

# Dutch experience

---

## Evaluation of societal quality of public sector research in the Netherlands

Barend van der Meulen and Arie Rip

*Dutch thinking about the issues, tools and practices of evaluation is explored, with special reference to societal quality. Indicators are identified and positioned through the review of 17 evaluation processes in the Netherlands. The context for the review process is examined.*

**P**RESSURES ON PERFORMERS of basic research to demonstrate user relevance, relationships with industry or utility, are visible in all national research systems, and while the terms often remain nebulous, the pressures leave traces in the evaluation of basic science. The Netherlands has gone quite far in this direction, both in thinking about the issues and the evaluation tools, and in actual evaluation practices. This article reports on the Dutch experience, and in such a way that generalisable lessons can be identified.

Evaluation has become more frequent, in particular because government sponsors stress accountability of researchers as well as the need to increase the chances that funding goes to research of strategic importance (Rip, 1999). And the scope of evaluations becomes broader: 'traditional' peer review is extended by including other actors than scientists working in the same area, and methods and tools are sought to evaluate the societal quality of research. The concept of 'societal quality' of research was introduced in the Dutch discussions in the early 1990s (see also Spaapen and Sylvain, 1994; and Van der Meulen and Rip, 1995) to suggest an analogy with 'scientific quality' of research. In both cases, quality is not limited to eventual impact; peers and other knowledgeable actors can form judgements about quality; and indications and indicators of quality can be developed and used. For scientific quality, there is a long tradition of review and decision-making based on evaluation of quality, and best practices have emerged. For societal quality, the situation is more complex, and it is not clear how to evaluate societal quality, especially for basic and strategic research.

The literature on the contribution of scientific research to socio-economic development, and to society

Both authors are at the Centre for Studies of Science, Technology and Society, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands; Tel: +31 53 489 3906; Fax: +31 53 489 4775.

more generally, emphasises the problems of measuring and evaluating (from Dunn and Holzner, 1987 to ETAN, 1999), but when read selectively and creatively also indicates real possibilities, and thus offers clues for developing methods for evaluation of societal quality of research. What is necessary is to locate existing tools and approaches in a conceptual framework which does justice to the richness of basic and strategic research in relation to possible applications (in the economics literature, for example, OECD, 1992, the phenomenon of spillovers of basic and strategic research captures one aspect of richness). At the same time, there is often no direct link of (and linear sequence from) research to utilisation and impact. In Irvine and Martin's well-known definition of strategic research this is very clear:

Strategic research: Basic research carried out with the expectation that it will produce a broad base of knowledge likely to form the background to the solution of recognised current or future practical problems. (Irvine and Martin, 1984)

Thus, research contributes to knowledge reservoirs, and actual uptake and use occurs through the new combinations that others make based on items in such knowledge reservoirs (Rip, 1997). This is a way to add a cognitive aspect to issues of transfer of knowledge and the need for well-functioning of networks for knowledge exchange. The third main element of the conceptual framework, eventual (indirect and sometimes direct) use and impact, has been discussed more extensively in the literature.<sup>1</sup> Together, the three elements span the space between scientific and scholarly research and uptake in society. Direct use, an aspect of societal quality on which policy-makers tend to focus, now appears as a particular case, and a shortcut through the space of research and uptake which is possible only when intermediating practices have become established. In the first part of this article we elaborate this conceptual framework further so as to identify and position various indications and indicators of societal quality.

For evaluation of societal quality, the articulation of a conceptual framework has to be combined with the development of practicable measures and indications (often proxies) of richness, of channels and interactions, and expected outcomes. Is it already possible to define what tools and methods are appropriate in what kind of evaluation process? Have best practices emerged from the experiences and experiments so far? Recently, we reviewed 17 evaluation processes in the Netherlands, and analysed how societal quality was evaluated and to what extent specific tools and methods had become part of the evaluation repertoire (Van der Meulen and Rip, 1997). In the second part of this article we report the main findings of the latter study.

In order to identify good practice it is important to look at the evaluation context as well as at the methods

used. Within two kind of evaluation contexts good practices have emerged: *ex ante* selection of projects with industrial relevance, and visitation of university research programmes and research institutes. Both seem to be robust enough to be practised in other countries as well. A third context, evaluation of national research programmes, is not very well developed in the Netherlands, despite the importance of such research programmes for policies to improve the relevance of basic research. In the concluding section we reflect upon the findings, also in terms of learning processes and their value in other evaluation contexts.

## Understanding societal quality

With the benefit of experiences in Dutch science policy, including the introduction of systematic evaluation of university research in the 1980s, and the discussion of societal quality of research started up by the sector advisory councils for research policy<sup>2</sup> in the late 1980s, we start with a broad view of societal quality. It captures such diverse phenomena as direct use of results of policy-oriented research, long-time collaborations with industrial counterparts as occur in engineering sciences and chemistry, understanding and development of solutions for urgent problems as the environmental sciences aim for, and the cultural importance of humanities. For science policy and the social legitimization of science, it is important to have such a broad definition. For evaluation, however, the challenge is to identify aspects that can be traced through indications and indicators, without overmuch reducing the richness and complexity.

Two broad categories are often used in discussions and when creating indications and indicators: relevance and impact. As a first step in our attempt to understand societal quality of research, we indicate how these categories work in evaluation:

*Relevance:* starting from a proposed, ongoing or concluded research project or programme, one enquires into its actual and envisaged linkages and promises. Relevance is particularly important in *ex ante* evaluations, but the promises should be checked in *ex post* evaluation.

*Impact:* the uptake of research (and the effects of such uptake), often as a combination of results of several projects, earlier findings, and experience of practitioners, can be studied as such, but for *ex post* evaluation, attribution to specific research projects and actions is necessary.

Relevance of scientific research can always be argued for. Every research project can be legitimated by being linked to external questions, be it industrial technology, a health issue, the environment, social issues or societal learning and culture. The necessity to legitimate research in terms of socio-economic contributions and user relevance in general has introduced additional promise dynamics across the research

system. Researchers, societal actors *and* policy-makers create hopeful expectations to argue and justify new investments. To mobilise resources, over-optimistic claims are made, and evaluations will therefore always show that the original aims were not met. At end of the 1980s, for example, biotechnology in the Netherlands had not realised the promises made when public stimulation programmes for biotechnological research were established. Policy actors felt uncomfortable about this, and managed the situation by adjusting expectations after the fact. The phenomenon of originally high expectations and disappointment and adjustment after the fact occurs quite generally and is, in fact, a necessary consequence of the quest for relevance in contemporary research.

An important question for evaluation of societal quality is thus how to check the robustness of expectations. Good practices, in this respect, must distinguish superficial claims for relevance ('the environment is an important issue, so environmental research must be relevant') from better articulated and consolidated claims. Inspection of such claims by experienced assessors helps — if one can find such assessors, and if the sector or the societal aspects are themselves sufficiently articulated to allow specific and traceable claims. Clearly, there are preconditions for evaluating relevance that lie beyond the evaluation itself. One precondition for systematic evaluation is a certain consensus on the potential of basic research in a specific area to contribute to socio-economic developments and other desirable goals. If structures and practices have evolved to transform such potential into actual contributions, checking the (actual or intended) use of such structures and practices in the research project will be an operationalisation of relevance. Such indications of relevance are being used increasingly in assessing research proposals, even while they might well neglect 'lateral' developments. Another possibility, only rarely used as yet, is to link up with systematic analyses of contributions of scientific research to crucial cultural questions, socio-economic issues and technological developments, as these are available in priority setting documents and science and technology foresight exercises. Such documents and insights are a form of 'intelligence' about the future (ASTPP, 1999), and research proposals may be required to position themselves in terms of this 'intelligence' (not necessarily always agreeing in detail).

In the medical sector, for instance, increasingly decisions on new investments in therapeutic means and instruments are not made by just looking at the therapeutic possibilities and perhaps comparing them with alternatives. Stories of the difference they will make, and indicators like mortality and morbidity of different diseases, are used to weigh different investments. Such considerations also influence priority setting of medical research, where information on the severity of a disease is complemented with assessment of critical problems in therapies and possibilities to improve these through research. Relevance of research can

---

## **Existing knowledge and the intensive discussions between policy-makers and researchers about the state of affairs provide background assessments against which the relevance of a particular piece of research can be judged**

---

then be shown by referring to these considerations. Similar analyses are possible, and are being done, for environmental research. Existing knowledge and the intensive discussions between policy-makers and researchers about the state of affairs (with regard to the available knowledge and the environmental quality) provide background assessments against which the relevance of a particular piece of research can be judged.

The second broad category, impact of research, appears to lend itself to measurement, but there are many limitations. There is the problem of attribution; for one thing because impacts can be quite diffuse ('impact accretion'), for another because impacts are co-produced and often combine results from several research projects and general insights. Attribution problems are exacerbated when the extent of impact is to be assessed, rather than identification of salient examples. Uptake and impact might occur years after the results have become available; studies of the Dutch Technology Foundation have shown that technical findings could have impacts only after ten years (or longer). After so many years, it is often very difficult to disentangle what has gone into the creation of an impact. Evaluation studies compromise by taking a snapshot after two or three years: sufficient time must have passed to capture a major part of eventual impact, but not so much that traceability to research projects becomes difficult.

Instead of measuring impacts themselves, the routes toward them can be mapped and evaluation can be based on assessment of the nature and density of 'travel' along these routes.<sup>3</sup> One way or another, knowledge has to find a way to 'users' (the quotes indicate that users are often projected rather than actual). If research results spread via courses, publications, workshops and so on, direct impact is difficult to measure, but indirect impact is probable. Specific relations are easier to trace, and might well offer better channels for uptake and impact. Collaborative projects, contract research, membership of advisory boards and committees are examples, and such direct relations with societal actors, industry and policy-makers are often used as indicators for societal quality.

We would argue not just that one must be careful in replacing impact assessment by relation assessment but also that the occurrence of such relationships is

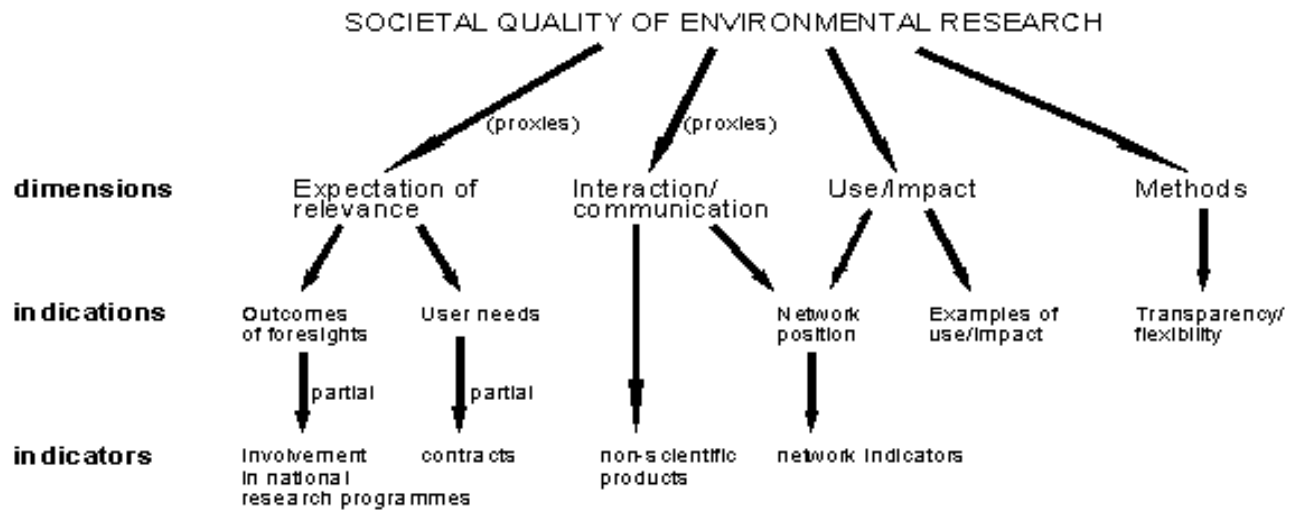


Figure 1. Defining 'societal quality of environmental research'

important for other reasons. Within these interactions, shared frames of reference are developed, and hybrid knowledge reservoirs are created and extended. Knowledge reservoirs carried by epistemic communities occur generally — scientific fields being obvious examples — but here they comprise scientific insights as well as practical knowledge and experiences, and insights and rules about how these two are interrelated and merge.<sup>4</sup> Such hybrid knowledge reservoirs are needed to create effective communication and enable societal actors to become real users. Indications of societal quality should thus include assessment of the existence and nature of knowledge reservoirs. Additionally, non-scientific members of the epistemic community that carries a well-functioning hybrid knowledge reservoir can be involved productively in evaluations as users or experts to make informed judgement about the societal quality of research

projects and programmes. Without these shared frames of reference there is little else to rely on for non-peers but their general knowledge and own preferences.

In discussing the commonly used, broad categories of relevance and impact, we have developed further indications of societal quality, based on an understanding of the processes involved, and we have gone some way in articulating the three elements (richness, reservoirs, and impact) we identified in the introduction as spanning the space between research and uptake in society. In a second step, we combined the general approach with data about evaluation practice in two research domains with articulated goals and user linkages: environmental research and research on neuromuscular diseases. We asked researchers in these domains about their experiences in relating research to society, about effective communication channels, and about ways and means to evaluate

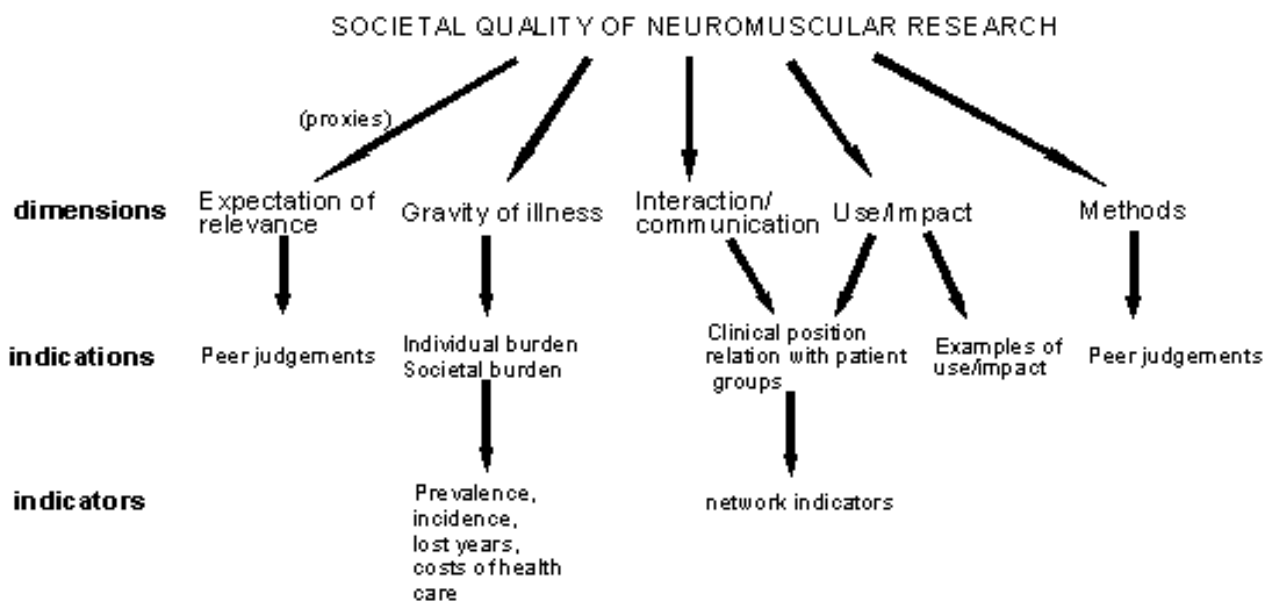


Figure 2. Unpacking 'societal quality of neuromuscular research'

Table 1. Policy context of the evaluation processes

Evaluation process	Policy context				
	Objective	Principal	Agent	Evaluator	Main user
VSNU visitation chemistry	learning	universities	VSNU	peer committee	universities
VSNU visitation law studies	learning	universities	VSNU	peer committee	universities
ECOS research schools	accreditation	OCW/KNAW	ECOS	peer committee	universities
KNAW – visitation of Institute for Social History	learning/ accountability	KNAW	KNAW office	visitation committee	institute + KNAW
NWO selection of proposals for priority programmes	selection	NWO	NWO	peers + expert committee	board NWO
NWO interim evaluation priority programmes	learning	NWO	programme board	programme board	programme board
NWO end-of-term evaluation of priority programmes	accountability	NWO	programme board	programme board	NWO
STW <i>ex ante</i> evaluation of projects	selection	STW	STW	peers + lay jury	STW
STW <i>ex post</i> evaluation of projects	learning/ accountability	STW	STW	staff	STW
final evaluation National Soil Programme	accountability/ learning	OCW	RMNO	consultant	RMNO
innovation-oriented programme carbohydrates <i>ex ante</i> evaluation of projects	selection	programme board	programme staff	peers	programme board
interim evaluation innovation-oriented programme carbohydrates	learning	EZ	consultant	consultant	
final evaluation innovation-oriented programme carbohydrates	accountability/ learning	EZ	consultant	consultant	EZ
innovation-oriented programme imaging technology <i>ex ante</i> evaluation of projects	selection	programme board	programme staff	peers	programme board.
Engineering University of Delft selection of centres of excellence	selection	university board	advisory committee	peers + advisory committee	university board
Experimental visitation National Institute for Public Health Studies	learning		visitation committee	visitation committee	Institute
TNO evaluation of institutes	learning/ accountability	TNO board	institute	peer committee	institute +TNO board

Key: ECOS = Accreditation Committee for the Research Schools; EZ = Ministry of Economic Affairs; KNAW = Royal Netherlands Academy of Arts and Sciences; NWO = National Research Council; OCW = Ministry of Education, Culture and Sciences; RMNO = Advisory Council for Research on Nature and Environment; STW = Technology Foundation; TNO = Netherlands Organisation of Applied Scientific Research; VSNU = Association of the Dutch Universities

societal quality. In our analysis, we identified *dimensions* of societal quality visible in practice, and *indications* and *indicators* linked to them.

Indications of quality are pieces of systematic information (up to little stories) that justify the attribution of quality. For each of the dimensions, the kind of information appropriate to it can be specified. Having such information is a first step in distinguishing empty or general claims for relevance and impact from substantial ones.

For some of the dimensions, indicators could be identified as well. Indicators are a sub-class of the class of indications; they can be presented as scores and represent the phenomenon one is interested in. Citation scores, for example, are a good indicator of scientific visibility, and a good proxy indicator of short-term scientific quality. For societal quality, it is more difficult to identify good indicators. For example, a research institute can have a record of products like reports, expert testimony, and advice, which show its activities in transferring results to the policy domain, and thus provide an indication of societal quality. To turn such data into an indicator, one should

understand the nature and extent of impacts of these kind of research products, and score the actual products in these terms.<sup>5</sup>

For environmental research, four dimensions of societal quality were identified, and for neuromuscular research five. (See Figures 1 and 2.) The terms used are indicative; the dimension 'methods' refers to the presence or absence of research approaches which enhance use of research; for example, transparency to outsiders, design orientation or action research. One conclusion is that the dimensionality and the possibility of indications and indicators are different in different research domains.

A further gloss on these findings is possible by considering research domains in which goals and user linkages are less articulated. Clearly, the extent to which socio-economic problems, knowledge needs and possibilities of research to contribute, are articulated varies across disciplines and research areas. Borrowing the concept of strategic (task) uncertainty from Whitley (1984), we can say that the extent of *socio-economic or more generally, societal strategic uncertainty* is different in different fields.<sup>6</sup> Societal

strategic uncertainty is unavoidable because of the open-ended character of research, but it is relatively low when the 'knowledge system' (Dunn and Holzner, 1987) is articulated and functions well. Uncertainty can be further reduced through foresight studies and other ways of enhancing 'intelligence' (ASTPP, 1999). It is also possible that societal strategic uncertainty is reduced locally and regionally rather than at the field level, through functioning missions and strategies of research organisations. Funding agencies are now also acting like brokers, creating linkages between research and societal actors.

Chemistry is a field with a lot of interaction between industry and universities. Industrial research managers as well as university researchers have shared ideas about the function of university research and how it can 'serve' industrial needs. In addition, there is a tradition of informal interaction and consultation, and joint representation towards policy actors. This must count as a characteristic of the field of chemistry. Even in the Netherlands, with its consultation culture, the field of chemistry stands out in this respect (Van der Meulen and Rip, 1998).

### **Research evaluation practice in the Netherlands**

In an earlier paper (Rip and Van der Meulen, 1995), we characterised the Dutch system as a patchwork: evaluation is by and large considered to be part of the 'housekeeping' of universities, research institutions and agencies, and delegated to those organisations. Thus, evaluation approaches have evolved according to the need of these organisations and sectors, each in its own and separate 'patch'. In keeping with the Dutch tradition of 'mediation' (Van der Meulen and Rip, 1998), evaluation is more related to R&D management and ongoing quality assurance and improvement than to measuring goal achievement by a sponsor. This delegation to the intermediary level implies that learning is organised within organisations, and approaches can vary. Since many evaluation reports (and in particular, those in the university sector) are public, there is use of information across organisations, and research groups, university departments and research institutes build up a reputational profile also on the basis of evaluations.

We reviewed 17 evaluation processes that occurred at different places in the research system (Van der Meulen and Rip, 1997). The processes were selected with an eye on their prominence in the overall evaluation system, as well as on the probability to find interesting experiences with evaluation of societal quality. We analysed the main evaluation processes of university research organised by the Association of the Dutch Universities (VSNU), by the Royal Netherlands Academy of Arts and Sciences and the National Research Council. In addition, we analysed the evaluation of several research programmes and of research institutes as well as an evaluation within the Delft

University of Technology aimed at selection of centres of excellence. By way of introduction to the evaluation processes, we provide a brief description of the policy context of each evaluation process in Table 1. For each of the evaluation processes, the objective of the evaluation, the organisation that has asked for the evaluation (the principal), the organisation that was responsible for the evaluation (the agent), the actors who did the actual evaluations and the users of the evaluation results, are listed.

Since the early 1990s, the VSNU has developed evaluation of university research. This is now an established practice (Verkleij, 1998), and other countries emulate parts of it. In principle all directly funded university research, organised in multi-annual programmes, in research schools (cf. below) and (inter)university institutes, is evaluated in a four-year rolling scheme, arranged by broad disciplinary areas in which all the research units (often complete departments or faculties) are evaluated. The evaluation procedure is a combination of self-assessment and evaluation by a peer committee. We selected two different fields of research, chemistry and law, each with specific linkages with economic and societal actors.

For the Royal Netherlands Academy of Arts and Sciences (KNAW), we chose the accreditation of graduate schools and a visitation of one of its own institutes. KNAW performs a specific role in evaluation of basic research. It has a general advisory role on evaluation of basic research, for example on new evaluation procedures and on the selection of peers for evaluation committees. The Academy sees itself as a protector of basic research and tends to give prominence in evaluations to scientific quality, although in specific cases it is responsive to requests to evaluate socio-economic relevance as well. It often hosts ad hoc evaluation committees. One recent example is the selection committee for the new Technological Top Institutes. For its own research institutes it has developed a evaluation procedure similar to that of the VSNU.

The KNAW hosts the Accreditation Committee for the Research Schools (ECOS). Research Schools in the Netherlands are formal (inter)university collaborations of the main groups in a research field. The research schools provide advanced training for PhD candidates and organise research of the research groups in five-year research programmes, with funding guaranteed by the universities from their block grants. The ECOS was established by the Royal Netherlands Academy of Arts and Sciences at the request of the Dutch Minister of Education and Science. The Academy has developed the accreditation protocol, defines the criteria for evaluation and appoints the members of the Accreditation Committee. Within the framework provided, the committee is autonomous and independent in its assessment of the research schools.

For the National Research Council (NWO), we looked at the selection, interim evaluation and final evaluations of its research programmes within the

PRIORITEIT funding scheme and the evaluation practices of the Technology Foundation. NWO had initiated the PRIORITEIT scheme in order to increase its capacity to stimulate research of societal relevance. In 1997 the grant scheme was terminated, but some of its programmes are still running. The programme budgets range from 5 to 10 million guilders, with a lower bound of 2 million guilders for a small programme on 'Ethics and Politics', and a peak of 67.2 million guilders for a co-funded programme on materials science. Programme themes vary widely; examples are language and speech technology, sensor technology, psychological fatigue at work, Dutch culture in European context, and nonlinear systems.

Within NWO, the Technology Foundation (STW) has a specific role. It was established in 1981 as an independent funding agency, and only later put under the umbrella of NWO. It awards grants to technological research (in a broad sense) at universities. There are no limitations as to subject or disciplinary area; the only requirement is a operational indication of expected *utilisation* of the research proposed. In addition to its *ex ante* evaluations, STW has developed methods for *ex post* evaluations in which results and impacts of projects are assessed five and ten years after the initiation of the project. The *ex post* evaluations of the project's utility are done by STW staff, using a four-point ranking of the relation with the user, the availability of a concrete product and the financial returns. Aggregated results of *ex post* evaluations are published periodically, and are used internally for evaluation of STW's own policy (Van den Beemt and Le Pair, 1991; Van den Beemt, 1997).

Funding of research programmes has been a main instrument of the Ministry for Education, Culture and Sciences (OCW) and the Ministry for Economic Affairs (EZ), responsible for science policy and technology policy respectively, to steer basic research towards socio-economic objectives. In the 1980s, the OCW initiated a series of national research programmes on policy-related issues, co-funded by other ministries, and in some cases, NWO. Most of the programmes have been evaluated, but in very different ways. From these evaluations we selected the one of the National Programme on Soil Research. As part of the evaluation, the Advisory Council for Research on Nature and Environment had been asked to evaluate the societal quality of the programme. The scientific quality was evaluated by an ad hoc committee organised by the Academy.

The Ministry of Economic Affairs funds a range of innovation-oriented research programmes (IOPs). The programmes aim at stimulating knowledge development, network creation and knowledge transfer in areas critical for Dutch industry, like biotechnology, catalysis, metal research and so on. The management of this scheme is delegated to SENTER, the ministry's agency for technology policy. *Ex ante* evaluation of proposals is done by a programme board. The interim and *ex post* evaluation studies are commissioned to external consultants. From the evaluation studies we

selected the evaluations done for the Carbohydrate Programme, a programme that was finished and had been quite successful, and for the new Imaging Programme, that had just finished its first selection round. For this *ex ante* selection procedure, the Imaging Programme had developed new criteria to increase the probability of utility of research results.

Apart from these main organisations responsible for commissioning and organising evaluation procedures, other bodies can be involved in ad hoc evaluations. Within research organisations, research institutes and universities, the board can organise evaluation processes for internal management purposes. From these ad hoc evaluations we selected those by the board of the Netherlands Organisation of Applied Scientific Research (TNO), one by the Netherlands Institute of Primary Health Care Research (NIVEL), and the evaluation process of proposals for internal interdisciplinary research programmes of the Delft University of Technology. TNO has set up a regular evaluation system for its own institutes. Its process is modelled after the university research visitation scheme (self-evaluation combined with peer review and site visit); the criteria used are linked to TNO portfolio management. NIVEL emphasises within its research management its contributions to health policy and health management and wants to be evaluated along these dimensions, in addition to evaluation of the scientific quality. The evaluation process at the Delft University of Technology aimed at selection of research programmes that could effect technological breakthroughs in and across industrial sectors.

The Dutch evaluation system (or patchwork, cf. Rip and Van der Meulen, 1995) relies on delegation of responsibility for research evaluation to the intermediary organisations, in particular the Royal Netherlands Academy of Arts and Sciences, the Association of Dutch Universities (VSNU) and the Netherlands Research Council — with some specialisation as to the kind of evaluations. These intermediary organisations are closer to the research performance level than central government agencies and ministries, and part of their self-image is that they serve the research system.

While the policy context of evaluations in the 1980s focused on accountability and quality, by now the central concern is the longer-term value of research programmes and how this can be enhanced.

---

**The review of evaluation practices aimed not just at finding 'how-to-do-it' methods, but also whether and how evaluation of societal quality would be useful and whether it had become part of evaluation repertoires**

---

Table 2. Experiences of evaluation of societal quality

Evaluation process	Societal strategic certainty			Aspects of societal quality				
	Definition of societal objectives	Society-oriented mission	Kind or research (generally)	Expectation	Urgency	Interaction	Use	Methods
VSNU visitation chemistry	medium	not explicit	basic-strategic	✓		✓	✓	
VSNU visitation law studies	medium	not explicit	strategic					
ECOS research schools	low	no	basic	✓				
KNAW – visitation of Institute for Social History	low	not explicit	basic					
now selection of proposals for priority programmes	low	relevance	basic-strategic	✓	✓			
now interim evaluation priority programmes	low	relevance	basic-strategic	✓				
now end-of-term evaluation of priority programmes	low	relevance	basic-strategic					
STW <i>ex ante</i> evaluation of projects	medium	application	strategic	✓		✓		
STW <i>ex post</i> evaluation of projects	medium	application	strategic			✓	✓	
final evaluation National Soil Programme	high	basic knowledge	basic-strategic			✓	✓	
innovation-oriented programme carbohydrates <i>ex ante</i> evaluation of projects	high	innovation	strategic	✓				✓
interim evaluation innovation-oriented programme carbohydrates	high	innovation	strategic			✓	✓	
final evaluation innovation-oriented programme carbohydrates.	high	innovation	strategic			✓	✓	
innovation-oriented programme imaging technology <i>ex ante</i> evaluation of projects	high	innovation	strategic	✓				✓
Delft University of Technology selection of Delft interdisciplinary centres of excellence	medium	innovation	basic-strategic	✓	✓			
experimental visitation National Institute for Public Health Studies	high	improve health care	applied			✓	✓	
TNO evaluation of institutes	high	innovation	applied-strategic	✓			✓	

Key: ECOS = Accreditation Committee for the Research Schools; KNAW = Royal Netherlands Academy of Arts and Sciences; NWO = National Research Council; NWO = National Research Council; STW = Technology Foundation; TNO = Netherlands Organisation of Applied Scientific Research; VSNU = Association of the Dutch Universities

Quality is one, but of course still important, aspect. Thus, the use of evaluations in terms of accountability or allocation of resources is limited. Accountability is considered to be realised by having an evaluation system in place. Allocation of resources might be an effect but is not a matter of course and requires a decision by university research management. (In the beginning of the VSNU evaluation scheme, the government did have ideas of linking the evaluation results to allocation of resources, but this has not been implemented in the government funding allocations to universities.) The results of the evaluations by VSNU and ECOS have consequences within the universities. Some universities reward groups with excellent evaluations within the VSNU, and ECOS

accreditation implies protection of the research programme for (in principle) four years.

### Experience with evaluation of societal quality

The review of evaluation practices aimed not just at finding 'how-to-do-it' methods, but also whether and how evaluation of societal quality would be useful and whether it had become part of evaluation repertoires. For each of the evaluation processes we looked at the policy context, at the strategic uncertainty of the research in relation to societal goals, and if societal quality was evaluated, the dimensions that were addressed.

Strategic uncertainty was operationalised along three lines; first the extent to which societal objectives for the research projects, programme, or field had been defined already. We did this by checking the availability and content of any foresight reports and priority setting documents and the content of research programme documents. Second was whether a specific society-oriented mission was defined for the research to be evaluated and, third, the kind of research that was evaluated (basic, strategic, applied). To check which aspects of societal quality were addressed in the evaluation process we analysed the documents of the evaluation, both the documents submitted to the evaluators and the evaluation reports itself.

In this article we will not go into the details of every evaluation process. An overview of the 17 case studies is given in Table 2. After a brief discussion, we shall focus on the three main kinds of evaluations within the set: *ex ante* selection of projects and programmes, research visitations and programme evaluation. For each of these types of evaluation we look at the specific context, analyse the overall evaluation process and the methods to evaluate societal quality, and list the specific indications and indicators which were used.

Societal quality has become part and parcel of most evaluation processes. In 14 of the 17 cases, societal quality is evaluated. The three exceptions are the accreditation of research schools (which are meant to provide a scientific environment for PhDs), the evaluation of the International Institute for Social History (a KNAW institute that has foremost a scientific mission), and the *ex post* evaluation of NWO Priority Programmes. In the last case, the absence of evaluation of societal quality is surprising, as the general mission of the Priority Programmes is to develop knowledge that is relevant for society.

When societal quality is evaluated, it turns out that two or three dimensions are used at best. The main dimensions used are the expectation that the research will contribute to socio-economic developments (relevance), the interaction with (possible) users or other societal actors and the actual use of the results. Relevance was used in most cases, and its use correlated with context of the evaluation process and level of social strategic certainty. Interaction and use appeared to be often evaluated jointly. This is particularly striking in the cases of evaluation of engineering sciences, chemistry, and applied sciences (NIVEL, TNO). These are research domains where the linkages between research and society are already articulated and have been institutionalised.

Evaluation processes rarely take the urgency of the topic or problem into account, and do not assess methods in relation to societal quality. Exceptions to the former are the selection of NWO Priority Programmes and the selection of the centres of excellence at the Delft University of Technology. Both grant schemes aim to address urgent societal, economic and industrial issues. Exceptions to the latter

are the *ex ante* selection of projects within the two innovation-oriented programmes. In both cases the committee assessed the research methods as to their appropriateness in the context of industrial practices.

### *Research programme evaluation*

The first type of evaluation to be discussed in more detail is the interim and *ex post* evaluation of national research programmes. Five examples were included in the set of cases. As all these programmes aim at stimulating 'societal quality' of research, we expected systematic evaluation to occur. In the case of the NWO Priority Programmes interim and *ex post* evaluation of outcomes has not been developed systematically. The protocol for evaluations agreed upon by the board of NWO consists of eight evaluation questions which allow consideration of aspects of societal quality. The conduct of the evaluation is left to the discretion of the respective programme boards, which emphasise scientific results and quality.

SENER has developed a pragmatic approach to mid-term and final evaluations of the programmes: for each programme evaluations are focused on the specific issues considered to be interesting for this particular evaluation. These could be outputs, sometimes also impacts, programme management, or *verankering* (continuing commitments to objectives as specified in the programme, after the programme had concluded). In the case of the Carbohydrate Programme, specific attention was paid to the interaction with industry and the use of the results by industry.

For the evaluation of the societal quality of the National Research Programme on Soil, the Advisory Council for Research on Nature and Environment (RMNO) had commissioned a evaluation study in which a large number of possible users were interviewed about their relation with and their assessment of the programme. Although both the Advisory Council and the Ministry of Education and Sciences (which had asked for the evaluation) were satisfied with the results and the approach, other national research programmes (co-)funded by the Ministry of Education and Sciences have not been evaluated in this way.

We conclude that a best-practice approach has not yet emerged in the Netherlands. The lack of established procedures, methods and criteria for evaluation of societal quality in the context of *ex post* evaluation of research programmes has to do with the general insufficiency in programme development and *ex post* evaluation in the Netherlands.<sup>7</sup> *Ex ante* selection and evaluation through visitation are better developed, and evaluation of societal quality has found a niche there.

### *Selection of technological research projects*

For *ex ante* selection, in particular for innovation-oriented programmes (IOPs) and at the Technology Foundation (STW) where the expected utility of projects from the engineering sciences are evaluated,

**Table 3. Criteria and indications for societal quality in Dutch *ex ante* evaluations**

Criteria (evaluation process)	Indication and indicators
Utilisation (STW in IOP carbohydrates)	Utility in industry, society, technology or science Possibility for development of marketable product by a Dutch firm Importance for long-term technological development Impact on competitive position of Dutch industry Impact on patent position of Dutch industry Potential users
Newness (IOP-imaging)	Possible market position of the product
Application orientation (IOP- imaging)	Technological feasibility of the application Interest of firms in application Are all firms in production chain interested?
Economic relevance (IOP- imaging)	Importance of specific innovation for functioning of the firms Number of firms that profit
Multiplier effect (IOP-imaging)	Possibility that innovation can be transferred to other technologies
Relevance (priority programmes)	Relation with foresight studies or reports of advisory committees Interest from societal actors External (co-)funding Activities for knowledge transfers
Societal relevance (Delft University of Technology)	Relation with priority technologies Definition of the technological problem Technological innovation
User group (Delft University of Technology)	Relevance of user group and their capability to reflect on and enhance utility

Key: IOP = innovation-oriented research programme; STW = Technology Foundation

good practices have been established. In general there is both a (peer) evaluation of scientific qualities and an evaluation of utility. For the latter, evaluators are not asked to assess from a position as a user, with a possible interest in the project, but as an expert in the field who knows the state of the art of the technology. Importantly, relevance and utility, the two aspects of societal quality that are judged here, have to be demonstrated. For relevance, this can occur through references to foresight reports, advisory reports, analyses of patent position and so on. For utility, applicants for grants have to give evidence that potential users exist, that these users are interested, and present concrete actions of knowledge transfer or exchange. Table 3 presents the criteria and indications used in the *ex ante* selection processes we reviewed.

STW has developed its own procedures in which the expected utility of the research is as important as scientific quality, and once a project is funded, an utilisation committee follows its progress. STW's

open competition is not organised by submission deadlines, but by number of proposals (and by availability of funds). Every time 20 proposals have been submitted, each of the proposals is reviewed by five peers. The reviews are sent to the submitters who are able to comment. Project proposals, reviews and researcher's response, for each of the 20 projects, are sent to a lay jury ('lay' in the sense that they need not have any knowledge of the research domain of the proposals) which ranks the projects on scientific quality and on utility. The lay jury is set up anew for each round, has 10 to 12 members and selects in two rounds following a Delphi methodology. Jury members give their evaluations individually. After this ranking, the STW board selects the eight projects with combined highest scientific quality and highest utility. If funds are insufficient to award eight projects (the annual budget of STW is 85 million guilders), further submission rounds are postponed (Van den Beemt and Le Pair, 1991; Van den Beemt, 1997). STW monitors the course and the outcomes of the projects systematically, through utility committees for each project and through interim and *ex post* evaluations. Although these *ex post* evaluations confirm the difficulty to foresee utility and the time when results materialise, they also show quite high returns in terms of concrete results, patents, products and commercial success.

Other *ex ante* selection procedures were analysed, from the Delft University of Technology (TUD) and from IOPs. In these cases, programmes (for TUD) and projects (for IOP) were again evaluated on scientific quality and with the expectation that results would be useful for industry. Evaluation processes here were more like traditional peer review, but included experts from industry among the peers. Within the IOP framework, there is learning and further articulation of methods to evaluate expected utility. The Programme Board of the Innovation-Oriented Programme on Imaging Technology, after the first selection of proposals, felt their criteria were formulated too broadly, and sought more specific indications. Criteria like 'utility', 'newness' and 'economic relevance' remained, but information had to be provided about the technological state of the art, about markets, about the 'fit' to existing production chains, and so on. If this trend continues, the evaluation will resemble research selection practices in industrial research.

In general for engineering sciences, one finds well-established practices to select research programmes and projects and, within these practices, methods and tools to evaluate the expected utility for industry. We consider two main factors to be crucial for the development of these methods and tools. First, the success of STW and its systematic development of a consistent evaluation process, in which utility has been as important as scientific quality. Second, the relatively high societal strategic certainty within the engineering sciences, reflected in an orientation on industry, formal and informal priority setting in co-operation with industry and the availability of experts from industry.

### Research visitations

An important development in the research evaluation system in the Netherlands is the delegation of responsibilities for evaluation to intermediary organisations and, consequently, an emphasis in the evaluation approach on learning by the research management of the university department or research institute that is evaluated. The so-called research visitation approach is characterised by an open protocol, an extensive self-assessment report, and a priori formulation of strict criteria or indicators, but negotiation through disciplinary panels and with the board of university departments of research institutes. There is similarity with the approach to evaluation in the Nordic countries, but it is embedded in the system in its own 'Dutch' way. It is used for evaluation of university programmes by VSNU, for evaluations of the institutes for basic research of the Netherlands Research Council (NWO) and the Royal Netherlands Academy of Arts and Sciences, for the institutes of the Netherlands Organisation of Applied Scientific Research, TNO, as well as for ad hoc evaluations like the NIVEL evaluation and the evaluation of NWO in 1996.

The general approach is used within different contexts, and adapted to the specific needs. As an example, Box 1 presents the protocol for research visitations of the VSNU. We note the involvement of those that are evaluated, which reflects the idea that university departments themselves can and should profit from the evaluations. Indications and indicators for the evaluation criteria, including societal quality, are adapted to the specific requirements and characteristics of (academic) scientific knowledge production, and of the discipline involved. The approach appears to be effective in its aim to focus the evaluations to the needs of university research management. Conclusions of evaluation committees are never ignored, even while there need not be immediate consequences (Westerheijden, 1997).

In general, the combination of self-assessments and site visits by visitation committees appears to be very effective in linking evaluation to research management. The extent to which societal quality is evaluated will then depend on management needs (including credibility and legitimisation pressures) that are felt. Not surprisingly, evaluating societal quality was implemented more easily in the evaluation of institutes with clear missions to produce 'useful' knowledge and technologies — the NIVEL (Institute for Research on Primary Health Care) and the TNO institutes. Societal strategic uncertainty was much lower in these cases, not only due to the mission as such, but especially in the case of NIVEL, the way it was taken up in management. NIVEL's management has implemented procedures (internal reviews, strict rules about contracts to guarantee independent position, and so on) to secure scientific *and* societal quality of the research products. TNO has developed a portfolio approach to enable management of the technological and market position of TNO and its institutes.

#### Box 1. Protocol of the Association of the Dutch Universities research evaluations (translated)

1. The protocol includes a classification of the disciplines as well as a rolling scheme for evaluation of the disciplines.
2. After consultation with the involved departments, the Association of the Dutch Universities (VSNU) determines a time schedule for every evaluation.
3. The directly related standing disciplinary committee of the VSNU nominates two or more candidates as chairman of the evaluation committee as well as deciding upon a profile of the expertise of the committee members.
4. The VSNU appoints a chairman, after consultation of the Royal Netherlands Academy of Arts and Sciences (KNAW). The chairman, in consultation with the KNAW, puts together a committee of which the majority of the members are from abroad. The working language of the committee is English.
5. The directly related standing disciplinary committee of the VSNU specifies within the discipline-specific protocol the *terms of reference* for the committee.
6. Based upon the general protocol and the discipline-specific protocol, the involved university departments make self-assessments of their performances over the last five years and describe their future plans. The unit of evaluation are research programmes. Of each programme, five key publications are put up as part of the self-assessment. In addition, a profile or mission statement of the department is requested.
7. The evaluation committee is requested to judge, for each programme, its quality, productivity, relevance and viability on a five-point scale. For each programme, a brief explanation of the scores is given, which might nuance the general judgement. In addition, the committee gives an assessment of the state of the art of the discipline and of each department.
8. The committee's judgements are based on documents, complemented by interviews with programme leaders and the department's management. Especially in the engineering and natural sciences the committee will make site visits.
9. The report will be finalised and presented to the VSNU chairman, after the departments have had the opportunity to react on the draft report.
10. The costs of the evaluation will be covered by the universities involved in evaluation. The costs depend on the discipline and the discipline-specific protocol. The base costs are Dfl.27,000 per university.

For university research, evaluation of societal quality is less developed. In the case of university law programmes, the protocol was restricted in its scope because of concentration on scholarly output, and only passing consideration of societal quality. The chemistry evaluation, on the other hand, enlarged the scope and developed indications for each of the criteria in the general protocol. The chemistry departments decided to have the visitation complemented with a bibliometric analysis of the impact of their publications and asked for an assessment of quality and development of the chemistry sub-disciplines, rather than of the discipline as a whole. To evaluate the relevance of the projects, research groups were asked to provide indications about research topics and methods, expected contribution to progress of chemistry and other sciences, and expected impact and applicability for technological development. With regard to

**Table 4. Criteria and indications for societal quality in Dutch visitation evaluations**

<i>Criteria (evaluation study)</i>	<i>Indications and indicators</i>
Societal and technological relevance (VSNU self-assessments chemistry )	Relevance of research topic Amount of contract research Number of patents and technological innovations Consultancy relations with industry Opportunity for students to participate in scientific research, Objectives of the programme Participation in national and international programmes (EU, IOP, STW) Possible application domains
Societal and technological relevance (visitation committee chemistry)	Relevance of theme for industrial and technological development Contribution to technological innovation and patents Amount of contract research
Societal relevance (NIVEL)	Mission of institute with respect to application area Advisory committee User involvement in research programme Relation between research programme and main policy documents
Technological maturity (TNO institutes)	Predictability <ul style="list-style-type: none"> <li>• Related technologies</li> <li>• Technological certainty</li> </ul> Interest in technology <ul style="list-style-type: none"> <li>• number of publications</li> <li>• number of congresses</li> <li>• R&amp;D level</li> </ul> Applicability <ul style="list-style-type: none"> <li>• number of publications about applications</li> <li>• Collaboration with universities</li> <li>• Percentage innovating projects</li> </ul> Kind of activity <ul style="list-style-type: none"> <li>• Background of experts</li> <li>• Main research sites</li> <li>• Market</li> </ul> Productivity <ul style="list-style-type: none"> <li>• Financial results</li> <li>• Return on investment</li> <li>• Successful application</li> </ul> Accessibility <ul style="list-style-type: none"> <li>• Availability</li> <li>• number of producers</li> <li>• price</li> </ul> Type of patents Barriers for application <ul style="list-style-type: none"> <li>• Expertise</li> <li>• Education</li> <li>• Patents</li> </ul>
Relation with funding agencies and contract partners (NIVEL)	Satisfaction of main sponsors about research programmes Amount of contract funding of total funding
Competitive position (NIVEL)	Mission in relation with other organisation in public health care Network position
Market position (TNO)	Market reputation of TNO Customer loyalty <ul style="list-style-type: none"> <li>• Percentage continuations of contracts</li> </ul> number of equal competitors Threshold for new entries Kind of competition Role of suppliers Role of clients
Ethical permissibility (NIVEL)	Protocols for clinical trials Procedures for ethical assessments Database registration according to privacy laws Good conduct code for research institutes

*(continued)*

Table 4. (continued)

Criteria (evaluation study)	Indications and indicators
Expertise (NIVEL)	Experiences researchers within application domain Memberships of advisory councils
Output (NIVEL)	Participation in invitational conferences Strategies for dissemination of results (No.) articles in professional journals (No) lectures for non-scientific audiences (No) contributions to policy reports and advisory reports News coverage of research results (No) implementation workshops (No) visits from abroad
Technology position (TNO)	Patent position Licence position Network position Impact Fund-raising <ul style="list-style-type: none"> <li>• acquisition efforts</li> <li>• new contract partners</li> </ul> Time to market internal quality assurance <ul style="list-style-type: none"> <li>• accreditation of good laboratory practice</li> </ul>

Key: EU = European Union; IOP = innovation-oriented research programme; NIVEL = Netherlands Institute of Primary Health Care Research; STW = Technology Foundation; TNO = Netherlands Organisation of Applied Scientific Research; VSNU = Association of the Dutch Universities

productivity the visitation committee did not only look at scientific publications, but also to the number of patents.

What criteria and indications to evaluate societal quality were used in the cases of research visitations we analysed? Table 4 lists the criteria as formulated by the evaluators and evaluatees, and shows the indications and indicators used in practice.

It is clear that the visitation approach, as practised in the Netherlands, is a good vehicle to introduce considerations of societal quality. How this is done then depends on type of research and research organisation. Learning occurs over time, and sometimes also across types of organisation. Overviews like the one shown in Table 4 (and the case descriptions behind them, in Van der Meulen and Rip, 1997) will stimulate such learning. For example, we have pointed out that evaluation societal quality of university programmes can be improved, not by having a standardised set of criteria and indicators, but by expanding the self-assessment report taking into account the specificities of the research domain and the position and mission of the research group. Information could be provided on:

- specification of the strategy of the programme to realise societal quality;
- relevance of research programme with reference to foresight studies, advises, participation in national research programmes;
- transfer activities: contract research, courses, handbooks, workshops, etc.; and
- actual impact of research through patents, realised products, critical role in societal discussions.

A visitation committee can then assess the appropriateness of the strategy with reference to the state of the art of the discipline and its relation to societal practices, and the actual realisation of the strategy, in its own terms. A recent experiment to create indicators and a semi-quantitative research profile, as part of the research visitation of agricultural programmes at Wageningen Agricultural University, showed the limitations of the indicators (Spaapen and Wamelink, 1999). The experiment itself, however, is an indication of the interest in further development of evaluation of societal quality.

## In conclusion

In an earlier and detailed analysis (Rip and Van der Meulen, 1995), we characterised the research evaluation system in the Netherlands as a 'patchwork'. This characterisation is still applicable, but it is clear that there are some interesting 'patches' and that learning occurs within and across them. Our review of 17 cases of evaluation processes focused on the learning with regard to evaluation of societal quality: had methods and tools been developed for the evaluation of societal quality, in what practices, and to what extent do they become part of the evaluation repertoire of the research system?

For *ex ante* evaluation of projects and programmes within engineering sciences (broadly speaking), the STW has developed good practices. In other contexts, similar methods and tools can be and are used, spurred on by the example of STW or for reasons of their own. An interesting recent development is the interaction

and collaboration between the two main funding agencies for medical and health research: the Medical Sciences Council within NWO, focusing on basic and strategic research; and the Foundation for Health Care Research Netherlands (ZON) which has the improvement of public health care through research and implementation projects as its aim. In their *ex ante* evaluations, the question of societal quality is addressed by introducing relevance as a separate criterion. At the moment, the criterion of relevance is grafted on the traditional peer-review-based process (Rip, 1999). When working together, and profiting from the attendant economies of scope, they will learn how to contribute to the realisation of societal quality of the research they fund, and further evaluation methods can be developed including extended and integrated *ex ante* evaluation of proposals.

For *ex post* evaluation through research visitations, the consultation approach and the way this has become embedded in the intermediary layer of the Dutch research system are clearly successful. Evaluation aims to inform research management, at various levels, rather than to focus on accountability and informing the main sponsors only. The further development of procedures and methods is shaped by this context, and they function satisfactorily within this context. It is not yet clear whether the evaluation of societal quality of research can be grafted on the present VSNU approach of evaluating university research through self-assessment and site visits by peer committees. Two elements are important: the diversity in the situation of different disciplines, and sometimes also research groups within a discipline, has to be recognised, and the notion of societal quality has to be sufficiently broad. We have offered some pertinent suggestions in the first section of this article, including the need to locate research domains and research groups and institutes on the dimension of societal strategic (un)certainty.

Evaluation of national research programmes (*ex post*, but also *ex ante*) appears to be a weak spot in the Dutch system. Programme evaluation is often ad hoc, and the learning that is possible — for example, in the series of programme evaluations commissioned by the Ministry of Economic Affairs — has not occurred, or at least has not been capitalised. Part of the problem is that the evaluation of societal quality is quickly reduced to a few basic indicators, focusing on contributions to ‘wealth creation’ and sometimes ‘policy-making’ (to use the terminology common in the UK).

An important lesson about the evaluation of societal quality is that the search for indicators and criteria to be used across the board is backgrounded. This has happened because of specific characteristics of the Dutch policy context, with its dense intermediary layer and emphasis on consultation (Van der Meulen and Rip, 1998), but the value of the message does not depend on that context. Briefly, the mission and strategies of the unit or programme to be evaluated must be foregrounded. Indications and indicators are drawn

from an evaluation repertoire, as relevant to those missions and strategies, and to the concerns of the sponsor or sponsors of the research. This is not to say that units can neglect societal quality. Indeed, there is pressure to include societal quality, and in an operational way. But the operationalisation should take the nature of the field and the position of the research performing institution into account.

To apply the message productively, it is necessary that the concept of societal quality (and related concepts) is articulated further, and that methods to evaluate societal quality are developed and applied further. The ‘toolbox’ of the evaluation repertoire has to be filled. In the Netherlands, the Consultative Committee of the Sector Councils for Research, backed by the Ministry of Education, Culture and Sciences, have played a key role here. The Sector Council for Agricultural Research continues to be active in articulating linkages between research and relevant sectors of society, and has conducted background analyses and dedicated foresight exercises to this end.

There are pressures to show relevance of research to the various sponsors and to society at large. These will create incentives to evaluate societal quality, but also to focus on quick solutions to the legitimisation problem: publicise good examples, go for easy indicators. This will short-circuit the learning processes, and the best practices which result will not qualify as ‘good’. To be able critically to evaluate actual practices, it is necessary to position them against a general framework, as we have tried to set out in the first section of this paper. (This is a framework, not a standardised set of criteria.) The policy advice is then to induce learning processes from which better practices can evolve.

## Notes

1. See Dunn and Holzner, 1987; Van der Meulen, 1992; Spaapen and Sylvain, 1994; Spaapen, 1995; Garrett-Jones *et al*, 1995; and Turpin *et al*, 1996; *Scientometrics* special issue, 1995; ETAN, 1999.
2. For an analysis of the role of these advisory councils see Van der Meulen and Rip, 1998.
3. This approach is being developed systematically by the UK Economic and Social Research Council in its series of studies on ‘non-academic impact’ of social science research. A report presenting pilot studies is forthcoming, by Spring 2000.
4. Sociologically, hybrid knowledge reservoirs are the general category, and scientific fields somewhat distant from practice a particular and ‘purified’ version (Rip, 1997).
5. The ‘compass card’ method introduced by Laredo and collaborators of the Centre Sociologie de l’Innovation, École des Mines, Paris, and the similar ‘Societal Quality Research Profiles’ of research institutes presented by Spaapen and Sylvain (1994) enable reflection on, and evaluation of, the mission of research institutes along a number of dimensions, but do not address the question of societal quality of research (projects, results).
6. Whitley (1984) applied this (and other) concepts from the sociology of work and organisations to the sociology of scientific fields. Now that strategic science has become widespread (in practice as well as in ideology), hybrid research and application domains can be analysed in a similar way.
7. This comparative assessment of the Dutch situation is based

on discussions in two networks on evaluation and other science and technology policy instruments funded by the European Union. Results have been published in a special issue of *Research Evaluation* (1995) on evaluation in Europe and in the final report of the ASTPP network (ASTPP, 1999).

## References

- ASTPP (1999), *Distributed Intelligence in Complex Innovation Systems*, Final report of a European Union supported Network on Advanced Science and Technology Policy and Planning (Fraunhofer Institute for Systems and Innovation Studies, Karlsruhe).
- William N Dunn and Burkart Holzner (1987), 'Introduction: toward knowledge systems accounting', *Knowledge*, 9(2), pages 163-167. Introduction to a theme issue on knowledge systems and impact indicators.
- ETAN (European Technology Assessment Network) (1999), *Options and Limits for Assessing the Socio-Economic Impact of European RTD Programmes* (Office for Official Publications of the European Communities, Luxembourg). Working Paper prepared by an independent ETAN Expert Working Group for the European Commission, Directorate General XII, Programme Evaluation Unit.
- Sam Garrett-Jones *et al* (1995), *Using Basic Research: Assessing Connections between Basic Research and National Socio-economic Objectives. Part 1: Review of Current Theory and International Practices*, Commissioned Report 36 (National Board of Employment, Education and Training, Canberra).
- John Irvine and Ben R Martin (1984), *Foresight in Science: Picking the Winners* (Frances Pinter, London).
- OECD (1992), *Technology and the Economy: The Key Relationships* (Organisation for Economic Co-operation and Development, Paris).
- Arie Rip (1997), 'A cognitive approach to relevance of science', *Social Science Information* 36(4), pages 615-640.
- A Rip (1999), 'Aggregation Machines — A Political Science of Science Approach to the Future of the Peer Review System,' in *Policy Studies Annual Review 1998*.
- A Rip and B J R van der Meulen (1995), 'The patchwork of the Dutch evaluation system', *Research Evaluation*, 5(1), pages 45-53.
- Scientometrics* (1995), Special Issue on First International Conference on the Evaluation of Research, Technology and Development, *Scientometrics* (34), pages 317-541.
- J Spaapen (1995), *Research for society, towards a method for the assessment of research in a policy context*, PhD Thesis, University of Amsterdam.
- J Spaapen and C Sylvain (1994), *Societal Quality of Research: Toward a Method for the Assessment of the Potential Value of Research for Society*, SPSP Concept Paper No. 13 (Science Policy Support Group, London).
- T Turpin (1996), *Using Basic Research. Part 2: Socio-economic Connections to Academic Research in Australia*, Commissioned Report 45 (National Board of Employment, Education and Training, Canberra).
- F C H D van den Beemt (1997), 'The right mix: review by peers as well as by highly qualified persons (non peers)', in Australian Research Council Commissioned Report, No.54, *Peer Review Process*, pages 153-164.
- F C H D van den Beemt and C. Le Pair (1991), 'Grading the grain: consistent evaluation of research proposals', *Research Evaluation* 1(1), pages 3-10.
- B J R van der Meulen (1992), *Evaluation processes in science: the construction of quality science, government and industry*, PhD Thesis, University of Twente.
- B J R van der Meulen (1995), 'Understanding evaluation processes in research systems in transition', *Science Studies*, 8(1), pages 24-35.
- B J R van der Meulen and A Rip (1995), *Indicatoren en indicaties voor de beoordeling van maatschappelijke kwaliteit van onderzoek*, Eindrapport, Commissie Overleg Sectorraden (Centrum voor Studies van Wetenschap, Technologie en Samenleving, Universiteit Twente, Enschede).
- B J R van der Meulen and A Rip, (1997), *Maatschappelijke Kwaliteit van Onderzoek Tussen Verantwoording en Management: Een inventarisatie van beoordelingspraktijken*, Report commissioned by the Ministry of Education, Culture and Sciences (Universiteit Twente).
- B J R van der Meulen and A Rip (1998), 'Mediation in the Dutch science system', *Research Policy*, 27, pages 757-769.
- Adrian C L Verkleij (1998), *Because every situation is different ... A contribution to the discussion of the pros and cons of large-scale research evaluations*, Paper presented to the meeting of the Confederation of European Union Rectors' Conferences, Brussels, 6 February.
- F J M Wamelink and J B Spaapen (1999), *De evaluatie van universitair onderzoek: Methodiek voor het incorporeren van de maatschappelijke waarde van onderzoek* (Nationale Raad voor Landbouwkundig Onderzoek, en Commissie Overleg Sectorraden, Den Haag).
- D F Westerheijden (1997), 'A solid base for decisions: use of the VSNU evaluations in Dutch universities', *Higher Education*, 33, pages 397-413.
- R Whitley (1984), *The Intellectual and Social Organization of the Sciences* (Clarendon Press, Oxford).