities are defined only in very broad terms’ (cited in: Cave et al., 1988: 29). For Dutch higher education institutions the situation is no better than for the British: in this country, too, goals of higher education institutions are not, or only vaguely, articulated. The following statements therefore are tentative.
Officially, the goals of higher education are hardly a matter for discussion. They seem to be self-evident and everyone seems to agree on them. One of the goals of higher education appears in the title of a policy paper: *Hoger onderwijs voor velen*—higher education for the masses (Ministry of Education and Science, 1977). The massification of higher education is not (or not only) a natural trend, but (also) an explicit aim on the government's side. In this context it is remarkable that entrance to universities is free for all graduates from the appropriate types of schools (and others with effectively the same level of education), while HBO-institutions can apply entrance selection procedures.

Other goals at the moment are typically found in one breath—quality and efficiency. The macro-economic effect of higher education is important for the government as well: higher education is seen as a preparation for work (to paraphrase the title of Boys *et al.*, 1988). More precisely: higher education has to qualify students for the autonomous practice of a profession requiring academic qualifications. Finally, the government has been trying to get more insight into what happens in higher education and research. This might be called the derived goal of making higher education and research more accountable to society (Spaanen *et al.*, 1988: 12).

Before approximately 1980, equality (especially of entrance opportunity) was at least as important as quality in the government's set of goals. This goal has not disappeared, but it is stressed much less since quality has been placed in the limelight. The importance of quality and efficiency has been recognised especially since the early eighties, when a contracting higher education budget and an economic recession made it clear that—in the first place—without efficiency, the higher education budget would take up more and more of the government's resources without returning a sufficient proportion of graduates in good time. In the second place, it became clear that without quality, higher education would not work as a motor for economic recovery and growth, which is, of course, also a government policy goal, albeit not directly pertaining to higher education. In line with this, the government's goals regarding research in higher education institutions stress social relevance and more particularly their stress innovation for economic growth. This leads to an accentuation of applied and strategic research.

The 'traditional' goal of university education can be described as education for a small number of students. Such 'traditional' education is directed towards curiosity and individual development, which prepares the student for fundamental research, the central value of which is the discovery of new knowledge about why things are as they are. These 'traditional' goals are hardly ever heard about in the Netherlands, neither in the higher education institutions, nor in the government.

It cannot be expected that higher education institutions subscribe to the government's goal of efficiency, at least not in the sense that they will wish for smaller budgets for themselves. The budget maximisation goal of 'bureaus' (higher education institutions) in their relationships with 'politicians' (the Minister and the parliament) has been stressed in administrative science at least since Niskanen (1971). From a mild form of conflict, the budget problem can become acute when the government is trying to cut back severely, or is even trying to close certain faculties or institutions—provided the threatened parts of the organisation are valued positively at the central level of the institution.

The intermediate organisations are in most respects organisations of the central administrations of higher education institutions. Their views coincide most with those of the Executive Boards. Acting system-wide, on a national scale, some aspects of the views of the Ministry of Education and Science can be expected to emerge in the goals of VSNVU and HBO-Council as well.

The different categories of actors have different goals in their interactions with the higher education system. Apart from those mentioned above, one can think of goals of students (career preparation, individual development) and employers (useful graduates, applicable research results), etcetera. Moreover, categories of actors may be internally divided about goals. Consequently, not only is there no single authoritative definition of quality, but also there never can be one. What is going to be assessed in quality assessment depends, therefore, on who will do the assessment. In the next section I shall first address the question: who will want to do the quality assessment in higher education?
12.3 Why control quality?

Why would an actor in the higher education system be interested in quality? And which actors are interested, in fact?

Government certainly is interested in quality. And for a legitimate reason: it pays for 90% of all higher education expenditure and wants higher education to be accountable for the way tax money is spent. This leads to a view on quality assessment in relation to efficiency matters. Such feelings also appear to have given an important impetus to the quality movement in other countries, such as Great Britain (see, among others: Cave et al., 1988: 9–10, 30; Sizer, 1989a: 6; Pollitt, 1990; and chapters 2 and 11). The Dutch government, in 1986, introduced an Inspectorate for Higher Education to strengthen the external evaluation of education and to ensure that quality assessments will be an input into the policy development process, organised around the Higher Education and Research Plan, the HOOP planning cycle (see chapter 4).

Other actors are interested in quality control as well. The VSNU and the HBO-Council may want all higher education institutions (their constituent members) to have a good reputation in society, which would result in a higher reputation for themselves as well.5

Higher education institutions are, in principle, interested in quality. A better ‘product’ is always preferable, provided all other circumstances remain the same. These other circumstances can be interpreted as the costs and benefits of producing more quality, to which I shall return in section 5. Because of the lack of goal statements by institutions, insight into what are the institution's 'products' is wanting. Therefore, it is not clear what kind of quality assessment the institutions would advocate. As management tasks are important at the central level of institutions, efficiency will be an important element of it.

The interest in quality at the faculty and lower levels is rooted in the evaluation processes that already exist in the disciplinary fields. This implies that the interest in quality assessment at these levels depends on the scientific or professional traditions in the field, which differ according to the paradigmatic coherence of the field. It must, moreover, be expected that in the actual operation of higher education institutions, goals shift or are displaced just as it happens in other organisations. The goals regarding education held by individuals who are primarily interested in doing fundamental research will differ from those who teach to earn their living, and different still are the goals of those who are engaged in education because they feel happy to share their knowledge with a new generation. Moreover, the context of the organisation and of the government's policy generates pressures towards certain goals. In a 'hostile' environment (as with severe budget cuts and reorganisations), goals of survival predominate, while in a 'gentle' environment ample space exists to attend to higher goals, such as quality. But such higher goals that emerge in a 'gentle' environment may just as well be individual goals not related to higher education at all.

In short, quality assessment has to operate in a complex organisational environment where different actors have different objectives with quality assessment and where it has to compete with other goals and interests of all actors who are involved in higher education.

12.4 Processes of quality assessment in the Netherlands

From the more or less theoretical exposition of actors' interests and positions concerning quality assessment I shall now turn to the actual assessment processes that exist in Dutch higher education. This description must provide the basis for the analysis of the situation in the last section. The attention, therefore, will be directed towards the processes of quality assessment and their outcomes.

12.4.1 Education

Internal evaluation

Universities – Until the 1980s the only quality assessments of educational aspects were internal, such as voluntary assessment of courses or programmes by faculties or individual lecturers. Systematic procedures built into the standard operational procedures of the institutions were scarce (see also: VSNU, 1983: 15); long-term programmes or follow-ups were hardly known either. The interest for quality assessment that now exists can largely be explained by the attention for it shown by the government (Weusthof, 1989: 148). A central role in this process has been taken up by the VSNU. In the evaluation process designed by the VSNU both internal and external evaluation are used to organise a periodical, ongoing movement towards better quality
Quality assessment of education in universities (for a detailed description see chapter 6).

**HBO-institutions** - Research on the internal evaluation activities by HBO-institutions is not known to me. The situation until recently probably resembled that in universities: many did something about internal evaluation of education, but often in a non-committal way. Some institutions, however, have set up more extensive quality control systems; such developments can be expected in HBO rather than in universities because of the higher importance to HBO of the relationships with external constituencies (professional field, employers). The HBO-institutions are just beginning to operate as coherent organizations following the STC operation, so in most institutions a coherent institutional quality control policy still has to be –and, it appears, in many of them: is being (Hogeschoolbericht 74)– developed (see chapter 8).

**External evaluation Universities** - Before the VSNU procedure there existed very limited external evaluation of university education. Some external quality control can be discerned in the procedure that existed for ex ante approval of curricula by the Academic Council (the predecessor of the VSNU). And there is a procedure for approval by ‘sister’ faculties for the appointment of professors. One faculty in the Netherlands has experience with intensive external assessment: veterinary medicine of the University of Utrecht. The study programme is unique in the Netherlands. It lacks, therefore, comparison with ‘sister’ study programmes in our country. The faculty has been associated, since about 1975, with the accreditation procedure of the American Veterinary Medical Association.

The VSNU evaluation process was tried out in 1988 in four disciplines (see also chapter 6). The process consisted of a self-evaluation by each faculty and a visiting committee for all faculties of a certain discipline. The visiting committees were provided by the institutions with the report of a self-study, and other data and plans of the institution. The VSNU added to this a ‘fact book’ with centrally known data (partly performance indicators). This information did not function too well in the trial evaluations; the centrally collected data were incomplete, differed from what was known in the institutions, or did not lead to unambiguous interpretations (VSNU, 1988a: 14). The production of fact books was therefore discontinued after the first, experimental round of external evaluations (VSNU, 1988a: 3). This does not mean the external experts have to form their opinions without any objective data, but what data are needed (such as number of students, efficiency, staff) can be incorporated in the self-studies (VSNU, 1988a: 13, 14).

The self-studies are the central documents in the VSNU evaluation process. The VSNU intends the self-studies to be an instrument with a three-fold purpose: to stimulate internal quality care, to prepare the institution for the evaluation by the visiting committee, and to act as a factual basis for the visiting committee (see also section 6.4.2). As yet, it is not clear whether or not these three goals have been attained. Attainment of the first goal, the most important one, is the most difficult to observe. The VSNU has found signals that the trial evaluations have led to follow-up committees and other actions (VSNU, 1988c: 15); this can be interpreted as proof of ongoing interest in educational quality.

Some recent reconnaissance committees, in the way they have operated, have anticipated or imitated the VSNU visiting committees, by not only concentrating on research, but also judging education of the faculties in their discipline. Judgments on education by reconnaissance committees are based hardly ever on performance indicators. The reconnaissance committees that concentrated most on education were the most recent ones, those on Theology and on Architecture. The one on Architecture came into existence, among other reasons, because of the problems with the architecture course in the Technical University Delft, which had not been solved in spite of many attempts by the Executive Body and by locally appointed committees. The members of the Reconnaissance Committee on Architecture formed their judgments by visiting the faculty in question and comparing it with the other university faculty and the HBO-institutions that were visited also. The complete VSNU evaluation procedure can be said to have been followed, including self-studies and all. The proposal by the Reconnaissance Committee on Architecture was drastic: amalgamation of Delft with Eindhoven. The University of Delft wishes to limit the intervention to a reorganisation. The Minister of Education and Science, explicitly true to the new steering philosophy, has stated he will let the institutions seek a solution by themselves.
The Reconnaissance Committee on Theology made extensive use of performance indicators for research, and its judgements were firmly based on them (Reconnaissance Committee Theology, 1989; see also section 6.5.2). Good education was deemed to be dependent on good research, hence the committee stressed performance indicators for research, in that way evading the problem of finding or developing performance indicators for quality of education. The reconnaissance committee concluded from its data that no faculty of Theology in the Netherlands reached a sufficiently high scientific level, and that some faculties were further from this goal than others. Combined with the expectation of a dramatic decline in enrolment, this resulted in the recommendation to close the faculty of Theology of the University of Amsterdam. The University has assented to reorganise the faculty drastically, but does not intend to close it altogether.

The outcomes of the most drastic—and most important—recommendations of the two committees are, notwithstanding their different approaches, the same: both recommended closure of a faculty, both without success. As to their less-sweeping recommendations only a very general statement can be made at the moment, because the implementation processes are continuing: some of the recommendations are implemented, others are not.

Two other committees, organised by the Minister of Education and Science in relation to a budget reduction (the so-called SKG operation, effected in 1986–1987), evaluated research and education of the institutions active in the fields of Psychology & Education and Sociology, Political Science & Public Administration. The first of these committees (they were called visiting committees) made use of peer review methods in the broad sense of the word, and discussions with the institutions involved. The latter visiting committee based its judgement, as far as education was concerned, on one performance indicator. Its results were severely criticised by the field, while the visiting committee on Psychology & Education met hardly any criticism. This apparent difference in acceptance cannot be explained wholly by the difference in methods, however. The reactions have probably been influenced to a large degree by the fact that the visiting committee on Sociology, Political Science & Public Administration had to propose more drastic budget cuts than the other committee.

HBO-institutions—Some study programmes, even before the STC operation, were organised in national consultation bodies, in which questions regarding curricula could be dealt with by representatives of all institutions in a 'discipline'. Formal quality assessments were not performed. During the STC operation, a small number of Task Re-allocation Committees existed; their decisions (which sometimes even included termination of study programmes in certain institutions) were not based on quality assessment procedures either.

Since STC, the HBO-Council has launched two plans for quality assessment in the HBO-sector. The first plan was based on institution-wide quality assessment, but somehow that never evolved beyond the stage of pilot projects. In 1989 a new plan was published: the HBO-Council now wishes to set up a sectoral procedure remarkably like the vHNU procedure (see chapter 8).

12.4.2 Research

The situation as to quality assessment of research will be described for universities only. No formal quality assessment procedures exist for research in HBO, as far as I know. Nor is one necessary, for what little research is done in HBO is almost always applied contract research, which proves its quality by its success in obtaining effective demand.

Internal evaluation

For internal quality assessment of research in universities no formal procedures are prescribed by the government's regulations, though the existence of a committee concerned with research policy is prescribed for each faculty. Some universities, mostly in the eighties, have developed internal procedures for distribution of part of their research budget on the basis of competition by way of some local mix of quality criteria (project funding in 'research pools', or regular assessment of research projects).

Faculties usually had their own procedures to divide their research budget, in which quality assessment not often played a part—and if it did, it was only for a small proportion of the faculty's research effort. Following the retrenchment policies of the early eighties more procedures have been devised, however, for internal quality assessment (Spaanen et al., 1988: 54). The internal procedures are peer review procedures, in which performance indicators (especially publications, sometimes also
success in obtaining grants and contracts) form an important basis for legitimation.

**External evaluation**

For external assessment of quality of research one specific procedure exists, the Conditional Funding procedure. Some other actors in the higher education system (viz., the Netherlands Organisation for Scientific Research (NWO) and some reconnaissance committees) are involved in activities close to quality assessment as well, so I shall give attention to them too. Apart from these, traditional peer review should be mentioned here as well. Traditional peer review—assessment of manuscripts by journal referees, awards and (honorary) memberships, etcetera—however too little resembles systematic quality assessment to include it here more extensively.

**Conditional Funding** — The most important quality assessment procedure regarding university research is the Conditional Funding (CF) procedure, introduced in 1983. The first round of the CF procedure was an *ex ante* procedure. Now, this has been changed into *ex post* judgement with the possibility of suggestions for the future. The judgements are made by, in principle, anonymous peers (nominated by, e.g., the Royal Netherlands Academy of Sciences [KNAW] and NWO) about the quality of each CF programme. The judgements are based on objective data about quality and quantity of labour input, publications produced by the project, but also on the ideas of the reviewers about the quality of the research group involved and about the prospects for the programme.

Following the first round of the CF procedure, Blume, Spaapen *et al.* concluded that (Blume, Spaapen & P:ins, 1983; Spaapen *et al.*, 1988: 48 ff., 60 ff.):

- the internal decisions whether or not to include a particular research project in the CF programmes of the faculty were based, especially in the first instance, on informal quality appraisals of researchers by peers or administrators in the faculty;
- universities inserted mostly ‘safe’ research and refrained from risky research (research of which successful completion could not be foreseen); this implied relatively little conscious selection of research for coherent research programmes;

- the disciplinary organisation of the quality assessments has a restraining influence on interdisciplinary research programmes, while the administrative context (budgets are allocated to institutions, not to programmes) restrains inter-institutional co-operation;
- the external judges, coming from well-known academic organisations, screened off the fields from the government, so that one of the government’s goals, more insight into research, was not attained;
- the external appraisals, in the large majority positive, were partly guided by the fear that the disciplinary field as a whole would be endangered by negative appraisals;
- during the implementation process administrative criteria gained in importance relative to the academic criteria, which tended to stimulate procedures for ‘distributive justice’ instead of quality assessment, thereby making the CF a blunt policy instrument.

In fact hardly any reallocation of staff or budgets resulted from the CF procedure, because almost all programmes were judged to be good enough. This may mean the quality of research is indeed good enough, but it also indicates that the CF procedure is not an impulse to obtain ever higher standards of quality for research.

The CF procedure has been used by universities or faculties to implement negative internal decisions, for which legitimation existed insufficiently before the CF judgements were made (Spaapen *et al.*, 1988: 53–54). Administrative staff, namely, have great difficulty legitimising painful decisions based on quality judgements in the eyes of the academic staff. Academic staff will accept such judgements only if they are backed by high reputations in the academic dimension (Spaapen *et al.*, 1988: 43, 111). The CF judgements bear this mark of high academic reputation, and can therefore be very valuable to administrators.

The effect of the CF procedure has been to guarantee a minimum level of quality, but it has not led to growth of quality. In the CF procedure, peer review led to strategic behaviour on the part of the ‘peers’: they acted as a buffer between researchers and the government. Would performance indicators have resulted in more goal achievement? In the course of time, CF has become more explicit as to goals and instruments. Thereby the role of performance indicators increased (Spaapen *et al.*, 1988: 110). Yet only very simple performance indicators have
been used, such as publications emanating from the research projects. This was coupled to an informal norm, such as: each project should lead to x articles, or y dissertations or z other books. As long as these norms are not too high, many projects would have been judged positively even on this objective basis. Such performance indicators alone, however, would not have resulted in the discontinuation of less than a handful of programmes. Simultaneously, an effect of CF has been to increase power of administrative staff in faculties and at the central level of institutions vis-à-vis the academic staff in the research groups (Spaanen et al., 1988: 109–110).

NWO grants – The Netherlands Organisation for Scientific Research (NWO) decides on its grants for research projects by some sort of peer review procedure. The exact procedure depends on the disciplinary sub-organisation, but it usually includes anonymous referees, organised in sections of specialists, who judge the research project proposals. This results in a list of research proposals ordered by priority. Sometimes the applicants themselves are consulted. The final grant decisions are made by the committee heading the NWO disciplinary sub-organisation, depending on the priority lists from its constituent sections and the amount of funds. This, however, is not quality assessment in a strict sense, but appraisal of expected quality beforehand. Just as happened in the CF procedure these peer judgements will be largely influenced by reputations of applicants based on past performance (Blume, Spaanen & Prins, 1985: 75).

The results of the NWO grants procedure are, as a rule, felt to be satisfactory for all involved. I have not come across a mention of problems by reconnaissance committees and visiting committees, apart from complaints about too little money for NWO to distribute. As further proof of satisfaction the Minister of Education and Science, in 1989, announced plans to shift part of the research funds from the universities to NWO for redistribution (see also chapter 3).

Faculties with experience in NWO grant procedures had less problems adapting to the CF procedure than other faculties – though this did not always lead to more success in the CF, due to differences between the criteria of these procedures (see: Spaanen et al., 1988: 43–44).

Peers, performance and power

Reconnaissance committees – In the contexts of research policy and of higher education policy about a dozen reconnaissance committees have operated.

Seven out of ten committees applied performance indicators – in particular scientometric methods – next to peer review. In all cases the performance indicators included publication analyses, in two cases supplemented with citation data. The publication measures proved to be the most important; other indicators led to a reordering of institutions as to quality only in one case.

Although higher education institutions, intermediate organisations and governmental advisory bodies sometimes react negatively when reconnaissance committees use performance indicators, most of the time their reactions to these evaluation methods are positive. Apparently the objectivity of the method is appreciated, though the reactions are most positive when performance indicators are used ‘intelligently’, i.e., when the reconnaissance committee is prepared to deviate from the objective measures because of its ‘expert opinion’ or after consultation with the institutions in the field. When reconnaissance committees base their conclusions on performance indicators, without corrective possibilities – which coincided with the use of very few performance indicators – the reactions were as negative as when they relied on peer review methods only (two committees). Such negative reactions against rather ‘mechanistically’ used performance indicators appeared twice in the set of ten reconnaissance committees. A problem encountered by one of the two SKG committees, was that institutions criticised the many errors in the data used to construct the performance indicators. The reason this happened was that the SKG committees had to complete their advice in only three months’ time, so they mostly had to rely on centrally available data, which proved to be not sufficiently reliable. The VSNU discovered the latter fact too in its trial evaluations (see above).

As far as policy effects of reconnaissance committees are concerned, only two reconnaissance committees have been clearly ineffective: they are the two that refused to use performance indicators in an effort to protect their disciplinary fields (law, philosophy). Of the other reconnaissance committees, many recommendations have been used in the further development of higher education policy for those fields, the most visible being: programmes or institutions have been stimulated with extra money, faculties have been closed, or re-allocations have been made among sub-disciplines. Partly, these effects have
been reached under the aegis of the large restructuring and retrenchment projects in Dutch higher education (TVC and SKG). Also, the reconnaissance committees often stimulated discussions on strategic matters in their field. All these effects appear to be independent of the methods the reconnaissance committees used, which range from consultation of involved actors combined with peer review informed by performance indicators to rankings by way of performance indicators with slight corrections based on consultations or peer review.

12.4.3 Conclusions

A number of quality assessment procedures exist in the Netherlands, most of them devised and first implemented during the last decade. Almost all of the procedures are externally initiated; internal quality care was rather underdeveloped before the government or the VSNV and the HBO-Council, stimulated by the government’s plans, started their external quality assessment procedures. Initially, the quality assessment procedures were engrained onto the traditions that existed in the more developed sciences: discipline-based peer review was therefore the ‘archetype’. But very soon—often already at the conception of the newer procedures—performance indicators began to play a role as an intersubjective basis for judgements. Their use as more than a basis for judgements by peers is not accepted by the field.

Reconnaissance committees have, from the point of view of quality assessment, some effects. Their presence sets quality on the agenda in their fields. Some policy decisions by the Minister and by institutions are based on their recommendations. Reconnaissance committees are therefore not unimportant for quality, but they are too few and operate with too long and irregular intervals to have a systematic effect for quality care.

The VSNV reports some evidence of enhanced quality care as an effect of its quality assessment procedure. Such an ‘improvement orientation’ does not follow from the CF procedure, which has become more and more ‘control oriented’ and only ensures a minimum quality level. The criteria (performance indicators: publications, success in obtaining research grants or contracts, quality of staff as indicated by Ph.D.s) associated with CF have become more accepted as the academics got used to them (Spaapen et al., 1988: 43, 111). Nevertheless, as stated in the previous paragraph, academics still retain the idea that perfor-

mance indicators alone cannot accurately measure the quality of their research or education. On the other hand, peer review that leads to quality judgements without any intersubjective basis in performance indicators is not accepted in the context of large-scale quality assessment. It will be interesting to see whether the larger role of NWO grants does not lead to a more ‘bureaucratic’, more objective (performance indicators!) and control-oriented procedure for quality assessment and budget allocation.

12.5 Costs, benefits and power in quality assessment policy

Can a theoretical explanation be given for the empirical facts presented in the previous section? To try and do so, I shall start with the assumption that actors in the higher education system act subjectively rational, i.e., actors in the higher education system choose their positions concerning quality assessment processes with the expected consequences of the procedures in mind: how will the quality judgements be implemented in policy by the Minister, the institution or the faculty?

The consequences of a policy aiming at higher quality in the higher education system can be analysed at different levels, and consequently, for different actors. At each level different power and dependence relationships exist. I shall concentrate, in this section, on the vertical relationships, i.e., those between ministry and institutions and those between institutions and faculties. Yet the other levels (research groups, chairs and individual actors) cannot be left completely out of the analysis.

12.5.1 Costs and benefits of quality assessment

Costs and benefits can be interpreted in a narrow, financial sense, but one can also take the institution’s reputation and the relevant contexts into account. These contexts are the fields of the disciplines and external constituencies (notably, employers and professionals), and the administrative context, which includes the funding organisations. The latter context provides the nexus between the reputational and the financial aspects of costs and benefits.

Costs that can be expressed in terms of money include the costs of assessing the quality of a product (education or
Quality assessment

research), the costs of keeping up with or developing new knowledge or technology to improve the product and the costs of implementing these innovations in education and research. Apart from the first factor, these are the things academics (and HBO-teachers) are supposed to do anyway. So only the first factor, assessing quality, is a 'real' additional cost factor for higher education institutions. This cost factor should not be underestimated: a chairman and a secretary may spend half or even all of their time during a year for a reconnaissance committee. The VSNU calculates the visitation procedure, including the preparation of the self-studies, on average £ 350,000 per discipline (about £ 110,000); each institution has to reckon with £ 31,000 (almost £ 10,000) (VSNU, 1988a: 28-34). Quality care, it can be concluded, is not free, but the price is not prohibitive, either.

Another category of costs, less easily quantifiable, results from the fact that when budgets depend on quality assessments the institution is confronted with uncertainty about its future budget, because reallocation as a result of differences in quality is possible. The budget is the basis of a large proportion of bureaucratic relationships between the institution and its faculties, and perhaps even more so between the institution and the Ministry of Education and Science. Therefore, quality assessment is potentially threatening for institutions. It might also be promising, but in periods of science in a 'steady state' (Ziman, 1987) or even budget cuts the threats weigh more heavily than the probability of higher budgets. In such situations defensive behaviour on the part of institutions is to be expected.

In the sphere of scientific reputations, quality assessment can be just as promising or threatening. If the distribution of reputations over a fixed number of elements in the higher education system is regarded, the situation is a zero sum game: a gain in ranking for one institution is a loss for one or more other institutions. Whether it pays for an institution to invest effort in heightening its quality depends on the distance it perceives between its rank order position and the positions (and behaviour) of its competitors. If quality assessments of the higher education system as a whole are at stake (e.g., in international comparisons), all institutions have an interest in high quality. Yet in both cases one also has to take into account that institutions may value their scientific reputations differently: a university, for example, that aims at becoming a European top-five research university will estimate its scientific reputation higher than a university that is more geared towards education for undergraduates.

Peers, performance and power

To what consequences do these hypothetical statements lead in the Dutch situation? Until recently, there were hardly any financial, and little reputational incentives for higher education institutions and faculties to be more than minimally interested in quality assessment. As long as the environment 'buys' the products (graduates, doctorates, publications), the institution's quality appears to be good enough. And even if the environment does not, the institution can either shift the blame on the environment (e.g., point to the high level of unemployment found in other sectors of the labour market as well) or even try to change the environment (such as: start its own periodical in which to publish articles). If budgets depend predominantly on the number of students, and if potential students choose to study in a higher education institution on other grounds than its scientific or educational reputation, then there is no reason for the institution to invest in quality care. The low level of institutional quality assessment procedures which existed before about 1980 is therefore understandable.

Since the early eighties, however, the government emphasises it is interested in more quality and that quality assessment may have financial consequences. Some of the restructuring operations the government has organised under this aegis have been mentioned above. What are the financial consequences of these policy changes? Specifically: what are the benefits for institutions? There is an incentive to co-operate for institutions facing the threat of taking away part of the budget if its quality proved to be too low. Vaguely there also have been, on the positive side, promises of more resources for good quality ('vernieuwingsgelden'). Furthermore, with the so-called mission budget the universities should be enabled to stress their individual corporate identity. Yet on balance the negative financial consequences have dominated, especially when related to the growing number of students. This context of shrinking budgets makes institutions very intent on their survival, without which none of their 'higher' goals can be attained. This creates a tendency towards strategic behaviour: try to stave off policy proposals that reduce budget certainty, try to encapsulate peer evaluation committees (see especially the CF procedure, in which both strategies have been used), keep data secret or muddled (this may, though need not be, a reason why centrally available data
proved unreliable), try to 'score' on performance indicators, etcetera.

However, not only a period of budget reductions diverts the attention of actors in the higher education system away from quality per se. Any policy context of centralised resource allocation based on bureaucratic procedures is an impediment to a policy aimed at raising the quality of education and research and at more permanent quality care in the higher education institutions. The reason for this is that bureaucratic procedures tend to be geared towards legitimation of decisions, which means, in a democratic constitutional state, that objective rules must be obeyed. Objective rules may agree with performance indicators, but not with the more comprehensive, but subjective, quality judgements of peer review. It appears, however, that scientists deny legitimacy to decisions based on only performance indicators. Performance indicators also tend to distort the operational concept of quality, as has been exposed in chapter 1.

How successful the higher education institutions can be in thwarting the government depends on the power and dependence relationships between institutions and the government. The institutions are dependent on the government for approximately 90% of their resources. The government depends on the institutions for the provision of higher education and of research; it cannot easily organise alternative higher education institutions (Bijleveld & Goedegebuure, 1989: 107). Yet in this situation of mutual dependence the government has the upper hand, for the government has the capacity to set the rules of the game for the other participants. Consequently, it can be expected that the government will attain a larger share of goal attainment than its opponents, the more so as its opponents are engaged in games (such as the game of scientific reputation) among themselves as well. For quality assessment this implies that the government's goals will be given a more prominent place in the definition of quality that will be used in practice than other actors' goals, even if the government is not directly involved in the quality assessment process. The 'governmental' definition of quality that will be used in practice results in an even greater distance between the higher education institutions and quality care as a permanent internal goal: it is not their, but the government's quality assessment that is taking place. Symptomatically, out of twelve reconnaissance committees studied, six identified with the government's task assignment, two occupied an intermediate position and a minority of four identified with the field. Out of these four, two clearly were trying to defend their discipline against the government's intentions (retrenchment); they even refrained from making quality judgements.16 Another effect of the 'unbalanced' quality assessment is that it creates an uniform 'ecological niche' for all higher education institutions involved in a single quality assessment procedure. This creates a pressure towards uniformity for the higher education institutions, contrary to the variety and innovativeness the government wishes to stimulate to make the higher education system better adapted to its multi-faceted and ever-changing environment. The institutions then have few possibilities to stress their individual goals: in practice, they are not free to become, for example, a research university or an education-centred institution, because they all have to comply to uniform standards of quality.

12.5.2 The positions of faculties

For the financial relationships, the institutional level is highly important. As regards scientific reputations the proper level of analysis is the faculty or the research group, as these are the levels at which the disciplinary dimension interlocks with the institution. Here one must differentiate among various types of disciplines. For in some paradigmatically homogeneous disciplines it is possible to arrive at a national or even an international judgement about quality, while in paradigmatically heterogeneous disciplines the same is not possible (Whitley, 1974; Spaapen et al., 1988: 15; Van der Meulen & Westerheijden, 1990: 15). Paradigmatically heterogeneous fields are not only found in the humanities and the social sciences, but also in technical fields: in the field of mechanical engineering practically all professors recently refrained from judging research outside their own specialism for the Advisory Council for Research Policy (RAWB) (1989: 26), because there was too little communication between the specialisms.

Faculties accordingly occupy different positions in quality assessment procedures. Faculties in heterogeneous disciplines can withstand pressure more easily from discipline-wide quality assessment by saying that in their way of teaching or research the traditions are different so that such-and-such performance indicator does not apply, or that the visiting committee is one-sided. As stated before, in situations like this, peer review
procedures can hardly lead to discipline-wide judgements; performance indicators, not hindered by such scruples, will be however less acceptable to the scientists in the discipline. The other side of the coin is that a homogeneous discipline has a better bargaining position vis-à-vis the institution and the government (Spaapen et al., 1988: 17–18). Therefore, it can withstand pressures better than a heterogeneous discipline. Quality assessment procedures can then be geared more towards intrinsic goals of the discipline, instead of being a mix of intrinsic goals and, predominantly, policy goals (Spaapen et al., 1988: 18; De Weert, 1990: 68).

The quality assessment processes existing in the Netherlands, for research as well as for education, make use of both performance indicators and peer review, albeit to different degrees. Though there may be a tendency towards more use of performance indicators, the individuals involved in research and higher education in the Netherlands retain a preference for judgement by peers. This is understandable when it is considered that peers, even when performing assessments in a procedure in which the government's goals are more important than the discipline's goals, will remain members of the discipline. This gives the individuals involved the possibility to apply social control, which is far less available to them when performance indicators are handled by administrators in government organisations.

The fact that in procedures having effects for whole research groups, faculties or institutions the judgements must be based on some inter subjective data to be accepted by the field has been noticed before. Peer review without formal performance indicators is regarded as legitimate, though, in 'classical' peer review as found in journal referee systems; then, however, scientific argumentations between the anonymous authors and referees are often possible. Interestingly, peer review is also perceived to be a legitimate method for ex ante judgement of individual research projects (e.g., the NWO grant procedure), even if this has financial consequences (viz., subsidy for the project or not). The balance dips to the other side with judgement of multiple research projects, as in the CF procedure. In other words: the balance dips to the other side when decisions are not about single projects of individual scientists, but about entire organisations run by administrators. Both the magnitude of the costs and benefits, and the different organisational dimension (academic versus administrative) may explain the different reactions.

12.5.3 Power, and methods of quality assessment

It is remarkable that most of the considerations about the positions of institutions and faculties apply to quality assessment procedures as such, without much differentiation as to the precise methods: the policy game surrounding the quality assessment procedure is the relevant context for the actors. Bargaining and power play an important role in this context. The argumentation that leads to a committee's recommendations, i.e., the peer review or the performance indicators, hardly comes to the fore at all: the quality assessments 'lead their own life', almost independent of their validity and reliability. The methods used in quality assessment procedures are either applauded or condemned, depending on whether or not the actor values the recommendations positively (Van der Meulen & Westerheijden, 1990: 13–14). This does not mean the choice for one or another method is of no importance at all, but it does mean that in the context of higher education policy the possible organisational or political consequences weigh more heavily than the quality of the way they have been reached.

12.5.4 The dilemma: the stakes of the game

Is it not possible to have external quality assessment without such conflicts? Is it not possible to devise system-wide quality assessment that is not 'control oriented' but 'improvement oriented'?

I fear not. Within small units, say faculties, rather informal quality care systems may exist, because all those in the unit are involved in the procurement of good education or good research. Social control can then provide enough positive or negative sanctions. In larger social systems, the administrative dimension becomes dominant. If, then, the actors expect no real consequences (such as budget reductions or additions) to follow from the quality assessment, no one will take the trouble to go through such an elaborate procedure. Yet the moment such real stakes are introduced, quality assessment becomes an arena for the power game in higher education institutions, and between higher education institutions and the government, so strategic behaviour -power oriented behaviour- will dominate 'improve-
ment oriented' behaviour. This is, I suppose, the dilemma of quality assessment: without the expectation of real consequences, the incentives to organize quality assessment are lacking; with the expectation of real consequences, quality assessment will turn into a power game. Whatever system for quality assessment, on a larger scale than within faculties, is developed, it is doomed, I fear, either to remain without implementation, or to become corrupted by strategic behaviour.

Notes
1 This is the direction exposed in: De Weert, 1990: 60–61.
2 I shall write mostly about universities and their association, the VSNU; but especially since the HBO-Council announced it would organise external quality assessment for the HBO-institutions very much along the VSNU lines (see chapter 8), most of what I write applies to both parts of higher education. When necessary, I shall give, however, separate attention to the situation in the HBO.
3 The paucity of explicit goals is also mentioned in: Pollitt, 1990: 65–67.
4 For a comparable argument concerning the situation in the UK, see also: Pollitt, 1990: 65.
5 I ignore for the moment the interest of the VSNU or the HBO-Council in the success of the particular arrangement of quality assessment, in which they play a predominant role. This provides the organisations with an extra reason to want quality assessment to be successful.
6 Reconnaissance Committees are committees of scientists from a certain discipline, appointed by the Minister of Education and Science, to explore future developments in their discipline, or to judge the relative strengths and weaknesses of the Dutch scientists in the discipline, or both, in order to inform the development of policy in science and higher education. Visiting Committees are the committees for external quality assessment—at the moment, of education only—appointed by the VSNU; some government committees (also with a task assignment of judging quality of Dutch research and education in a discipline) have been named so too.
7 Data from research project 'Evaluation of reconnaissance committees' by Van der Meulen & Westerheijden (1990).
8 There are two technical universities offering a study programme on Architecture—Delft and Eindhoven. The faculty of Architecture of the Technical University of Delft has, despite its internal problems, a very high international reputation, according to the list published in Libération, 1989.
9 Which is a contested opinion in the literature. See, e.g., De Weert, 1990: 66.
10 Later, a very great proportion of research was entered into CF pro-

grammes to secure financing for staff, making selection based on quality almost impossible.
11 Data from research project 'Evaluation of reconnaissance committees', see also: Van der Meulen & Westerheijden, 1990, in which most reconnaissance committees that have existed in the Netherlands are included. In the text I omit from consideration two committees that operated in the 1970s and that had a task more directed towards the relationship between applied research and government policy.
12 The tenth reconnaissance committee did not collect data to construct performance indicators, but made use of other reports about its field, some of which included performance indicators (the CF reports).
13 The policy effects of reconnaissance committees cannot be measured easily, because, on the one hand, the implementation processes extend over several years and, on the other hand, these involve many of the key actors of the higher education system. Moreover, the recommendations by the reconnaissance committees are not always in a form that facilitates a review of their effects (e.g., very general recommendations, addressed to many actors). Nevertheless, the effects mentioned in the text can be discerned.
14 Data from research project 'Evaluation of reconnaissance committees' by Van der Meulen & Westerheijden (1990).
15 See the section on reconnaissance committees, above, and Spaapen et al., 1988: 111.
16 In one of the two cases, the reconnaissance committee on Economics, the paradigmatic structure of the discipline, with many disparate sub-disciplines, was not conducive to overall quality assessment anyhow.
17 A comparable conclusion regarding the CF procedure was drawn by Bijleveld & Goedegebuure, 1989: 108.