PBL WORKING PAPER 19 **JUNE 2014**

Evaluating the societal quality of research carried out by PBL Netherlands Environmental Assessment Agency

Bert de Wit*[1], Femke Merkx[2]

1.PBL Netherlands Environmental Assessment Agency, P.O. Box 30314, NL 2500 GH The Hague, The Netherlands, e-mail: bert.dewit@pbl.nl
2. Kenniscocreatie, onderzoek en advies, The Hague, e-mail: Femke.Merkx@kenniscocreatie.nl .

***Corresponding author**: *bert.dewit@pbl.nl*

Abstract

The evaluation of the societal quality of research is an important issue for PBL Netherlands Environmental Assessment Agency as its mission is to provide policy relevant knowledge. There are several methodological problems with regard to the assessment of the societal quality of research. PBL decided to make use of the ERiC Guide for Evaluation of Research in Context which has been proposed by various Dutch research organisations. This paper describes how the ERiC-Guide was used and what indicators have been constructed taking the mission of PBL into account.
We also present some lessons that we have learned by using this Guide in connection to the 2012 international scientific audit of PBL. Both from the perspective of PBL and the audit committee that visited PBL, suggestions for possible improvements of indicators and procedures are made which still need to be assessed for their practicality. Finally, we present our conclusions which can be used for future evaluations.

***Key words**: societal quality of research, evaluation methods, indicators, policy analysis*

1. Introduction

The aim of this paper is to record the experiences with the evaluation of the societal quality of research by PBL Netherlands Environmental Assessment Agency and formulate some conclusions that can be used for future evaluations. We describe the process of identifying suitable indicators for the evaluation of the societal quality of research, the lessons learned and suggestions for improvements. These experiences date from 2012 when an international audit committee reviewed the scientific quality of the work of PBL. Societal quality of research is created when research is connected to societal practice. Interactions between knowledge producers and users are supposed to enhance the uptake and use of the resulting knowledge by stakeholders (cf. Meijer, 2012, Spaapen and Drooge, 2011).

The evaluation of the societal quality of research has gained attention over the past decade in the Netherlands. Initially the Committee of Sector Councils in The Netherlands stimulated the development of ideas on how to evaluate the societal quality of research, as this topic was one of their primary concerns (Spaapen, Dijstelbloem and Wamelink 2007). The subject was taken further on board by other Dutch organisations, such as the Rathenau Institute and the KNAW, The Royal Netherlands Academy of Arts and Sciences. In collaboration with NWO, (the Netherlands Organisation for Scientific Research), the VSNU (Association of Dutch Universities) and the HBO-Raad/Vereniging Hogescholen (The Netherlands Association of Universities of Applied Sciences), a prototype of a Guide for the evaluation of the societal quality of research was produced and tested in various research institutes. Since 2010, this guide is available: *Evaluating Research in Context (ERiC)* as part of the Standard Evaluation Protocol (SEP)(KNAW 2009). The SEP was originally focused on the scientific ratings of publications and researchers. The ERiC Guide is considered as a way to produce evaluations that take the missions of research institutes and their context into consideration. KNAW has produced specific reports on the evaluation of the engineering and social sciences and the humanities (cf. KNAW, 2013). Also in other countries, methods to evaluate the societal quality of research have been developed. An example of such an evaluation method is the RQF[1] case study method that has been officially designated by the former government of Australia for evaluating the societal quality of research. However, this political decision has not been implemented.

In the discussion on the societal quality of research sometimes the question is raised whether this kind of evaluation should be addressed separately or as an indispensable part of the scientific review procedure. For research organisations like PBL, that have a societal mission, it is obvious that it is an indispensable part of the scientific review procedure as PBL is a research institute with a mission to produce policy-relevant research[2]. PBL is an independent institute for policy analysis in the field of spatial planning, nature and the environment. PBL produces knowledge for its clients, often from the Ministry of Infrastructure and the Environment, but also from other ministries, parliament and political parties. The knowledge produced is also used frequently by other groups in society (see the *PBL self-evaluation report* 2012). The evaluation of the quality of the research of PBL cannot be limited to the scientific performance (ratings of publications, H-index, etc.) of the researchers and research groups of the institute. Equally important is the question to what extent PBL succeeds in fulfilling its mission to perform policy relevant research.

This motivated PBL to look for methods to evaluate the societal quality of its research in addition to the more conventional methods for evaluating scientific quality.

In the following sections we will discuss

- the use of PBL products;
- the impact of products;

[1] RQF= Research Quality Framework

[2] See: http://www.pbl.nl/en/aboutpbl. PBL is funded for a large part by the Dutch government; the researchers who work for PBL are officially linked to the Ministry of Infrastructure and the Environment.

- the ERiC method for the evaluation of the societal quality of research and how it has been operationalised for PBL;
- the lessons learned, both by those preparing the PBL audit and by the audit committee;
- possible additional, more specific indicators;
- possible alternative evaluation approaches.

But first, we want to discuss what we know about the use of PBL products from earlier evaluations.

2. What do we know about the use of PBL products?

To get an idea of the use of PBL products, it is necessary to know what position PBL has in the science and political systems in The Netherlands and who are its main target groups and audiences. Halffman and Hoppe (2003) characterise policy analysis agencies like PBL as the 'linesmen of politics', demarcating the political playing field in Dutch politics. They structure the political discussion. Their function is above all national. National policy makers, politicians and parliament are the primary target groups, but societal groups, businesses and researchers are often also interested in PBL products. It should be noted that PBL also plays a role at the international level, for example working for the OECD[3], the European Commission and the IPCC. So tracking the use of PBL products should not be limited to Dutch audiences. The audiences for PBL products may be: policy makers or politicians, societal groups or businesses, or researchers, at a national, international or local level.

Over the last decade the policies on spatial planning, nature and the environment have been decentralised to a large extent in The Netherlands. As a result provinces and municipalities are the main authorities responsible for the elaboration and implementation of these policies. PBL at the moment is trying to find out what the on-going process of decentralisation means for its services and products and what it might reasonably be expected to do for provinces and municipalities The growing importance of these other governance levels besides the national level makes it more difficult to find out how PBL services and products have been used. It means that a larger number of actors in different positions and at different levels of government and public administration can be considered as potential users.

PBL produces various products. PBL produces for example *policy evaluations* (on spatial planning, nature and the environment), but also *outlooks* based on scenario studies, reports on *strategic issues* and *methodological* reports. Policy evaluations are typically intended to inform political discussions in parliament on the extent to which policy goals have been attained and what may explain possible implementation deficits. Outlooks are intended to play a role in the formulation of strategic policy alternatives in the initial stages of the policy cycle. Methodological reports have still another goal, viz. providing information on the methods and analyses used in PBL studies. The type of products also depends on the preferred dissemination strategy. For some important studies, special presentations have been organised for high ranking officials. In rare cases, the output of a project has been a presentation for policy makers and no official report has been published. But as PBL is part of the public research infrastructure in The Netherlands, it usually produces reports that are available to the general public on its website.

Reports can be used in different ways. Below we mention some examples of the use of PBL reports to illustrate the diversity of uses and audiences. Formal, as well as informal information on the use of PBL's products and on the role they play in the political discussions and the formation of opinions can be obtained for example by registering official reactions of ministers or politicians to reports (formal) or by collecting reactions of policy makers and other audiences, e.g. during presentations (informal).

[3] For example, PBL contributed to the OECD Environmental Outlook to 2050.

An example of a report that played an important role in political discussions, is the *Quick Scan of variants of the Ecological Main Structure* (PBL 2011). This report appeared at a moment when nature conservation policy was heavily debated in society and the government had announced a major shift in nature policy. The government no longer guaranteed that the Ecological Main Structure would be completed in 2018 as was originally planned. The budget for its construction was reduced. The quick scan of variants gave the discussants an impression of the ecological consequences of several alternatives to the planned Ecological Main Structure. This report was one of the most frequently cited reports of PBL on the internet in the period 2008-2012. Another example is the report on *Urban Outskirts* (Hamers *et al.* 2009), a report that was readily used by the policy makers that requested it. They welcomed the concept of urban outskirts as a transition zone between 'red' and 'green' with a lot of activities taking place in it. They asked the researchers to organise a conference on urban outskirts to draw attention to the issue and even invited the researchers to become a member of a jury that had to decide on proposals for regenerating urban outskirts. A third example is the report on *particulate matter* (MNP[4] 2005). This report on a politically much debated subject in air pollution policy ('nightmare dossier') has been used by many people who are somehow involved in this issue as a point of reference for reframing policy. The report also led to a revision of the monitoring system on particulate matter of RIVM, the National Institute for Public Health and the Environment (De Wit *et al.* 2014).

Sometimes, the actual use of reports differs from what the authors expected. Both in the sense that non-intended target groups use a report or that it is used in a way for which it was not intended. General scenarios that have been produced by PBL and the CPB Netherlands Bureau for Economic Policy Analysis for national policy makers (such as *'Welfare, Prosperity and the Quality of the Living Environment'* 2006) have been used by local authorities to produce forecasts of the number of houses, offices and premises needed in the future. Of course, this is not the intention of scenarios, that *describe* possible futures and do *not predict* them. One of the scenarios had even been translated to a spatial scale of square meters, which was not allowed from a scientific point of view, thus creating a false sense of certainty for political decisions (PBL internal communication 2011). An example of the use of products by originally non-intended target groups are the *SCENE* scenarios developed by RPB[5] for the national level. These scenarios have been used for discussions on the future development of the province of North Brabant with representatives of various societal groups and businesses (internal communication 2011).

On the basis of what we know from PBL researchers about the use of PBL products we conclude that their use depends on the type of product and the target groups, but as shown above, products are sometimes used by non-intended audiences, or in a way that is not anticipated. Therefore we need an additional method to find out who are the users of PBL products. We cannot confine the analysis to the intended audiences only. In order to identify the actual use of PBL products and its effects on users we have to look for proxies, since a detailed investigation of these issues over the various audiences requires too much effort, while some information moreover will probably never come to the ears of PBL researchers as the actual use may be coincidental and not the consequence of planned interactions.

3. Impact of research

In the discussion about the societal quality of research, some discussants refer to the concept of '*impact'* as a possible yard stick to measure the societal quality of research. The British system of funding of research considers 'impact' of research an important factor in funding decisions (KNAW 2013). Impact can be understood in various ways and may be demonstrated in various spheres of utilization. What counts as a proof of 'impact'

[4] MNP is one of the predecessors of PBL Netherlands Environmental Assessment Agency.
[5] RPB is the other predecessor of PBL Netherlands Environmental Assessment Agency

may consequently vary considerably. Duryea (2007) (Australia) defines 'impact' not only as 'use' of research products in a broad sense, including citation by various audiences, but also as influencing political decision making or changing societal practice. Meijer (2012) links 'impact' to societal effects of research and to the mission of the research institute.

Ideally, a research organisation like PBL that produces policy relevant research, would like to have an indication of its impact. However, because of problems of attribution and temporality it is often difficult to point out what the impact of a specific product is or has been. The problem of attribution is based on the fact that political decisions are rarely influenced by one report or one scientific advisor only (Hoppe 2008; referring to C. Weiss 1991). Therefore it is difficult to attribute political decisions to a specific report. Meijer (2012) notes that systematic datasets that can be used for this kind of evaluation are often lacking.

The problem of temporality refers to the fact that scientific reports may only after some time be adopted and used by policy makers. It is difficult to point out the influence on policy making of 'outlooks', because it may take some time before policy makers pick up the messages from these types of reports that tackle long term political problems (Dammers 2000; WRR 2010). The temporality problem does not occur to the same extent for various PBL products, as policy evaluations for example fit in a strict time schedule of parliament's agenda. And the attribution problem is also partly relieved by the fact that ministers officially react to policy evaluations and members of parliament give their opinion on these evaluations. But in general one can say that it is rather difficult to find out what the impact of PBL products has been.

As both the concepts of 'use' and 'impact' entail their own problems in research evaluations, it is wise to consider what is available in terms of research evaluation methods and guides. When the Supervisory Board of PBL in 2011 asked PBL to organise a scientific evaluation of its products and suggested to follow the procedures laid down in the Dutch Standard Evaluation Protocol for research institutes, PBL decided to make use not only of the Standard Evaluation Protocol for the evaluation of the scientific quality of its research, but also of the ERiC guide for evaluation of the societal quality of its research. The ERiC guide seemed a workable option for evaluating the societal quality of research.

4. Evaluating the societal quality of PBL research: approach and indicators

4.1 The ERiC-approach

As mentioned in the Introduction, the ERiC guide for the evaluation of the societal quality of research of university institutes is available since 2010. According to the ERiC guide a mix of self-evaluation, case studies and indicators should be used for getting an idea of the societal quality of research carried out by a research institute.

The ERiC guide recommends to produce an organisation-wide self-evaluation report. PBL produced such a report in which the mission of PBL is described as well as its embedding in the Dutch science system and political system, PBL's strategy and the institute's strengths and weaknesses. Furthermore, the report provides information on research programmes, research funds, research collaboration networks and gives an impression of the use and users of PBL products. In addition to the organisation-wide self-evaluation study eight detailed case studies have been produced on projects that have been selected by the audit committee.

Indicators for the societal quality of research need to be easy to understand and should be linked to the main activities and the mission of a research institute. For a research institute like PBL, it makes for example no sense to think of patents as an indicator of the use of research results, as PBL is not an institute of applied research that works for industry. Its primary function is to provide policy makers and politicians with the knowledge they need. So, indicators should be found for example in the number of

citations of PBL reports in parliamentary discussions or references made by policy makers in policy documents, etc.

The underlying idea of the ERiC Guide is that 'productive interactions' between researchers and 'stakeholders' can be considered as a proxy for (future) impact. These interactions are seen as the pathway by which impact is produced, either because the people concerned were involved in the research process (as clients, as stakeholders or as persons who influenced the research agenda) or because they realise they can do something with the results ('valuation'). The benefits can be both societal and/or economic. The definition of productive interactions is:

> *'Productive interactions are exchanges between researchers and stakeholders in which knowledge is produced and valued that is both scientifically robust and socially relevant'*
>
> (definition: cf. SIAMPI approach,
> http://www.siampi.eu/Pages/SIA/12/642.bGFuZz1FTkc.html)

Both the exchange on its own and the result of the exchange are societally relevant and can be considered as proxies of societal impact (Meijer 2012).The word 'productive' indicates that the interaction is expected to lead to efforts by stakeholders to use research results for their purposes. The scientifically robust character of the products refers to the 'credibility' of the research results and the social relevance to their 'saliency' in terms of Cash et al. (2003). To produce scientifically robust research results, scientific review is necessary. 'Salient' research is research into real world problems that produces results that can be readily picked up by the people who want to tackle these real world problems.

There are several ideas about what 'productive interactions' might look like in practice. Interactions can be direct or mediated through publications, exhibitions, design, models, financial support, etc. In the scientific literature, Spaape and Drooge (2011), De Jong *et al.* (2011) and the CWTS research centre of Leiden University (2012) present various possible elaborations of 'productive interactions'. We will discuss these various subdivisions of 'productive interactions' in more detail in section 6.
PBL for its self-evaluation has used the following categories, which are defined in the ERiC guide:

. direct interactions (contacts between persons);
. indirect interactions[6] (contacts through some kind of 'carrier' (texts, exhibitions, models, data, films);
. financial interactions when stakeholders engage in research by financing part of the research.

For a policy analysis agency like PBL this subdivision seems quite manageable.
For the practical application of these categories of interactions to PBL's situation, a reflection is needed on what may count as an indicator for these interactions, seen from the perspective of the mission of the institute.

4.2 Indicators of productive interactions

Indicators of productive interactions must be relevant (linked to the mission of the institute), measurable, reliable and reproducible ("valid"). The mission of PBL excludes beforehand the production of commercially exploitable products, patents, and co-funding by businesses. The following possible indicators of productive interactions in the context of PBL's mission have been listed and used for a pilot (Table 1)

[6] In the ERiC guide (2010) a distinction is made between artefacts (exhibitions, models, websites, etc.) and publications (reports, papers, protocols, educational material, etc.), but we put them together under the heading of indirect interactions.

Table 1. List of indicators of productive interactions

Indicators based on **direct interactions**

- presentations for Parliamentary Committees, activities/projects for the Dutch or European Parliament
- activities on request of political parties
- activities for policy makers in various ministries; presentations for policy makers
- activities for Cabinet Sub Council meetings
- activities and presentations for authorities at a provincial or municipal level and for the European Commission
- presentations during conferences on societal issues, invitations as a key note speaker, etc.
- national and international contacts by participation in relevant networks
- co-production of the research agenda in dialogue with the principal research funders and stakeholders

Indicators based on **indirect interactions**

- discussions about PBL products or references made to PBL in parliamentary documents and in policy documents, references to PBL in parliamentary questions, etc.[7] TV and radio interviews and interviews in the printed press
- references to PBL in the social media, etc.
- citations of reports and activities of PBL on the internet
- website statistics. Number of visits to the website or special thematic websites and number of downloads of reports
- co-production of educational and information material

Indicators based on **financial interactions**

- information on the funding of PBL projects. PBL depends for a large part on government funding, but it can acquire external funding to a maximum of 25% of the total budget. The external funding may come from organisations on other governance levels (EU, OECD or provincial).

When this list of indicators had been drawn up, a pilot was run to see whether there were sufficient data available pertaining to these indicators and if not, whether they were easy to obtain. The research agenda of PBL for example is partly but actively shaped by the discussions among parliament and government, resulting in various assignments for evaluation and in depth studies about issues that in some cases are highly debated in politics as well as in news media and in society. The research agenda is the result of interactions between PBL's Director, several director generals and the secretary general of the funding ministry. So, information on the co-production of the research agenda was available. Data on the year by year work programmes of PBL and multiannual strategic

[7] Politicians can also refer to PBL can without having a specific report of PBL in mind.

programmes could also be produced readily. Parliamentary data banks could be used to trace references to PBL in parliamentary documents.

For some other indicators however, no data were available or data were incomplete. Data about media attention were not complete nor were the data on website visits and downloads[8]. It turned out that in practice, the information that was available about indicators of indirect interactions, was in most cases rather fragmentary.
The Netherlands Institute of Social Research (SCP), a sister policy analysis institute, experienced similar problems in 2008 (KNAW-commissie 2008). To get a more complete and more informed impression of the use of its reports, SCP employed a contextual response analysis. A contextual response analysis consists of an analysis of references made to specific reports in a selected number of databases. In this case, the number of references were scored on the internet at large and also in parliamentary databases and databases containing newspapers articles and radio and television interviews. Not only the number of references was registered, but also the category of actor that referred to a report. Based on this information, references to reports on the internet and in the media were analysed, thus filling the gaps that had been noted with regard to information on the use of SCP reports. PBL decided to commission a similar analysis (see annex II of the *PBL self-evaluation report 2012*).

4.3 The contextual response analysis

For the contextual response analysis, PBL reports[9] were categorised in one of the following four domains:

1. Environment and Nature
2. Sustainable Development
3. Housing, Urban Issues and Spatial Planning
4. Mobility and Infrastructure.

A distinction was made between the following categories of audiences (or product users):

A) government
B) science
C) users active in news media and knowledge dissemination
D) users active in opinion making (including politics, interest groups).

The contextual response analysis yielded a lot of information on the references made to reports on the internet, references to PBL in parliamentary papers, in the printed media and on radio and TV. For the indicator 'references on websites', a subdivision was made into those organisations that repeatedly made references to PBL reports and on the other hand the one-time only references. That distinction makes it possible to see whether PBL reports are mainly used by repetitive so-called frequent users or whether there are a lot of infrequent users.

4.4 Case studies

The ERiC guide envisages not only an organisation-wide self-evaluation, but also a detailed description of case studies. The audit committee selected eight PBL projects for closer scrutiny (see section 6). The project descriptions were streamlined according to a template. Issues on project management were addressed, such as the team and project leader, resources, researchers and possible collaboration with other research institutes. The main research findings and conclusions from project evaluations were presented, but

[8] The reasons for this lack of complete data were the transition to a different registration system in 2010 and the aftermath of the merger of RPB and MNP in 2008
[9] For the contextual response analysis a selection of 40 reports was used that were published in the 2008-2012 period.

also information on the number of publications, references in scientific journals, scientific presentations and the status (H-index) of the researchers that worked on the project. For some projects, the researchers could provide detailed information on the use of the resulting report(s), for other projects this information was not readily available. Instead, the results of the contextual response analysis were then taken as an indication of the use. Citations in parliamentary documents, the printed press and TV, but also the interactions with the client and possible other target audiences are mentioned in the project descriptions. The project descriptions can be found in Annex 1 of the *self-evaluation report* (PBL 2012).The way the audit committee used this project information to get an impression of the quality of the research, is described in section 5.

4.5 The client satisfaction survey

Though not part of the ERiC approach, PBL also initiated a client satisfaction survey. A client satisfaction survey gives an impression of how reports and activities of PBL are valued by clients and other relevant groups in society. It is a welcome supplement to the quantitative data collected by the contextual response analysis, as 'citation' in itself does not give information whether or not a report satisfies the knowledge needs of a client or stakeholder. Nor does it give an answer to the question whether PBL is doing the right things. Citation is an indication of the success of dissemination efforts.
The client satisfaction survey was carried out by the integrity officer of PBL (see *self-evaluation report PBL* 2012). The survey shows that policy makers value PBL reports in different ways, depending on their position in the organisation. A trends report like *The Energetic Society* has been highly appreciated by various policy makers because it opened new perspectives to them. Also the *Nature Outlook 2010-2040* (PBL 2012) was valued because of the 'refreshing views' on nature policy alternatives (citations from the client satisfaction survey). Some policy makers are more interested in facts and figures of policy evaluation reports (the 'Balances' of the environment, spatial developments, nature), while other policy makers are more interested in reports on specific issues like environmentally damaging subsidies or the consequences of an ageing population for physical planning. NGO's (for example 'Natuurmonumenten', a nature conservation organisation), use the figures from PBL policy evaluation reports to put pressure on politics and administration.

5. Lessons learned from PBL's perspective

In retrospect (after the 2012 audit), the following lessons can be learned with regard to the evaluation of the societal quality of PBL's research.

5.1 Use a variety of indicators of productive interactions

From the results of the contextual response analysis, one can conclude that a high score of a particular product on one particular indicator might be accompanied by a low score on another indicator. A report that has been often cited on the internet, may have relatively few downloads from the PBL website and may even not be cited in newspapers, radio or television. Of all 40 reports that were analysed, the *Quick Scan Variants EHS* report got the biggest number of unique internet references by persons and organisations. On the other hand, even though the *Quick Scan Variants EHS* report covered a politically sensitive issue, the attention in the printed media was very limited. Another example: there may be a weak correlation between the number of website hits for a specific report on the PBL website and the number of references to the same report on the internet. This could indicate that other channels than the PBL website may play an important intermediary role in the dissemination of information about PBL reports. For example, knowledge centres which frequently cite from PBL reports, seem to constitute an important additional route for dissemination of PBL products.

The conclusion is obvious: do not use one indicator as the only indicator of productive interactions.

5.2 Infrequent users are an important part of the audiences

The audiences of PBL reports vary considerably. The contextual response analysis showed that there are quite a lot of infrequent users of the 40 reports. Infrequent users are people or organisations that refer occasionally to a particular product of PBL. Especially the reports in English have mostly infrequent users. However, even recurring reports, such as the Nature and Environmental 'Balances' (policy evaluation reports) are cited for a large part by infrequent users.

A comparison of the results of the contextual response analysis with a preliminary survey among PBL managers points out that they have only a limited impression of the use of reports. That is not surprising, as the findings from the contextual response analysis indicate that there is a large group of infrequent users of reports. The contextual response analysis is very useful to get an overall picture of the number of references, and what sort of groups refer to certain types of reports. It clearly produces additional information with regard to the question which target groups have been reached.

5.3 The ERiC approach can be used, but requires specification and interpretation

The systematic approach of evaluation outlined by the ERiC guide for evaluating the societal quality of research can be applied to the context in which PBL functions, if the indicators of productive interactions are adapted to PBL's mission and PBL's context (see e.g. table 1). The indicators will however not be a sufficient source of information. Indicators provide figures which on their own are difficult to interpret for external auditors. Evaluation is not restricted to selecting indicators and collecting relevant data, but also involves interpretation of these figures. For this reason PBL researchers and management wanted to add their interpretation to the collected indicator information, which illustrates that evaluation is always an evaluation from a certain perspective.

The systematics of the ERiC guide also envisage a SWOT-analysis. The analysis of strengths and weaknesses explicitly asked for addressing the developments in the context of the agency, considering possible consequences of changes at the science-policy-society interface (for example budget cuts, openness) and the national and international positioning of PBL as a government funded research institute. Having a unique opportunity to get feedback from a group of distinguished scholars, the PBL management also asked the audit committee to reflect on the strategic choices that PBL had made.

Although the systematics of the ERiC approach make no explicit provision for a client satisfaction survey, it proved to be a valuable supplementary source of information. It not only provided information on the degree of satisfaction with the products that PBL has produced, but also on the needs of clients of PBL and their appreciation of the role of PBL and of its position in the Dutch policy advisory system. Such a more qualitative approach can provide important additional information compared to the approach by means of quantitative indicators.

6. Lessons learned from the perspective of the audit committee

Discussions and reflections on methods and procedures for evaluating the societal quality of research hardly ever consider the perspective of the reviewers. This is a serious shortcoming as in nearly all evaluative procedures it is a committee of peers that integrates the various types of information that are collected during (self)evaluation into an overall assessment of quality. How do committee members deal with all the information that is provided to them? How do they spend their time reviewing different sources of information and what type of information is considered most important? In

this section we look back on the 2012 PBL evaluation from the perspective of the review committee and we draw lessons for future improvement. This section is based on the experiences and observations of one of the authors (Femke Merkx), who served as the secretary of the international audit committee that carried out the 2012 PBL evaluation.

The 2012 PBL evaluation committee consisted of eight international experts, some of them working in academia, some of them affiliated to foreign institutes similar to PBL. They visited PBL for five days, which is a considerable investment of time for people who are in some cases leading research institutes or research groups with heavy obligations in terms of research and teaching. Although a five-day site visit to evaluate PBL seems a substantial amount of time, in practice it appeared to be hardly enough to accomplish the task.

During the visit the Committee interviewed a large number of people, both from within the Agency as well as representatives from PBL's clients and representatives from collaborating research institutes and universities. Also interviews were held with researchers who had been working on the eight projects that were reviewed by the Committee.

With hindsight a few things took considerably more time than was expected. We discuss them here, because they provide important lessons for future evaluations, for PBL as well as for evaluation of societal quality within policy analysis agencies more generally. First of all, there was discussion about the criteria for evaluation, especially concerning the criteria for evaluating scientific quality. Although most schemes for the assessment of societal quality of research assume that this type of evaluation is additional or complementary to the assessment of the scientific quality of research and that these two assessments can be separated, in the case of PBL the assessment of scientific quality and assessment of societal quality appeared to be interrelated. In preparation to the site visit, committee members had been asked to each assess two PBL-projects that had been selected for detailed review. They were asked - in line with the Dutch standard evaluation protocol (SEP) for university research - to assess scientific quality[10] and to look additionally at societal relevance and societal quality[11]. During the committee meetings the notion of scientific quality was subject to ample discussion. Some members argued that a policy analysis agency like PBL should not be assessed in a classical manner on the criterion of scientific quality, because notions that we normally associate with scientific quality, like novelty and contributing to theory development, are not necessarily relevant criteria to assess the quality of PBL's work. Accordingly, whether or not a project leads to a publication in an academic journal can as such not be considered as a distinguishing criterion for scientific quality at PBL. Some policy relevant studies use existing theory and methodology and the results of such studies may be difficult to publish as they do not constitute novel scientific insights. Yet they can be scientifically sound as well as highly relevant for policy makers.

Furthermore, in order for policy advice to have impact, the right timing is often crucial, which means that scientific quality needs to be weighed against timely advice. The Committee brought up this consideration in the assessment of a project on *Environmentally harmful subsidies*. On the one hand the Committee was of opinion that the analyses that were made could have been performed more thoroughly, on the other hand the Committee recognized that there was pressure to deliver the report in time in order to have policy impact. The issue of timeliness was also discussed when assessing

[10] Following PBL's mission and core values (self-evaluation report, 2012) committee members were asked to assess scientific quality by evaluating the following seven aspects: 1) The analysis is scientifically sound; 2) Approach, methods or analysis were published or presented for academic peer-review; 3) Theory and concepts that are used or developed are state-of-the-art; 4)Third parties are able to reproduce the results of PBL; 5)Academic quality of the PBL researchers involved; 6)Scientific quality control of data and analysis provided by external collaborators; 7) The advisory part of the report is accompanied by a clear statement on the conditions under which they are issued

[11] Following PBL's mission and core values (self-evaluation report, 2012) Committee members were asked to assess societal quality by evaluating the following eight aspects: 1) Policy Relevance; 2) Societal Relevance; 3) Quality of science-policy interactions; 4) Timeliness of the advice; 5) If relevant, an integrated approach was chosen; 6) The societal context in which policy takes effect is taken into account; 7)The report addresses and includes the appropriate level(s) of governance (local, regional, national, European, international); 8) Communication of results.

the project *Roads from Rio+20: Pathways to achieve global sustainability goals by 2050*. The Committee concluded that the report was published too late to have immediate impact at important meetings where policy was being prepared.

Eventually the Committee did assess scientific quality, but using its own interpretation which was informed by the specific role of PBL as a policy analysis agency. In the evaluation report it says: 'This aspect (of scientific quality) concerns among others the appropriateness of the chosen approach, the proper use of pre-existing scientific literature and the quality of the argumentation. In addition, it was examined if conclusions follow logically from the study and if uncertainties are addressed in a proper way.'

Here is not the place to go into further detail about the notion of scientific quality in relation to PBL's work. What is important in the context of this article is that the societal mission of PBL, i.c. providing policy relevant studies, influences the way in which scientific quality is interpreted. In other words, whereas most studies on the assessment of societal quality consider the assessment of scientific quality and the assessment of societal quality separately, for the PBL audit committee taking PBL's societal mission as a starting point for the assessment influenced the criteria that were used in assessing scientific quality.

Even so, the Committee still assessed societal quality on itself. It was recognized that PBL carries out various types of studies and that assessment criteria should vary according to the type of study. The notion of 'productive interactions' is for example very different for a scenario study like *Nature Outlook 2020-2040* than it is for a study on *Environmentally harmful subsidies*. During the making of the Nature Outlook, discussions with all kinds of stakeholders took place to find out what future visions they had about nature. From these discussions four visions were produced and further elaborated into scenarios. Then, model calculations were made to assess the impact of these visions on various aspects of nature in the long term. Also, an example was produced of what might be the consequences of the implementation of different nature visions in a specific area to make the scenarios more concrete for stakeholders. So in the case of the Nature Outlook extensive interactions took place in the phase of actual research. This was evaluated positively by the audit Committee, since the involvement of stakeholders can be appropriate to address various ways of normative framing. For other types of studies however interactions during the research stage are not per se an indication or prerequisite for societal quality. They may even be considered undesirable as they might compromise the researcher's independence. This could be the case in policy evaluations, where people involved in designing and implementing policy may have interest in an evaluation leading to a positive judgement. So it can differ from project to project how and when interactions form a relevant criterion for the assessment of societal quality.

Varying the assessment criteria in line with the type of study was possible, since the Audit Committee studied a selection of eight projects in detail. But when reporting productive interactions on the aggregated level of the entire institute the relevant variety disappears from view. This situation could be improved if various project categories could be distinguished for which different types of interactions indicate societal quality. In section seven we discuss some relevant distinctions between projects that could be used to build such a project categorisation.

Experiences from the Audit Committee also revealed that apart from using productive interactions as a proxy for societal quality, expert judgment by the Committee members played a role in assessing societal quality. This was the case for the project *Roads from Rio+20: Pathways to achieve global sustainability goals by 2050*. Based on her own expertise and insight, one of the Committee members argued that societal quality of this project was not optimal as the project used a Western way of framing governance which she considered less relevant for many developing countries.

Not so much time was spent discussing the aggregated list of indicators that were reported in the self-evaluation report. Partly this can be explained by a lack of time, partly also because there was little disagreement about the overall societal quality and relevance of PBL's work. Therefore there was not much reason to elaborately discuss the aggregated list of indicators that was reported in the self-evaluation report. This

apparent lack of discussion on these indicators does not imply that these figures did not influence the overall impression that Committee members hold from PBL. At one point a Committee member did refer to the large number of meetings that PBL holds with parliament and the large number of references to PBL studies in Parliamentary documents as an indicator of societal quality. Also the lack of interaction with industry and private companies was noted. So, in addition to reporting productive indicators on a project level, it makes sense to report indicators for societal quality on an aggregated level. Yet, as stated above, the aggregated use of indicators could be further improved and be made more informative by distinguishing between project categories.

As noted before, some of the tasks of the Committee took more time than was expected. The review committee spent quite some time discussing evaluation criteria. It shows that committee members have no standards readily available for assessing policy analysis agencies. It also appeared that context matters. Not all projects can be assessed using the same criteria. Furthermore, as Committee members were all from institutes abroad, the Committee spent quite some time in understanding the specific Dutch institutional and policy context and the role of PBL within that context. This type of contextual understanding is necessary to make a proper assessment of the societal quality of PBL's work. While working abroad, two of the Committee members were Dutch by origin and for that reason knew more about the Dutch context than others. This was an important factor in the contextual understanding. Furthermore one of the committee members was a leading expert in science-policy interfaces. Her role was important as well, as she could point at some relevant differences between the Dutch science-for-policy culture in comparison to the American culture of scientific policy advice. The lesson to be learned is that indicators for societal quality do not speak for themselves. Therefore a self-evaluation report should not solely report on indicators for societal quality (like productive interactions) but also provide argumentation why these indicators point to societal quality and how they contribute to the overall picture of the societal quality of research.

7. Possible refinements of the 'productive interactions'

In section 4 we described the sort of productive interactions we have taken into consideration and the various indicators for direct, indirect and financial 'productive interactions' that have been used, while in section 5 the lessons learned have been presented. In section 6 we concluded that different types of PBL studies require different assessment criteria adapted to their context. Here we consider the question whether the indicators that have been used can indeed be refined to become more tailor-made. For this purpose we discuss the ideas that have been put forward by De Jong *et al.* (2011) about indicators and the stage in the research process and the ideas of CWTS (2012) about target groups. But it is also clear that for a policy analysis agency like PBL, the policy process itself might also be a source of inspiration for producing more tailor-made indicators.

7.1 The stage of the policy process

A possible refinement of indicators of productive interactions could be to see for what stage in the policy process products are produced. As Kropp and Wagner (2010) argue, the knowledge needs of policy makers in the Agricultural Ministry in Germany change according to the stage of the policy process. The authors found that in some stages, information about for example uncertainties is welcomed by policy makers, but not in other stages.

When the use of strategic reports is studied, it might be feasible to analyse whether specific key elements of the advice can be traced back in policy documents or whether ministries have officially or in informal ways reacted to the strategic report. Perhaps the reaction was to put the issue on the political agenda; or the report has been used by NGOs or business to put proposals forward for new policy arrangements. In addition to

an analysis of documents, also interviews can be useful means to trace the actual use in policy documents and in political discussions. For strategic advice to be of use, the timing of reports and the way the knowledge is transferred, are also important factors, as illustrated in section 6 by the example of the report on *Environmentally harmful subsidies*. The *Nature Outlook 2010-2040* of PBL produced in 2012 (see section 6) was rather soon picked up by national and regional authorities and societal organisations. One can argue that in this case, the timing of the Outlook was perfect. There was a policy window that provided an opportunity, because there had been drastic changes in nature policy by the previous government, and nature policy had become controversial. So, for indicating the societal quality of research, timing and timeliness of reporting and advising is an important factor, which is however sometimes beyond control of researchers.

For policy evaluations like the *'Balances'*, it would seem more appropriate to see whether these evaluations are referred to in parliamentary documents (like parliamentary questions), and in political discussions in the media and on the internet. The supposed pathway to impact might lead to different kinds of use as mentioned above, but also to different audiences having more interest for specific types of products. The findings of the client satisfaction survey also hint at this (see 4.5).

7.2 The type of activity

The type of activity that is asked for more or less determines with whom researchers interact and how frequently.

When a policy evaluation is asked for, interactions between scientists and policy makers are limited because there is always a risk that external observers might get the impression that policy makers try to influence the outcome. But discussions on the format of the evaluation report and the question what policy items should get special attention, are necessary. The kind of policy evaluation (a strict evaluation or a more process oriented evaluation, with or without participation of societal groups) also has to be discussed.

For a methodological report like the one about the Working Group II report of the IPCC assessment, a lot of interactions with scientists and their organisations have taken place, as well as an inventory of the criticisms by the public. So, in that case the interactions with policy makers were very limited, but the interactions with scientists were numerous and the interactions with the broad public limited to an on-line survey of criticisms.

When it comes to strategic reports, interactions with policy makers are very much instrumental and needed to get an idea what product the policy makers are thinking of. The involvement of stakeholders during the research process can also be instrumental for the production of a strategic report and can be used as an indication of productive interactions, if stakeholder involvement serves the aim of the strategic report. The process of the *Nature Outlook 2010-2040* serves as a good example of this.

We conclude that for an evaluation of the societal quality based on the notion of productive interactions, it is necessary to take the type of activity into account. And the type of activity is directly linked to the type of products PBL produces.

7.3 The stage in the research process

De Jong *et al.* (2011) link 'productive interactions' to the 'knowledge dynamics' of the research process. They distinguish the following stages in the research process:

- agenda setting
- research collaboration
- knowledge dissemination
- impact

We discussed the interactions in the agenda setting stage already in section 4 when we discussed the way the work programmes and research agenda of PBL are produced.

These interactions are important, but it is not clear at first sight what possible refinements could be interesting as this is for a part an informal process.
And with regard to research collaboration, we mentioned earlier that interactions with clients and stakeholders for some activities are necessary, but for other activities should be avoided. So, one cannot say in general that for a policy analysis agency like PBL collaboration with stakeholders in research is an important indicator of societal quality.
In general, for the dissemination of research results, several sorts of 'productive interactions' are important. Sometimes, for more strategic products, special outreach activities are undertaken. An example is the way PBL's director presented the *Energetic Society* (2011) to various audiences of policy makers. Information on such outreach activities could be presented in project evaluations and could be used as an indication of the societal quality.

7.4 Refinement of target groups

The Centre for Science and Technology Studies of Leiden University (CWTS 2012) in its research programme explores the possible refinement of the target groups with which productive interactions take place. CWTS makes a distinction in private sector contacts, interactions with other professionals, policy makers and students, with the education sector and with the general public. As PBL is a policy analysis agency and works for governments and not for the private sector, its primary target groups are policy makers, politicians and parliamentarians. Other target groups are the press, societal groups, businesses, international institutions (OECD, EC, World Bank, etc.), other research institutes and the general public. Only for those projects that clearly deal with a broad societal topic for which interaction with societal groups and business is necessary, it makes sense to see what sort of interactions have taken place with these groups.

It might be interesting for PBL to see whether intermediary organisations (advisory bodies, knowledge centres and the like) and representatives of specific interests have a preference for a specific type of reports, for example strategic reports, as they might be more interested in strategic advice than other groups in society. If this hypothesis proves to be correct, it means that the reactions of these audiences are particularly interesting as an indication of the societal quality of strategic reports.

At the moment, there are no committed relationships between PBL and the educational sector apart from PBL professors that occupy a university chair on behalf of the agency or are working part-time for the agency. In the near future, the emphasis on public access of information and enhancement of transparency might lead to more data and models being made available to a greater public. Then it would become interesting to see what educational spin off is produced and which groups make use of it.

7.5 Synthesis of possible refinements in 'productive interactions'

We have summarised possible refinements of indicators for productive interactions in Table 2 below as possible additional indicators (to the ones already presented in Table 1). As the type of report asked for by the client(s) more or less determines the interaction patterns between researchers and other actors involved, with whom they interact and why, it can be taken as a starting point. Thus, we have linked the type of report to the stage in the policy process:

Table 2. Possible additional indicators for the societal quality of research

Type of report	Stage in policy process	Possible additional indicators
Strategic reports (incl. outlooks)	Problem recognition, framing, agenda setting	Interactions with policy makers, involvement of

		stakeholders (why and how) References in parliamentary data base Reactions of political parties References on the internet by interest groups and knowledge centres Timing (window of opportunity) Outreach activities of researchers Presentations for policy makers (national or international) Presentations in parliament (national or EP) Presentations for other audiences (stakeholders)
	Policy formulation	References in policy documents (incl. text references) Official reactions of the client(s) Involvement of researchers in (inter)departmental deliberations on a subject
Policy evaluation reports	Policy evaluation	Official reactions of ministers Reactions of parliament Reactions of provinces, municipalities References on the internet by interest groups
Other products (for example methodological reports)	All stages	Collaboration and interactions with other research institutes /researchers (if appropriate) Involvement of stakeholders (if appropriate, why and how); reactions of these stakeholders Educational spin-off of reports and models

8. Discussion of alternative methods for evaluating the societal quality of research and their consequences

A critical review of the methods employed by PBL for its self-evaluation should also consider other methods for evaluating the societal quality of research. The ERiC approach that PBL used, can be seen as a mixture of the case study approach, the indicator approach and the self-evaluation approach (Rand 2010). Separately, these methods cannot be seen as serious alternatives to the ERiC approach because they each have their own shortcomings and difficulties, for example with regard to the selection of case studies, indicators or the attribution and temporality problems associated with impact assessment.

Depending on the goals of a project, evaluations of research can be done ex-post or ex-ante, or during the research process, as a formative evaluation. If the goal is to raise the awareness of those involved in the project or educate them, a formative evaluation procedure is more adequate. But in most cases, PBL projects do not have such a goal. Consequently it would be too far-fetched to make such a specific category of formative evaluations.

In the next section, we discuss a new approach that still has to be elaborated, the I2S framework (Bammer, 2013).

8.1 The I2S framework for reviewing integrative applied research on real world problems

The experiences from the audit committee point to case study analysis as the preferred approach for the assessment of societal quality. There is however an important disadvantage when looking at case studies. The case study approach is time consuming. Both for those who prepare the case study reports, but also for committee members that need to familiarize themselves with the details of a project. For that reason the case study approach necessarily confines itself to a limited number of studies. That raises the question to what extent an assessment that builds largely on the assessment of a selection of case studies is representative for the institute at large. This in fact was also a point of discussion in the audit committee. In its report the Committee stated that '*although the Committee had no reason to think [the selection] was not [representative, they] had no means to assess whether this sample (of case studies) was indeed representative'.*

One way of dealing with this problem is to not only assess quality per se, but also look at the institutional procedures that are in place to guarantee quality. Such was done indeed for scientific quality. The Committee recommended *'for future scientific audits […] to keep a record of internal and external review procedures. That would give the Committee an additional source of information to assess scientific quality and its control.'* (Report of the audit committee 2013). A similar approach could be taken for the assessment of societal quality. That would however require a framework for reporting and reviewing decisions made in relation to societal quality.

A recent publication by Gabriele Bammer (2013) provides some interesting starting points for such a framework. Bammer sketches a broad framework of issues that need to be discussed when aiming for societally relevant research, in particular when dealing with complex real world problems.[12] These issues range from how to frame the problem, to what academic and stakeholder expertise to involve, how to deal with uncertainties, how to integrate various types of knowledge and how to make sure that research can support practice and policy change.

Bammer's framework serves several purposes. First it provides a shared reference for organising, collecting and discussing various concepts and methodologies that can be of use in integrative applied research for complex real world problems. Thus it is positioned by Bammer as a framework for a new field of study, coined Integration and

[12] Many, though maybe not all of PBL's studies concern complex real world problems

Implementation Sciences (I2S), which aims to provide the methodological toolbox for integrative applied research. Bammer suggests that all projects that aim for an integrative applied research style, should keep record of the decisions made and methods used relating to knowledge synthesis, dealing with uncertainties and having impact on policy and practice. Keeping record serves two purposes: 1) It serves the purpose of mutual learning between projects and research fields and 2) It serves an evaluative purpose as these records can be used from a methodological point of view to assess the societal quality of a project.

Implementing such a procedure for societal quality control would on the one hand form a positive sign that societal quality is taken seriously and on the other hand it would provide an audit committee with a better structured set of information about the societal quality of projects. It would be interesting to set up pilots in various institutes to see what the adoption of Bammer's framework would mean in practice.

9. Conclusions

1. In principle the ERiC guide for the evaluation of the societal quality of research can be very well used for assessing the societal quality of PBL's research. The ERiC method employs a mix of self-evaluation at the institute level, case studies and indicators for societal quality of research. The selection of indicators that are considered relevant for the various activities, products and target groups of PBL, is linked to the mission of PBL as a policy analysis agency.
2. When evaluating the societal quality of research, not only individual projects should be assessed (the 'case studies'), but also indicators for the societal quality of research at an aggregated level are needed.
3. The selection of indicators should be done in a careful way. A variety of indicators is needed to obtain an overall impression of the 'productive interactions' that have occurred with regard to the products and services of PBL. These productive interactions between researchers and 'stakeholders' or 'actors' can be considered as pathways to produce (future) impact of a study, and might result in different kinds of products and services for the use and interest of various audiences.
4. The audiences of PBL reports vary considerably. A contextual response analysis is very useful to get an overall picture of the number of references/citations, and what sort of audiences refer to what sort of reports.
5. A report on the societal quality of research should not only present quantitative information on indicators, but also provide argumentations why these indicators point to societal quality.
6. A client satisfaction survey (which is not part of the ERiC method) provides useful qualitative information, for example on how the role of PBL in the science-policy system is perceived. This information is a welcome supplement to the quantitative information obtained from the various indicators of societal quality.
7. A possible improvement of indicators could be established by distinguishing various product categories and linking different types of interactions to these categories as indications of societal quality. In order to do this for various product categories and services of PBL, the supposed pathways to impact should be made more explicit.
8. Possible additional indicators for productive interactions that might be considered for future research evaluations have been listed. It would be interesting to find out whether such a further differentiation of indicators of productive interactions is possible, especially from a practical point of view.
9. The experiences of the audit committee warrant the conclusion that the committee preferred case study analysis to get an indication of the societal

quality of research. The case studies method however brings up the issue of representativeness.

10. An alternative method to evaluate the societal quality of research might be found in the Integration and Implementation Sciences Approach, based on a framework of issues that come up when doing integrative applied research on real world problems. However, this approach has not yet been applied for research evaluations, so experiments are needed to see whether it works in practice.

Funding

This work was supported by PBL Netherlands Environmental Assessment Agency as part of the project on the international audit 2012.

Acknowledgements

We thank Willemijn Tuinstra (PBL), Ed Dammers (PBL), Peter Janssen (PBL) and Jack Spaapen (KNAW) for their valuable comments.

References

Bammer, G., 2013. *Disciplining Interdisciplinarity. Integration and Implementation Sciences for Researching Complex Real-World Problems.* http://epress.anu.edu.au/titles/disciplining-interdisciplinarity.

Cash, D.W., Clark, W.C., Alcock, F., Dickson, N.M., Eckley, N., Guston, D. H., Jäger, J. and Mitchell, R. B., 2003. 'Knowledge Systems for Sustainable Development'. *PNAS* 100(14): 8086-8091.

CPB, MNP and RPB, 2006. *Welfare, Prosperity and the Quality of the Living Environment.* http://www.welvaartenleefomgeving.nl/context_UK.html. Den Haag

CWTS, 2012. *Merit, expertise and measurement.* CWTS Research Programme 2012-2015. Leiden.

Dammers, E., 2000. *Leren van de toekomst. Over de rol van scenario's bij strategische beleidsvorming.* (*Learning from the future*, in Dutch). Eburon Academic Publishers, Delft.

De Jong, S.P.L., Van Arensbergen, P., Daemen, F., Van der Meulen, B. and Van den Besselaar, P. , 2011. 'Evaluation of research in context: an approach of two cases'. *Research Evaluation* 20 (1), 61-72. doi: 10.3152/095820211X12941371876346

De Wit, B., Kunseler, E., Tuinstra, W., Hage, M., Petersen, A., Tennekes, J., 2014. 'Aspects influencing the roles of PBL researchers working at the science and environmental policy interface. An illustration from PBL practice'. *PBL Working paper 20*, June 2014.

Duryea, M., Hochman, M., Parfitt, A., 2007. 'Measuring the impact of research'. *Research Global* Feb. 2007, p. 8,9, 21.

ERiC , 2010. *Evaluating the societal relevance of academic research in context.* http://www.rathenau.nl/en/publications/publication/evaluating-the-societal-relevance-of-academic-research-a-guide-1.html

Halffman, W. and Hoppe, R., 2003. 'Science/policy boundaries: a changing division of labour in Dutch expert policy advice'. Paper presented to the Sociology of the Sciences Yearbook conference. Basel, December 2003

Hamers, D., Nabielek, K., Piek M. and Sorel, N., 2009. *Verstedelijking in de stadsrandzone. Een verkenning van de ruimtelijke opgave.*(*Urbanisation in the urban outskirts. An exploration of spatial problems,* in Dutch). PBL, Den Haag.

Hoppe, R., 2008. 'Na 'doorwerking', naar 'grenzenwerk'. Een nieuwe agenda voor onderzoek naar de verhouding tussen beleid en wetenschap'. (*After impact assessment towards boundary work,* in Dutch*)*, *Bestuurskunde*, 2008-2, 15-26.

KNAW, 2009. *Standard Evaluation Protocol 2009-2015.* http://www.knaw.nl/Content/Internet_KNAW/publicaties/pdf/20091052.pdf

KNAW-commissie, 2008. *Evaluatie van het Sociaal en Cultureel Planbureau (SCP*) 2004 – 2006. (*Evaluation of the Netherlands Institute for Social Research,* in Dutch*).* KNAW, Amsterdam. Zie ook: Prins, A., 2008. *Evaluatie in meervoud. Een kwantitatieve analyse van gebruik en gebruikers van het werk van het Sociaal en Cultureel Planbureau 2004-2006.*(*Evaluation in plurality,* in Dutch) http://www.knaw.nl/Content/Internet_KNAW/actueel/bestanden/Bibliometrisch_onderzoek_SCP_output.pdf

KNAW, 2013. *Towards a framework for the quality assessment in the social sciences. Appendix 2 Societal relevance as a criterion for quality assessment*. KNAW Amsterdam.

Kropp, C. and Wagner, J., 2010. 'Knowledge on stage: Scientific Policy Advice'. *Science Technology & Human Values* 35: 812, originally published online 21 May 2010. DOI: 10.1177/0162243909357912

Meijer, I., 2012. *Societal returns of scientific research. How can we measure it?* CWTS-WP-2012-014.

MNP, 2005. *Fijn stof nader bekeken. De stand van zaken in het dossier fijn stof (A closer look on particulate matter,* in Dutch*)* MNP report 500037008, Bilthoven.

OECD, 2012. *Environmental Outlook to 2050: The consequences of inaction.* http://www.oecd.org/env/indicators-modelling-outlooks/oecdenvironmentaloutlookto2050theconsequencesofinaction.htm

PBL, Stockholm Resilience Centre and Stockholm University, 2009. *Getting into the right lane for 2050, A primer for EU debate.* http://www.pbl.nl/sites/default/files/cms/publicaties/500159001.pdf

PBL Netherlands Environmental Assessment Agency, 2011. *Herijking van de Ecologische Hoofdstructuur. Quick Scan van varianten.* (*Quick scan of variants of the national ecological network,* in Dutch). PBL, Den Haag.

PBL Netherlands Environmental Assessment Agency, 2012. *Nature Outlook 2010-2040*. PBL, Den Haag.

PBL Netherlands Environmental Assessment Agency, 2012. *Uitkomsten klantwaarderingsonderzoek* (*Results of a client satisfaction survey. Main points in the PBL self-evaluation report,* in Dutch*)*. PBL, Den Haag.

PBL Netherlands Environmental Assessment Agency, 2012. *Self-evaluation Report May 2008 – May 2012.* PBL, Den Haag. http://www.pbl.nl/node/56185

Report of the audit committee, 2013. *International Scientific Evaluation of PBL 2008-2012.* http://www.pbl.nl/sites/default/files/cms/International-Scientific-Evaluation-of-PBL-2008-2012.pdf

Spaapen, J., Dijstelbloem, H, Wamelink, F. , 2007. *Evaluating research in context. A method for comprehensive assessment.* COS, The Hague.

Spaapen, J.B. and Drooge van, L., 2011. 'Introducing 'productive interactions' in social impact assessment'*. Research Evaluation*, 20, 211-218. ISSN 0958-2029.

Weiss, C., 1991. 'Policy Research: data, ideas, or arguments?' In: Wagner, P., Weiss, C., Wittrock, B. and Wollmann, H. (Eds): *Social Sciences and modern states: national experiments and theoretical cross-roads,* Columbia University Press, Cambridge, pp. 307-32.

WRR, 2010. *Uit zicht: toekomstverkennen met beleid.* (*Out of Sight. Future studies and policy,* in Dutch). WRR Verkenning 24, Den Haag.