# QUALITY ASSESSMENT IN THE DESIGN AND ENGINEERING DISCIPLINES

Royal Netherlands Academy of Arts and Sciences
PO Box 19121, NL-1000 GC Amsterdam
T +31 (0)20 551 0700
F +31 (0)20 620 4941
knaw@bureau.knaw.nl
www.knaw.nl

First edition: November 2010
Second edition: April 2011
pdf available on www.knaw.nl

Basic design edenspiekermann, Amsterdam
Typesetting: Ellen Bouma, Alkmaar
Printing: Bejo druk & print, Alkmaar

Foto cover: Nationale beeldbank/Taco Gooiker

ISBN: 978-90-6984-620-0

The paper for this publication complies with the ∞ iso 9706 standard (1994) for permanent paper

Mixed Sources
Productgroep uit goed beheerde bossen en andere gecontroleerde bronnen.
www.fsc.org Cert no. CU-COC-804134-N
© 1996 Forest Stewardship Council
FSC

This publication is made from FSC-certified paper with number CU-COC-804134-N.

# QUALITY ASSESSMENT IN THE DESIGN AND ENGINEERING DISCIPLINES

## A SYSTEMATIC FRAMEWORK

Royal Netherlands Academy of Arts and Sciences
Advisory report KNAW TWINS Council

# FOREWORD

One of the great merits of science is its natural diversity. It is a diamond with many different facets. Its variety can be seen in the subjects studied, the research methods used, and the way results are ultimately reported. At its most profound, it reflects the rich variety of phenomena that fill our world and the depth and breadth of our minds.

Science deserves to showcase its merits in a way that is as objective as possible but also respects, if not celebrates, its variety. The highest standards of quality assurance are needed not only to build support for research beyond the scientific community, but also because transparent assessment of research is beneficial for scientists themselves. The challenge is to ensure that the method used to measure quality is compatible with the intrinsic value of the relevant field of research. There is much work ahead in that respect.

One crucial first step is to survey scientific publications and their citations. According to this criterion, the Netherlands has world-class researchers, an assertion recently confirmed by the report *Science and Technology Indicators*, published every other year at the behest of the Dutch Ministry of Education, Culture and Science. In terms of citation impact, the Netherlands is at the top of the world rankings, along with the United States, Switzerland and Denmark. Our research community is also exceptionally productive.

It is of course gratifying to hear that Dutch scientists are among the best in the world according to these measures, but it is important to realise that research quality cannot be measured *solely* on the basis of scientific publications and citation impact. In many fields, citations tell only half the story or less, and the standard assessment methods miss out on large areas of research as a result. For example, the products of the design and engineering disciplines consist not only of peer-reviewed journal articles, but also conference proceedings, designs, software and structures. In the humanities, many publications are in book form, or in a language other than English, making them virtually invisible in terms of citation impact. These disciplines come in for a harder time when the customary quality assessment measures are used. It is no accident, then, that the Academy has been asked to advise on practical assessment criteria precisely for these disciplines.

In 2008, the Academy argued in its advisory report *Kwaliteitszorg in de wetenschap; van SEP naar KEP* [Quality Assurance in Research: From SEP to CEP] that any new quality evaluation protocol must offer enough flexibility to accommodate differences between disciplines. Researchers who wish to emphasise the societal and cultural relevance of their work or its economic value should be able to do so, for example. The Academy was therefore delighted to advise on this matter. The present advisory report responds to the call for a more flexible assessment method for the design and engineering sciences. I am pleased that the committee has concluded that a separate set of criteria is not necessary for these disciplines. Research quality and societal relevance are sufficient. The discipline-specific factors are reflected in the indicators to be used to assess these two criteria. This is an important guiding principle that emphasises both the universality and the diversity of research at one and the same time.

The committee indicates that it has very deliberately not selected or weighted indicators. That is something that scientists themselves must do in the various assessment situations, in close consultation with university administrators. I hope that 3TU Federatie and design and engineering researchers will waste no time in accepting this challenge so that from now on, Dutch science can sparkle in all its diversity.

Robbert Dijkgraaf
President of the Royal Netherlands Academy of Arts and Sciences

## Second edition

In the second edition some textual corrections have been made. Recommendation 4 on page 45 has been moved to the section about research funding bodies. In table 3.1 the header ‘Other output’ has been replaced by ‘Designed artefacts’.

# CONTENTS

# SUMMARY

## Background

3TU Federatie – an alliance between the three Dutch universities of technology – has asked the Royal Netherlands Academy of Arts and Sciences to advise on the criteria used for ex-ante and ex-post assessments of research output in the design and engineering disciplines. Scientists in design and engineering regularly encounter problems in the assessment of the quality of their research output, whether that assessment takes place within the context of an external evaluation, an academic appointment or promotion, or an application for funding. The quality indicators used in such situations are borrowed from the more basic sciences (publications in ISI journals, impact factors, citations, the Hirsch Index) and are, in the eyes of these scientists, inadequate. First of all, they argue, their output consists not only of peer-reviewed international publications, but also of conference proceedings, designs and works of engineering. Secondly, they point to the fact that their research is generally more context-specific and multidisciplinary than that carried out in the more basic sciences. Design and engineering journals therefore have a lower impact factor and are consistently given a lower rating.

The Royal Academy installed a committee to look into this question, chaired by Prof. A.W.M. Meijers (Eindhoven University of Technology). The committee's assignment was to draft an advisory report on the criteria to be used for the ex-ante and ex-post assessment of:

- design and engineering activities in technical disciplines that can be considered scientific in nature;
- research in the design and engineering disciplines.

The criteria had to satisfy the following requirements:

- they had to be useful to organisations that fund research activities (NWO, etc.);
- they had to be useful to universities in assessing academic staff (researchers/designers);
- they had to be credible within an international context.

In the present advisory report, the committee proposes a set of pertinent criteria and reports on its findings. The report is based in part on interviews conducted with 27 scientists working in various disciplines and on a survey of the systems used and lessons learned abroad.

## General criteria with discipline-specific indicators

One important point for the committee is that the assessment standards for a particular discipline should not differ completely from those used in other disciplines. That would make such standards arbitrary and opportunistic. The committee has therefore attempted to develop an assessment framework consisting of two elements:

1. generally applicable quality criteria;
2. discipline-specific indicators for those criteria. These indicators provide empirical information on the extent to which a person, group or research proposal satisfies the quality criteria. Such information may be either quantitative or qualitative in nature.

Contrary to what it had expected at the start of its advisory processes, the committee has concluded that assessments of research output quality in the design and engineering disciplines can be based on the two criteria used to assess output in other scientific disciplines: 1) research quality and 2) societal relevance. In other words, supplementary criteria are unnecessary for design and engineering. The committee's international survey has shown that these two criteria are also considered credible for these disciplines in other countries.

Although no additional or different quality criteria are necessary, that does not mean that every scientific discipline can be assessed in the same way. Assessing quality should be a question of fine-tuning owing to the differences between disciplines (including their publication cultures), categories of scientific activity (designing, research) and assessment situations (external evaluation, appointment, research proposal). Such fine-tuning is expressed in the discipline-specific indicators selected under the criteria 'research quality' and 'societal relevance', and the relative importance assigned to each indicator. Table 3.1 summarises the indicators that the committee regards as important for the design and engineering disciplines.

## Following stage

The committee has explicitly *not* chosen to narrow down the choice of indicators for the design and engineering sciences, or to determine their relative importance. That is the following stage, and the necessary decisions must be taken by the scientists themselves (peers) in each of the various assessment situations, in close consultation with university administrators. The committee advises the board of 3TU to create sufficient scope for discipline-specific quality assessment in the technical sciences, and to ask the design and engineering disciplines to determine the indicators and their relative importance for assessing quality in those disciplines.

## Difficult to compare different fields

It became clear during the interviews that virtually no one has encountered problems when quality assessments of research output are conducted by internationally respected peers. The problems arise when comparisons are made *between* disciplines, or when the disciplines themselves are too broad. The committee believes that quality benchmarks can only be meaningful when they are made *within* disciplines. As appealingly simple as administrators may find the 'one size fits all' approach to assessing quality, it does not do justice to significant differences between disciplines and will therefore always give certain ones an unfair advantage. The committee advises research funding bodies such as the NWO and the STW to devote more attention to programmes focusing on disciplines that do not fit easily into existing categories or the present quality assessment method. This applies in particular to the design and engineering sciences; the quality indicators used for these disciplines must do them justice.

## Peer review crucial

The committee believes that proper quality assessment should be based on peer review, and it applauds the fact that most research funding bodies make use of the peer review system. It is vital, however, to select the right peers, as this will influence the outcome of the assessment. The peer selection processes used by the NWO, external evaluation committees and appointment advisory committees should be more objective and transparent. Peers involved in assessing quality in the design and engineering sciences must themselves be assessed according to a broad range of indicators, and not only on their publication behaviour.

## Importance of publications

This advisory report proposes using indicators other than peer-reviewed publications to assess quality. The committee emphasises, however, that peer-reviewed publications are also important for the design and engineering disciplines, and will remain so. Such publications encourage mutual quality checks, help disseminate knowledge, and contribute to the 'scientification' of disciplines. Scientists working in these disciplines or in sub-disciplines do not always publish as a matter of course, however. The committee advises them to lend their full support to a culture of peer-reviewed publications. It is important to identify publication methods that are appropriate for the discipline or sub-discipline concerned.

# 1. INTRODUCTION

## 1.1 Background

Scientists in the design and engineering disciplines have long found assessment of the quality of their research output problematical, whether that assessment takes place within the context of an external evaluation, an academic appointment or promotion, or an application for funding. The quality indicators used in such situations are borrowed from the more basic sciences (publications in ISI journals, impact factors, citations, the Hirsch Index) and are, in the eyes of these scientists, inadequate. First of all, they argue, their output consists not only of peer-reviewed international publications, but also of actual designs and works of engineering. Secondly, they point to the fact that their research is generally more context-specific and multidisciplinary than that carried out in the more basic sciences. Design and engineering journals therefore have a lower impact factor and are consistently given a lower rating. As a result, experienced researchers in the design and engineering disciplines meet with little recognition in the scientific community, and often have trouble arranging external research funding or external evaluations, or gaining academic appointments.

## 1.2 Committee members and committee's task

In view of the foregoing problem, 3TU Federatie – an alliance between the three Dutch universities of technology – asked the Royal Netherlands Academy of Arts and Sciences to advise on the criteria used for ex-ante and ex-post assessments of research output in the technical disciplines of design and engineering. The Academy's Council for Technical Sciences, Mathematical Sciences and Informatics, Physics and Astronomy and Chemistry (TWINS Council) assembled a committee consisting of the following members:

- Prof. A. van den Berg, University of Twente
- Prof. R. de Borst, Eindhoven University of Technology
- Prof. P.P.M. Hekkert, Delft University of Technology
- Prof. A.W.M. Meijers (chairperson), Eindhoven University of Technology
- Prof. R.A. van Santen, Eindhoven University of Technology

Dr E.E.W. Bruins (STW Technology Foundation) served as an observer. A. Korbijn (Royal Academy) acted as the committee's secretary, and Dr J.B. Spaapen (Royal Academy) as an adviser. C.S. Tan (Royal Academy) assisted the committee during its interviews and international survey.

**Assignment**

The committee's assignment was to draft an advisory report on the criteria to be used for the ex-ante and ex-post assessment of:

- design and engineering activities in technical disciplines that can be considered scientific in nature;
- scientific research in the design and engineering disciplines.

The criteria had to satisfy the following requirements:

- they had to be useful to organisations that fund research activities (NWO, etc.);
- they had to be useful to universities in assessing academic staff (researchers/designers);
- they had to be credible within an international context.

In the present advisory report, the committee proposes a set of pertinent criteria and reports on its findings.

## 1.3 Procedure

The committee began by interviewing 27 representatives of various technical disciplines. The purpose was to investigate problems that the interviewees had encountered with the present system of assessment and to explore which aspects the stakeholders felt should be included in any new set of assessment criteria. A list of the problems they noted is given in Section 2. The interviewees and their specialisations are listed in Appendix 1.

To explore how this problem is tackled abroad and what lessons the Netherlands can learn from other countries, the committee also collected relevant information on the United Kingdom, Australia, Finland, the United States, Norway, Sweden, France, Germany and Austria. The results are summarised in Section 4; more detailed information is presented in Appendix 3.

Based on the information it had gathered, the committee analysed the problems that arise in assessing quality in the design and engineering disciplines and formulated a proposal for suitable assessment criteria. The main findings and an initial version of these criteria were passed on to the interviewees with a request for their written comments. The final proposal for the criteria is described and explained in Section 3.

## 1.4 Relationship with other initiatives

### 1.4.1 Evaluating Research in Context (ERiC)

The three universities of technology are taking part in three pilots being carried out within the context of the ERiC project (*Evaluating Research in Context*). ERiC is a follow-up to a previous project run by the Consultative Committee of Sector Councils for Research and Development (COS), which focused on how to measure the value of research for society. Part of that project involved developing a measuring method, referred to as the sci_Quest method. In 2006, a decision was taken to continue working on this method so as to explore and improve its usefulness in practice. The ERiC platform was set up for this purpose; its members are the Netherlands Association of Universities of Applied Sciences (HBO-raad), the Royal Netherlands Academy of Arts and Sciences (KNAW), the Netherlands Association for Scientific Research (NWO), the Association of Universities in the Netherlands (VSNU), and the Science System Assessment department at the Rathenau Institute. The Dutch Ministry of Education, Culture and Science acted as an observer. The project involved implementing three pilots within the Faculty of Electrical Engineering at Eindhoven University of Technology, the Faculty of Architecture at Delft University of Technology, and the Faculty of Mechanical Engineering at The University of Twente. The purpose of the pilots was to develop a method for assessing the relevance to society of the research conducted by these faculties. The indicators that emerged from the Architecture and Electrical Engineering pilots are given in Appendix 2. The authors took the results of the ERiC pilots into account when composing the present advisory report.

The present report is both narrower and broader in scope than the ERiC project. To begin with, this report focuses on a smaller number of disciplines, i.e. the design and engineering disciplines. Secondly, it not only considers how to assess the societal relevance of the output concerned, but also how to assess scientific the quality of that output.

### 1.4.2 Standard Evaluation Protocol

The Academy, the NWO and the VSNU have adopted the *Standard Evaluation Protocol* 2009-2015 (SEP) for the evaluation of scientific research [VSNU, KNAW and NWO 2009]. According to SEP 2009-2015, an assessment consists of an external evaluation conducted once every six years and involving a self-evaluation report and a site visit, and an internal midterm review midway between two external reviews. The SEP uses

four assessment criteria and two levels of assessment: the institute (or faculty or research school) as a whole and the underlying groups or programmes. The four assessment criteria are:
1. Quality
2. Productivity
3. Societal relevance
4. Vitality & Feasibility

The purpose of assessing scientific research by applying the SEP is to account for the institute's past performance and to improve its quality in future. One difference between the present advisory report and the SEP is that SEP only concerns entire institutes or the underlying groups or programmes. The present advisory report, on the other hand, also looks at the assessment of research proposals and persons. In view of the SEP's importance in the Netherlands, the authors of the present advisory report have taken pains to identify criteria that complement those of the SEP.

### 1.4.3 Assessment criteria in the humanities

The present method of quality assessment is also problematical in the humanities. For example, citation indices only count articles, and not books, and of those articles only the English-language publications. Contributions to public debate are also not taken into account. At the request of the Cohen Committee on the National Plan for the Future of the Humanities, the Royal Academy will join with researchers in the field and users of research data in the humanities (such as the NWO) to design a simple and effective system of quality indicators. At the time of writing, the advisory report on the humanities had not yet been completed.

### 1.4.4 Disciplinary boundaries

The committee's assignment was to recommend criteria for assessing the quality of (1) design and engineering activities that can be categorised as scientific and (2) research in the design and engineering disciplines. This two-pronged assignment came about because the present system also appears to be inadequate for assessing design and engineering activities (for example instrumentation design) within more 'science-like' disciplines. In addition, some of the research conducted within the design and engineering disciplines is multidisciplinary and unique in nature.

The text box below explains what is meant by design and engineering activities. The committee defines the design and engineering disciplines as those disciplines in which design and engineering activities play an important role. Examples are industrial design, architecture, information science, mechanical engineering, chemical engineering, biotechnology, civil engineering, marine engineering, electrical engineering and aerospace engineering. The committee believes that the criteria and indicators proposed in Section 3 are in any event appropriate for these disciplines.

**TECHNOLOGICAL DESIGN**

Design does not involve acquiring a better understanding of reality (true or reliable statements about that reality); rather, it is about creating something new. Designers add something to reality that did not exist before, and that is valuable or functional for users.

Design is a generic activity that takes place in many different domains, from art (a new painting or sculpture), to law (a new piece of legislation) to industry (a new chemicals plant). The design and engineering disciplines are concerned with designing new technological artefacts. These may be products or processes that can take physical form (buildings, telephones, power plants, molecules, chips) or abstract algorithms (software). It is also possible to design services (Internet transactions), living organisms (genetically modified crops), or environments that shape human perception (virtual reality).

There are many different ways to describe design processes. Some descriptions emphasise rational decision-making based on scientific models, with the range of solutions being narrowed at each pass; others try to include the non-rational, creative or art-like aspects of the design process. The American Accreditation Board for Engineering and Technology describes design as follows: '*Engineering design is the process of devising a system, component, or process to meet desired needs. It is a decision-making process (often iterative), in which the basic sciences, mathematics, and the engineering sciences are applied to convert resources optimally to meet these stated needs*' (Accreditation criteria for engineering curricula 2010-2011, www.abet.org).

The options to choose from in this process are not predetermined; they are developed within the process itself.

A design process results in a product: the design itself. This is a representation of the ultimate artefact in a form (prototype, computer model, scale model, drawing) that can be used to conduct all sorts of tests (performance, user aspects, producibility, costs, etc). Many engineering disciplines therefore view a design as a platform for research. Some describe an engineering design as analogous to an experiment in the natural sciences [Hee, van and Van Overveld, 2010].

# 2. PROBLEMS ASSESSING QUALITY

## 2.1 Introduction

Research comes in many different forms. We can attribute such variety to the various research traditions and approaches taken in the humanities, science and the social sciences. Each of these disciplinary clusters in its turn conducts many different types of research. The differences are often described in terms of the research aim, either a quest for fundamental understanding or a bid to solve practical problems related to use. This is why we divide research into basic research and applied research. Discussion in the literature shows that these are not mutually exclusive categories. Table 2.1 shows a customary classification of scientific research.

*Table 2.1 Classification of scientific research, borrowed from [Stokes, 1997, p. 73].*

| | | Considerations of use? | |
|---|---|---|---|
| | | No | Yes |
| Quest for fundamental understanding? | Yes | Pure basic research | Use-inspired basic research |
| | No | | Pure applied research |

Dutch universities of technology undertake a broader and more complex spectrum of research activities because design and engineering are part of their academic portfolio. They are essential activities for these universities; it is, after all, part of their mission to help solve societal problems or generate economic opportunities, and design and engineering in particular serve to build the necessary bridge to the real world.

With scientific research taking on so many different forms, as we saw above, we may well question whether a single quality yardstick can be used to assess them all. Can the same criteria that we use to assess basic research (in the natural sciences) also serve to evaluate the quality of applied research, or of design and engineering activities? Practitioners in the design and engineering disciplines usually answer this question with a resounding no, and it indeed touches on the heart of the controversy concerning the assessment of quality in these disciplines.

The discussion has been brewing for many years. In 2000, the Consultative Body for the Engineering Disciplines (DCT) published a report in which it recommended criteria for assessing engineering design [*Criteria voor de Beoordeling van het Ontwerpen in de Construerende Disciplines*] [DCT, 2000]. Another advisory report on quality assessment in the design and engineering disciplines was published in 2009 [*Kwaliteitsbeoordeling Ontwerp- en Constructie Disciplines*] [Schouten, 2009]. Both reports emphasise that engineering design should be assessed as a research activity. The present advisory report focuses not only on design and engineering activities but also on the research related to those activities.

## Growing urgency

The importance of a good quality assessment framework in the design and engineering disciplines has grown more pressing over time. To begin with, it is increasingly the case that research projects and academic staff on temporary appointment are financed solely by external sources. Design and engineering compete with other disciplines in that respect, forcing them to battle disciplines in the natural sciences that are a better fit with the existing quality criteria. Secondly, quality assurance has grown increasingly important at universities (external evaluations, internal midterm reviews), with administrators attaching consequences to the outcome. In that case, the criteria used to assess research quality are crucial. This problem is not exclusive to the design and engineering disciplines. Applied multidisciplinary research generally gets lower marks on traditional quality criteria. The disciplines concerned are often more restricted fields in which multidisciplinary journals with lower impact factors play an important role, and in which citation patterns are usually asymmetrical: frequent references in multidisciplinary journals to articles in monodisciplinary journals, but much less so the other way around.

In the following section, we describe the problems that leading scientists in the field have encountered in their work, based on their interviews with the committee (see Appendix 1).

## 2.2 Problems encountered

### Difficult to compare different fields on quality

Most of the interviewees who work in the technical sciences find that it is difficult to assess design and engineering using traditional methods based on publications in ISI journals with high impact factors. The extent to which the interviewees themselves were troubled by this – for example when applying for research funding – depends on the field in which they work. Almost none of the interviewees, on the other hand, had encountered problems when the quality assessment was conducted by peers working in the same field. Those working in the discipline are well aware of who the top scientists are. The problems arise when disciplines are compared with one another.

### There is no single, clearly designated funding channel for the design and engineering disciplines

One of the situations in which the quality of research output in different disciplines is compared is in the competition for research funding. This can be problematical when disciplines differ too much from one another. Researchers in the design and engineering sciences encounter problems because more basic research is generally accorded a higher status. Many of the interviewees see the lack of a clearly designated funding channel for the design and engineering disciplines as a shortcoming. Researchers who say that they do not find the current system of quality assessment problematical often obtain funding from business and industry or from more specific funding programmes such as the Point-One or Innovation-Oriented Research Programmes (IOP).

With EUR 100 million having recently been transferred from the direct to the indirect funding mechanism, the lack of a single funding channel at the NWO for the design and engineering disciplines has become even more urgent. At the moment, say a number of the interviewees, these disciplines simply do not have a fair shot at obtaining funding from the NWO.

An additional complication for design and engineering is that in a number of specific areas – for example civil and hydraulic engineering – the intermediary research organisations have been abandoned, have become less knowledge-intensive, or no longer fund research. This reduces the number of opportunities for contract research; at the same time, indirect funding programmes are offering a diminishing number of opportunities that require the involvement of stakeholders in civil society. That too makes it more urgent to have a single funding channel within the NWO.

### Little acknowledgment that publication cultures differ considerably

None of the interviewees dispute the importance of publishing for knowledge dissemination and for the evaluation of output by means of peer review. The emphasis on peer-reviewed articles in ISI journals, however, pays little regard to the fact that disciplines differ considerably when it comes to their publication cultures. In information science, for example, some peer-reviewed conference proceedings are more prestigious than publication in an ISI journal. The number of publications that a researcher can produce each year also varies considerably from one discipline to the next. In engineering, researchers need to spend a considerable amount of time on designing and engineering before they can even consider publishing their results. Various interviewees reported increasing pressure to publish in high-impact journals such as *Science* and *Nature*, even though fellow researchers in their discipline do not consider these journals important.

Interviewees who work in such disciplines as industrial design and architecture indicate that for them, publications are becoming more important as part of a strategy of 'scientification'. It is important, however, to bear in mind that the form and medium in which articles are published must be appropriate for the relevant discipline.

## Declining role of design and engineering in technical disciplines

Design and engineering are the foundations of the engineering sciences and represent a bridge to actual practice. A design or work of engineering integrates knowledge from various disciplines in order to create the functions of a system that will operate in a real-life situation. Many of interviewees observed that the 'scientification' of some technical sciences (for example civil engineering) has led to a 'mono-science-like culture' in which the design and engineering of technical systems receive less attention than is actually necessary, and that there is often a huge disparity with real-life situations. That is an undesirable situation.

## Growing gap between university administrators and practitioners in the field

Many of those interviewed feel that the quality assessment of research output at universities is becoming a bureaucratic tool. The growing gap between practitioners in the field and university administrators makes it difficult for the latter to have a clear view of the quality of research in a particular discipline. That is why there is a need for 'objective' and, preferably, quantitative information about the quality of different research units. Comparisons between such units are then based on bibliometric analyses or other numerical indicators. Design and engineering are at a disadvantage in that respect, compared to other disciplines that are a better fit with the existing quality criteria. That disadvantage affects not only the way the disciplines are funded, but also their scientific reputation and the recognition they are given.

## 2.3 Evaluation of perceived problems and conclusions

### Differences between disciplines

The committee agrees that it is pointless to compare the research quality of different disciplines. Such a comparison assumes that there are quality indicators that apply or that should apply equally for the relevant disciplines. That is a false assumption, however. The research traditions, publication habits, and activity and output categories (journal articles, proceedings, books, artefacts, designs) are too diverse. Value judgements – for example about the comparative quality of researchers working in various disciplines based the number of ISI journal articles and impact factors – are therefore not meaningful. Neither does the popular Hirsch Index offer a basis for comparison. At most, it is possible to compare data that has been standardised for a particular discipline. We can then say that scientist A falls into the top 5 percent in discipline X, or that scientist B belongs to the top 25 percent in discipline Y.

Quality benchmarks are therefore only meaningful when they are made *within* disciplines. As appealingly simple as administrators may find the 'one size fits all' approach, it does not do justice to significant differences between disciplines and will therefore always give certain disciplines an unfair advantage. This opinion corresponds with previous conclusions by the Academy's KNAW Quality Assurance Committee [KNAW-commissie kwaliteitszorg, 2008].

### Peer review

Within disciplines, the assessment of research quality is usually by means of peer review. No matter how scrupulous and differentiating such reviews are, however, they can never be entirely objective or do justice to all the differences and controversies within a discipline. Personal bias cannot be avoided, and subjective elements will always play a role. Different peers will therefore reach different conclusions, at least to some extent. There is also the danger of a small number of peers dominating a discipline, and of previous assessments influencing later ones [KNAW, 2008]. Despite these limitations, the committee believes that the peer review system is in fact the best quality assessment method. The discipline must, however, be sufficiently homogenous in nature and international enough in scope to make independent peer review possible. If these conditions are satisfied, then a peer review system should also be relatively trouble-free in the design and engineering sciences.

### Bibliometric research

Along with the growing gap between administrators and practitioners, the limitations of peer review have prompted the trend towards bibliometric research. It is an illusion to think that an entirely quantitative/bibliometric approach to research quality is possible, however, or that it can replace the peer review method, not in the least because ultimately, peers are needed to assess and interpret the bibliometric data. At most, the committee believes, bibliometric research can supplement peer review. It is important to realise that both methods have their limitations. Quality assessments should always be conducted with a great deal of care for that reason.

### Random indication

An additional reason to proceed with caution is that quality assessments are always mere random indications; as time passes, people and their work can vary in quality. That is something that assessment committees must be aware of.

### Publications are also important in design and engineering

The committee attaches great value to the publication of research results in peer-reviewed journals, even in such disciplines  as industrial design and architecture. That is because, first of all, assessment by peers precedes publication of such results. Secondly, publications make knowledge available to others and ensure that it is disseminated. The publication method must be appropriate for the discipline concerned, however. A unilateral focus on publications in monodisciplinary journals that have a high impact factor is undesirable, especially for a university of technology where multidisciplinary and application-driven research is also important. The committee has been pleased to note the growing importance that scientists in such disciplines as industrial design and architecture now attach to peer-reviewed publications, as part of their research output.

### Criteria steer the direction of research

The criteria used to assess the quality of research output also influence decision-making on the type of activities universities choose to develop. They cast a long shadow because researchers tend to anticipate the criteria that will be used to evaluate their output and to determine whether they will receive funding, and how much. Quality criteria are therefore far from neutral; rather, they set the standard in a discipline. Besides their role in the ex-post evaluation of research output, they also have an ex-ante influence in steering the direction of research. In that respect, the committee shares the concern of a number of interviewees that a unilateral focus on criteria appropriate for the natural sciences may have led in some technical disciplines to undue pressure being placed on design and engineering and the related multidisciplinary research. In view of the mission of a university of technology, it is very important to strike the right balance between the different categories of research described previously (Section 2.1, Table 2.1) and design and engineering activities. That will only happen, however, if the quality criteria used are tailored to that purpose.

## Research funding bodies

Research funding bodies such as the NWO and the STW must be aware that the 'one size fits all' approach does not do justice to the differences between disciplines and will therefore always give certain disciplines an unfair advantage. A number of the interviewees believe that the NWO does not offer the design and engineering disciplines equal opportunities. More generally, they find that multidisciplinary or interdisciplinary research does not fit effortlessly into the NWO's structure, and that it must compete with disciplinary research that more readily meets the quality criteria set. There are also vested interests that new disciplines are forced to compete with. The committee believes that the NWO has acknowledged the problem by initiating cross-disciplinary programmes. This initiative merits further development, for example by launching programmes that bring together NWO and STW disciplines. With respect to the STW, the various interviewees indicated that the yardstick for assessing the societal relevance of research is too severely limited to the short-term interests of the businesses on the user committees. According to the committee, it would be a good idea to take a more differentiated approach in this case as well and to also support research projects of long-term relevance, even if no businesses can be found for that purpose as yet. That is also important for the design and engineering disciplines.

## Importance of a transparent objective

Finally, the committee considers that the quality criteria should be fine-tuned to suit the objective of the relevant quality assessment. It is one thing to assess a research proposal and quite another to review the past performance of a research group or to evaluate the performance of a member of staff awaiting promotion or a permanent appointment.

### CONCLUSION 2.1

Quality comparisons are only meaningful when they are made within disciplines. As appealingly simple as administrators may find the 'one size fits all' approach to assessing quality, it does not do justice to significant differences between disciplines and will therefore always give certain ones an unfair advantage.

### CONCLUSION 2.2

Those working in the design and engineering disciplines encounter relatively few problems when the peer review system is used to assess the quality of research output or to identify outstanding scientists.

### CONCLUSION 2.3

It is important to universities of technology that the criteria for research quality should be fine-tuned to reflect the different types of research and activities that take place there. A unilateral focus on criteria derived from the natural sciences makes it difficult to strike the right balance between basic research, applied research, and design and engineering activities.

### CONCLUSION 2.4

The publication of research results in peer-reviewed journals, books and proceedings is of huge importance in the design and engineering disciplines. The publication method must be appropriate for the discipline concerned, however.

# 3. ASSESSMENT CRITERIA

## 3.1 Introduction

This section identifies criteria and indicators that the committee believes are suitable for assessing quality in the design and engineering sciences.

***Comment concerning terminology***
The term ‘quality’ is an evaluative concept indicating how good something is. It can be applied to objects, processes and persons. Different situations call for different quality assessment criteria. For example, we assess the quality of an article for publication differently than the quality of the computer on which we write it. Quality assessments are relatively meaningless if we do not indicate the criteria on which we have based our evaluation. We usually cannot observe directly how well something meets a certain assessment criterion. That is why we work with indicators, which provide empirical information on the extent to which an object, process or person satisfies a particular criterion. That does not necessarily have to be quantitative information (for example the number of articles someone has published); it can also refer to qualitative information (for example what peers think of the journal in which someone has published an article).

### Basic principles for assessment

***Take disciplinary differences into account***
As discussed previously, disciplines differ significantly from one another. Even within the relatively limited field of the design and engineering sciences, however, there is considerable variety. There is broad agreement that the present system of quality assessment does not take this variety sufficiently into account. In the committee’s opinion, however, that is something that an assessment system must certainly do.

### *Differentiation required...*

In view of the differences between disciplines and between assessment situations, the committee believes that differentiation is required in assessments of research quality. This means:

- differentiation between disciplines/sub-disciplines;
- differentiation between types of activity within a particular discipline (research or designing and engineering);
- differentiation between the objects of assessment (a member of staff, a research proposal, or a research group).

### *...but within an overall framework*

On the other hand, the committee also believes that the assessment standards for a particular discipline/sub-discipline should not differ completely from those used in other disciplines. That would make them arbitrary and opportunistic, and perhaps also create the impression that the design and engineering disciplines are subject to 'milder' requirements than other disciplines. There should be an overall assessment framework that allows for the variety described above. The committee has therefore attempted to identify overall quality assessment criteria that also apply in a broader sense, along with indicators that may differ or vary in importance from one discipline to the next.

### *Peer review crucial*

As explained in the previous section, the committee regards peer review as crucial to assessing the quality of publications, researchers, research groups and research proposals. The discipline must, however, be sufficiently homogenous in nature, broad (and international) enough in scope, and have a sufficient number of suitable peers available.

### *Bibliometric data should supplement, not replace, peer review*

The committee believes that bibliometric data should be used to supplement peer review, and not replace it, as discussed earlier in this report. When determining the value of bibliometric indicators (citation scores, impact factors, the Hirsch Index, etc.), it is by no means a trivial matter to define precisely what is regarded as a discipline – indeed, this is a crucial point.

### *Transparent objectives*

Quality assessments, for example external evaluations, are often used as a ranking and PR tool, and not to identify potential improvements. The purpose of the quality assessment should be made clear before the assessment is conducted, and the assessment itself should be viewed in the light of that specific purpose.

## 3.2 Assessment criteria

### *Five criteria narrowed down to two*

In the first instance, the committee considered five different assessment criteria, in line with a proposal put forward by the Royal Academy of Engineering (United Kingdom) for evaluating quality in the technical sciences. These criteria were: 1) publications, 2) impact, 3) innovativeness, 4) involvement of external stakeholders, and 5) reputation of scientists concerned. There was considerable support for these criteria during the consultations between the committee and practitioners in the Netherlands. When considered in greater detail, however, some of these criteria actually turned out to be underlying indicators for other criteria. For example, publication in leading journals is not an independent criterion but an indicator for the criterion 'research quality'. The involvement of external stakeholders in research is also not a criterion in itself, but rather an indicator for the criterion 'societal relevance'. The committee gave considerable thought to whether innovativeness should be categorised as an independent criterion. Ultimately, it decided that innovativeness combines two aspects, scientific originality and applicability for society, and is therefore not a criterion in and of itself. The first aspect is expressed in scientific publications and the second can be shown by indicators measuring societal relevance. Finally, the committee believes that 'reputation' is a derived variable encompassing both scientific and societal aspects.

In the end, the committee concluded that essentially, there are only two criteria suitable for assessing the quality of activities carried out in the design and engineering sciences:

1. research quality
2. societal relevance

This conclusion is remarkable in that the proposed criteria do not differ from criteria currently applied in other areas of science and scholarship. Contrary to the committee's original assumption, then, it will not propose a separate set of criteria for assessing quality in the design and engineering sciences. The foregoing two criteria have broad support in the scientific community and are similar to what is customary at international level (see Section 4). They appear to offer an overall framework for assessing quality in every scientific discipline.

Within this overall framework, however, the committee would wish to see differentiation with respect to the indicators used to determine how well something satisfies the two assessment criteria. Specifically, that means:

1. differentiation between disciplines: the nature and relative importance of the indicators used will differ from one discipline to the next; in the design and engineering sciences, for example, there are other indicators of research quality than numbers and citations in peer-reviewed journals.

2. differentiation between types of activity within a particular discipline: research, designing, or engineering.
3. differentiation between the objects of assessment: output or person. The indicators used to assess a funding application for a research project often differ from those used to assess a candidate for an appointment or promotion.

Table 3.1 shows the two assessment criteria and related indicators that the committee believes are suitable for assessing quality in the design and engineering sciences. As mentioned earlier, not all indicators are equally relevant or have the same relative significance for each discipline/sub-discipline, activity or assessment situation. The table attempts to show the most important types of indicators. Indicators of the quality of education and teaching are not included, and are beyond the remit of this advisory report.

## 3.3 Explanation of indicators

In this section, we provide a more detailed explanation of the indicators in the table where necessary.

### 3.3.1 Research quality

***Scientific publications***
As mentioned earlier, scientific publications are also an important indicator of research quality in the design and engineering sciences. The publication method may differ from one discipline to the next, however (e.g. journals, proceedings, electronic publications).

***Designed artefacts***
The committee believes that design artefacts (products, processes, software) should also be considered in quality assessments of research output in the design and engineering sciences. It wishes to make a few comments in this connection, however. The artefacts it has in mind are those that make a scientific contribution to the discipline, and not artefacts designed within an educational context. One of the essential characteristics of a scientific design is that it generates new knowledge that can be applied generically. The documentation that accompanies the artefact should be enlightening in that respect. If a design generates knowledge, then that knowledge can usually also be published in peer-reviewed journals, a guarantee of quality assurance. The fact that this knowledge *can* be published does not necessarily mean that it has been published. There may be many good reasons for not doing so, for example because publishing is not customary, or may even be prohibited, in the business environment in which the knowledge was generated.
As an indicator for research quality, artefacts can play a role in the appointment of professors who have worked in industry. Because such candidates are important to

*Table 3.1 Quality indicators for the Technical Sciences*

<table>
<tr><th></th><th>INDICATORS FOR OUTPUT</th><th>INDICATORS FOR PERSON</th></tr>
<tr><td>SCIENTIFIC QUALITY</td><td>Scientific publications<br>Articles in peer-reviewed journals (no. and type of journal)<br>Articles in peer-reviewed conference proceedings (no. and type of proceedings)<br>Scientific books published by leading publishers or significant contributions to such books (no. and type)<br>Citations of individual articles<br>Impact factors of journals in which articles are published<br><br>Designed artefacts<br>Peer-reviewed artefact (design) + documentation. This also includes software design<br><br>Research impact (ex-post)<br>Use of scientific products by other researchers (artefacts, methods, measuring instruments, tools, standards and protocols)<br><br>Potential research impact<br>Possible contribution to development of theories and models, methods, operational principles or design concepts</td><td>Recognition by scientific community<br>Membership of prominent organisations such as academies of sciences<br>Prestigious grants (VENI, VIDI, VICI, or ERC Grants)<br>Honorary doctorates<br>Visiting professorships<br><br>Editorships<br>Chief/full editorship of international scientific journal/book/conference proceedings<br><br>Considered expert by peers<br>Advisory capacity in scientific circles (NWO, external inspections, etc.)<br>Keynote lectures at science conferences<br>Membership of programme committees<br>Participation in international assessment committees for scientific programmes/institutes or scientific advisory councils/institutes<br><br>Research impact across the course of career<br>Person’s citation score<br>Contribution to developing a ‘school of thought’</td></tr>
<tr><td>SOCIETAL RELEVANCE</td><td>Use of results by external stakeholders (ex-post impact)<br>Contribution to solving societal problems<br>Market introductions and new projects in industry<br>Income generated by use of results<br>Spin-offs with industry<br>Patents used<br>Artefacts used (designs, software)<br><br>Use of results by profession (ex-post impact)<br>Use of artefacts, methods, measuring instruments, tools, standards and protocols<br><br>Involvement of external stakeholders in scientific output (potential societal relevance)<br>Businesses or civil-society organisations involved in guiding research projects (e.g. in user committees)<br>Contract financing by potential users (e.g. industry)<br>Public financing related to societal questions<br>Valorisation grants<br><br>Contribution to knowledge dissemination<br>Professional publications and papers, non-scientific publications, exhibitions and other events related to research results</td><td>Considered expert by external stakeholders<br>Advisory and consultancy work (focused on users)<br>Leading position in industrial research (e.g. managing director of R&D department)<br><br>Considered expert by profession<br>Oeuvre prizes (e.g. architects)<br>Retrospective exhibitions<br><br>Contribution to knowledge dissemination<br>Activities focusing on popularisation of science, education and contribution to public debate<br>Training of professionals<br>PhDs with their first job in relevant practice</td></tr>
</table>

universities of technology, such scientific achievements must also be considered. It is essential, however, to ask peers to assess the scientific merits of a small number of well-documented design artefacts.

### *Research impact*

Research impact means the extent to which research output is used by other scientists. One of the most common methods of use is the citation. Another indicator is how much new research products are used by fellow researchers. By research products, we are referring here to artefacts, methods, measuring instruments, tools, standards, protocols, etc. This criterion does not include the use of scientific results by non-scientists; that type of use falls under societal relevance.

### *Potential research impact*

This indicator plays a role in assessing the quality of research proposals. They are assessed in part for their potential contribution to the discipline.

### *Recognition by the scientific community*

Expressions of recognition are an important gauge of the research quality of an individual. Such recognition must be gained within a peer-review process, however. Examples include the award of science prizes (Spinoza) and other distinctions, honorary doctorates, membership of prominent organisations such as academies of sciences, and prestigious grants and other funding (VICI, ERC Grants).

### *Extent to which peers regard an individual as an expert*

The extent to which a researcher is regarded by fellow scientists as an expert is an important indicator for quality because practitioners generally know precisely who are the leading scientists in their field. Such scientists are invited to give keynote lectures at science conferences, become members of programme committees, and to act in an advisory capacity in scientific circles.

### *Research impact across the course of a career*

This means the research impact of a person throughout his or her career. That impact may become evident from that individual's citation score. It can also be seen in how much peers believe that person has helped develop a particular school of thought.

## 3.3.2 Societal relevance

### *Use of results*

This covers the indicators that measure the use of scientific knowledge or products beyond the field of science. It is usually only possible to assess actual use after the research has been completed. Such use can take many different forms: it may involve making a contribution to solving a societal problem, or it may mean commercial use by a business. The use of results by the relevant profession constitutes a separate category.

***Potential impact***
The societal relevance of research can be assessed on a number of different timescales. The time horizon for applied research is shorter than that for basic research. It is notoriously difficult, however, to identify the societal relevance of long-term research. In such cases, it is possible to work with indicators that express the degree of interest that stakeholders in civil society have in the research. The stakeholders consist not only of businesses, but also of government ministries or international non-profit organisations. In the case of short-term research, co-financing can be taken as an indicator of stakeholder involvement. That involvement also considerably increases the chance that the results will in fact be used.

***Contribution to knowledge dissemination***
Whether assessing research output or evaluating an individual, the contribution made to knowledge dissemination must also be included as an indicator of societal relevance. Knowledge dissemination helps solve societal problems or create economic opportunities, after all. It is also part of the mission of a university of technology.

## 3.4 Relative importance of the criteria and the indicators

Assessment is a question of fine-tuning. Before an assessment is carried out, a decision must be taken as to which of the two criteria and which of the indicators are most important in that given situation. The criteria and indicators relevant for a full-time professor who heads a research group will differ from those used for a part-time professor whose job involves bridging the gap between academia and real-life situations. The criteria and indicators for an individual VIDI grant will differ from those for a funding application submitted in an open NWO competition, or from an STW funding application. And we have not even mentioned differences between disciplines. In short, there is no one method for conducting a quality assessment.

This advisory report proposes a systematic framework for assessing quality in the design and engineering sciences, one that takes disciplinary differences into account and that has been developed to serve in a variety of assessment situations.

## 3.5 Profiles

The assessment of a person or research group according to the above table must not result in a single score – that would, after all, be denying the existence of different quality criteria and indicators. The result should be a profile, which can be shown in diagram form. The most obvious diagram would be one that offers an overall score for each of the quadrants of the table (instead of a score for each of the indicators in that quadrant). One example would be to use a scale running from low to high (numbers would suggest a false accuracy in this case). Some other form of representation may also be appropriate, however.

The use of profiles raises the issue of which profiles are desirable, and at what level of aggregation (persons, groups). The profiles can vary per discipline and per research group. There can also be differences within a group. A part-time professor working in industry will have a different profile than a full-time professor who heads a faculty. The purpose, however, is to see that the group as a whole is assigned the desired profile. Figure 3.1 gives an example of two profiles, one that is explicitly scientific in nature and one that is mixed.

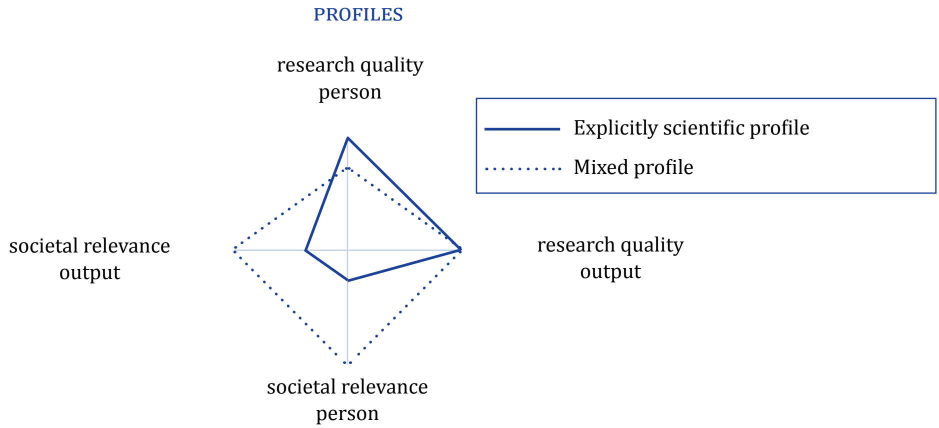


*Figure 3.1 Example of an explicitly scientific profile and a mixed profile.*

## 3.6 Selecting peers

One significant issue in the peer review system is how to select the most suitable peers. The problem arises in connection with external evaluations, but also when assessing research proposals or appointing academic staff. In small countries and small disciplines where everyone knows everyone else, there is the risk that peer review will result in an 'old-boy network'. It would be better, then, to use a more objective method for determining which peers are suitable in a given situation. The table of criteria and indicators can be useful in that respect. Specifically, peers can be defined as fellow scientists who get high marks on indicators that typify a particular discipline. Candidates for assessment committees can then be judged on the basis of that score. In addition, peer autonomy can be encouraged by including international peers on committees.

## 3.7 Conclusions

CONCLUSION 3.1

The quality of research output in the design and engineering disciplines can be assessed on the basis of two criteria: 1) research quality and 2) societal relevance. It is not necessary to identify a separate set of criteria for these disciplines. Each of these two criteria is related to a set of indicators showing how well a person or output scores on the relevant criterion.

CONCLUSION 3.2

Assessing quality is a question of fine-tuning owing to the differences between disciplines, categories of scientific activity (designing, research) and assessment situations (external evaluations, appointments, research proposals). The present advisory report offers a systematic assessment framework that will do justice to these differences by attributing a certain significance or weight to indicators of research quality and societal relevance (Table 3.1).

# 4. SURVEY OF LESSONS LEARNED ABROAD

## 4.1 Introduction

Scientific research is an international affair, and in that respect it is important that the criteria proposed in the previous section are also credible from an international perspective. The committee additionally wanted to make use of lessons learned abroad. It therefore looked at how a number of other countries tackle the quality assessment of scientific research, what criteria they use, and what they have learned in the process. Where possible, the committee looked specifically at the design and engineering sciences. The United Kingdom (UK) and Australia have extensive national research evaluation systems. In both countries, the issue of suitable criteria generated considerable discussion, and there was a great deal of background research. The committee therefore investigated the situation in both these countries in detail. In Finland, the committee looked at how Aalto University assesses research quality; this university took a very thorough approach and its evaluation system devotes considerable attention to the design and engineering sciences. Finally, the committee looked briefly at how a number of countries relevant to the Netherlands deal with quality assessment. Information on the individual countries can be found in Appendix 3. Section 4.2 reviews the most important findings and conclusions.

## 4.2 Summary of lessons learned abroad and conclusions

Most foreign assessment systems cover the entire spectrum of science and scholarship and only the tackle specific problems associated with design and engineering indirectly. There are nevertheless a number of relevant observations that can be made within the context of the present advisory report.

### *Use of bibliometric data*

The assessment methods used in the UK, at Finland's Aalto University, and in Australia attach considerable value to bibliometric analyses. The issue of whether these analyses are in fact applicable has clearly generated considerable discussion in these countries, particularly with respect to the design and engineering sciences. Although the UK's Research Excellence Framework at first emphasised the use of bibliometric methods, that emphasis was ultimately toned down. The consultation rounds made clear that such indicators were not robust enough and had too little support from scientists in the field. They have not disappeared entirely from the assessment system, however; they now serve as background information for review committees. In the Australian system, on the other hand, bibliometrics still plays a prominent role, although such analyses are differentiated by discipline and only apply in specifically defined fields. In the design and engineering sciences, they apply in every discipline except the architectural disciplines. The engineering sciences are evaluated in a similar manner to physics and chemistry. The Finnish evaluation committee indicates that the Architecture, Design, Media and Art Research Panel could not actually make use of bibliometric analyses in its reviews owing to the small number of articles published in ISI journals.

### *Possibility of including non-traditional information*

The evaluations produced by both Aalto University and in accordance with Australia's ERA (Excellence in Research for Australia) method allow 'non-traditional output' to be included in the evaluation. This covers films, websites, exhibitions and creative 3D work (architecture, design, games, software). Here again, this exception appears to apply mainly to the disciplines of architecture and industrial design. The evaluations are carried out by peer review committees that assess case studies submitted by the units under review. The Australian system stipulates specifically that the submitting units must themselves identify the research component of the output. The experience gained at Aalto University in Finland shows that this takes some adjustment. The evaluation committee pointed out that while the university administrators had indeed made this possible, when it came right down to it the units were urged to submit mainly scientific articles.

### *Focus on societal impact from an international perspective*

In addition to the research quality of the output, all of the foreign evaluation systems discussed here also consider its societal impact. There is still considerable discussion of the way in which that is supposed to happen.

### *Reasonably comparable criteria used for assessment*

There is a reasonable level of consensus concerning the criteria to be used to assess the quality of research output. Although the precise description differs from one system to the next, the following criteria are virtually always used: research quality, societal relevance and (to a lesser extent) productivity.

### *Method used to assemble peer review committees not specified*

Peer review committees play a crucial role in the various evaluation processes. This means that the quality of the assessment process depends largely on the quality of these committees. Given the enormous variety of fields that some panels are obliged to consider, the question is whether the members of the peer review committees are in fact actually 'peers' to the extent required by particular research projects. What is notable is that most of the assessment procedures discussed here do not stipulate the way in which the peer review committees are assembled or how the members are selected.

#### CONCLUSION 4.1

The key criteria in the assessment systems of trendsetting countries such as the United Kingdom, Australia, the United States and Finland are research quality and societal relevance.

#### CONCLUSION 4.2

Various foreign assessment systems allow results such as creative output, websites, and other products to be included in the assessment of quality in the design and engineering sciences. That possibility is often limited to the more 'artistic' fields, however.

#### CONCLUSION 4.3

Virtually all of the assessment systems considered devoted considerable effort to bibliometric information. The significance of that information differed from one system to the next.

#### CONCLUSION 4.4

The assessment framework proposed in this advisory report is credible within an international context.

# 5. CONCLUSIONS AND RECOMMENDATIONS

Section 2 described problems in the present system of assessing the quality of research output in the design and engineering disciplines. Section 3 identified a new assessment framework that avoids these problems. Section 4 considered that framework within an international context. This final section discusses the scope of this advisory report and its implications for existing practice. It also identifies a number of follow-up steps. The section ends with a set of recommendations. First, however, we repeat the key conclusions below.

### CONCLUSION 2.1

Quality comparisons are only meaningful when they are made within disciplines. As appealingly simple as administrators may find the 'one size fits all' approach to assessing quality, it does not do justice to significant differences between disciplines and will therefore always give certain ones an unfair advantage.

### CONCLUSION 3.1

The quality of research output in the design and engineering disciplines can be assessed on the basis of two criteria: 1) research quality and 2) societal relevance. It is not necessary to identify a separate set of criteria for these disciplines. Each of these two criteria is related to a set of indicators showing how well a particular output, unit or person scores on the relevant criterion.

### CONCLUSION 3.2

Assessing quality is a question of fine-tuning owing to the differences between disciplines, categories of scientific activity (designing, research) and assessment situations (external evaluations, appointments, research proposal). The present advisory report offers a systematic assessment framework that will do justice to these differences by attributing a certain significance or weight to indicators of research quality and societal relevance (Table 3.1).

CONCLUSION 4.4

The assessment framework proposed in this advisory report is credible within an international context.

## Scope of this advisory report

In accordance with the request submitted to the Academy by the three Dutch universities of technology, the assessment framework is intended mainly for the design and engineering sciences. However, the committee anticipates that the framework will also be useful for other technical disciplines. That is first of all because the two criteria identified (research quality and societal relevance) correspond closely to the existing practice of quality assessment (SEP, assessment of research proposals, etc.). Secondly, it is because the proposed framework is flexible enough to do justice to relevant differences between disciplines. That flexibility is the result of the indicators used to determine how well research output satisfies the two quality criteria. These indicators need not be identical or even equally significant in every area of science. Discipline-specific indicators can even be added. The proposed framework is therefore not only useful for the design and engineering disciplines, but also, for example, for the more basic sciences (i.e. the natural sciences. After all, the quality indicators customarily used there (publication in high-impact journals, etc.) are included in the indicators in Table 3.1. The framework is therefore generic in nature, but to become operationally valuable, it must be made discipline-specific by selecting pertinent quality indicators and assigning them a relative importance. Scientists working in the more basic disciplines can, for example, indicate that the quality of their output can be assessed by considering the indicator 'publication in high-impact journals'.

## Compatibility with existing practice

There are various situations that require quality assessments to be made: the evaluation of a research proposal; a performance review of a research group; or the evaluation of a staff member being considered for a promotion or permanent appointment. The committee believes that the advice given here will be relatively easy to apply in the existing assessment practice. The SEP used for external evaluations is also compatible with the proposed assessment framework.

Rejection of the 'one size fits all' approach to assessing quality means that policy-making will probably come to play a more important role in decision-making situations involving comparisons between disciplines, for example in the NWO's competitions for research funding in heterogeneous fields or in the award of scientific prizes (Spinoza). Such decisions cannot be made and justified merely on the basis of a quality assessment. Comparing quality between different disciplines is like comparing apples and oranges (person $p_1$ or research proposal $r_1$ has a value score of $v_1$ on indicator set $i_1$ and person $p_2$ or research proposal $r_2$ has a value score of $v_2$ on indicator set $i_2$). At best, it is possible to compare scores that have been standardised for the relevant disciplines (person $p_1$ scored among the top 5 percent of scientists on indicator set $i_1$ in discipline $d_1$ whereas person $p_2$ scored among the top 25 percent of scientists on indicator set $i_2$ in discipline $d_2$). Even here, however, the meaning is not entirely obvious.

## Request for advice by 3TU

The committee applauds the fact that the Governing Bodies of the three Dutch universities of technology have raised the issue of the status of design and engineering activities and the related research. These activities are under pressure because the quality assessment framework is not sufficiently compatible with them. As a result, the balance between basic research, applied research, and design and engineering activities has been or is at risk of being upset. This is an undesirable situation, given the mission of the universities of technology. If the three Governing Bodies adopt the advice given here and communicate it to their academic staff, they will have made clear that outstanding design and engineering activities are important to universities of technology and a vital component of the technical sciences.

The first step in the process of implementing the advice is for the design and engineering sciences to determine which quality indicators described in Table 3.1 apply to them and their relevant importance. These indicators must be credible within the discipline in an international context. Only peers working in that discipline can determine which indicators should apply. Most obvious would be to utilise existing cycles of quality assurance for this purpose. The board of 3TU could, for example, ask a discipline to define the quality indicators prior to an external evaluation, so that the evaluation committee is able to use these indicators in its research assessment. A committee consisting of international peers could also be asked to evaluate these indicators.

RECOMMENDATION 1

The committee advises the board of 3TU to create sufficient scope for discipline-specific quality assessment in the technical sciences, and to ask the design and engineering disciplines to identify the indicators and their relative importance for assessing quality in those disciplines. These indicators must be credible in an international context.

## Research funding bodies

The committee believes that the method used for quality assessment should not in itself be advantageous or disadvantageous for particular disciplines. This should be a key factor in the assessment system utilised by the NWO/STW. The assessment framework set out in this advisory report can help determine quality indicators that take disciplinary differences into account.

By introducing cross-disciplinary programmes, the NWO is demonstrating that it is in any event aware of the importance of programmes that do not fit within existing frameworks and for which other quality indicators may apply. The committee believes that this approach should be extended, for example by starting up programmes that combine NWO and STW areas. That is very important for the design and engineering disciplines, which do not have a dedicated funding channel at the NWO. The committee advises the STW to take a more differentiated approach to assessing the societal relevance of research projects. Even those projects that do not immediately interest businesses because of their long-term nature should be eligible for funding. That would be of benefit to the design and engineering disciplines, given the fact that various intermediary research organisations in a number of areas have been shut down (see Section 2).

The committee applauds the fact that most research funding bodies make use of the peer review system. It is vital, however, to select the right peers, as this will have a major influence on the outcome of the assessment. Specifically, assessment committees have very little leeway to deviate from referee reports. It would be a good idea to make the peer selection criteria more objective and transparent. That would improve the quality of the peer review process. Table 3.1 may be useful in that context. Peers involved in assessing quality in the design and engineering sciences must also be assessed according to a broad range of indicators, and not only on their publication behaviour.

RECOMMENDATION 2

The committee advises the NWO and the STW to devote more attention to programmes for disciplines that do not fit easily into existing disciplinary categories and the present quality assessment method. This applies in particular for the design and engineering sciences. The quality indicators used must do justice to these disciplines. The committee advises the STW to support projects that will be relevant to society in the longer term and in which businesses have not yet expressed a tangible interest.

RECOMMENDATION 3

The committee advises the universities and the NWO/STW to make the peer selection process more transparent and objective. That will improve the quality of peer review process. The assessment framework and discipline-specific quality indicators described in this advisory report can play a role in this context. It is important to select the right peers when convening external evaluation and review committees and when selecting referees to evaluate research proposals.

### *Scientists in the design and engineering sciences*

A quality assessment system that is fine-tuned to the design and engineering sciences requires the scientists working in these disciplines to make an active contribution, for example in order to identify the quality indicators, decide on their relative significance, and operationalise them. They will also need to participate as peers in NWO and STW research proposal assessment procedures. The committee has noted that this now only happens on a limited scale.

The assessment framework described in this advisory report makes it possible to use other indicators of research quality than peer-reviewed publications. Nevertheless, the committee and many of the interviewees believe that peer-reviewed publications are also important for the design and engineering disciplines, and will remain so. Such publications help foster mutual quality checks and the dissemination of knowledge. Scientists working in these disciplines or in specific areas of these disciplines do not always publish as a matter of course, however. It is therefore important to look for methods of publication that are appropriate for the discipline or sub-discipline concerned. New media can play a role in that respect.

### RECOMMENDATION 4

The committee advises scientists in the design and engineering disciplines to do their utmost to promote a culture of peer-reviewed publications wherever necessary. Such publications can serve to verify results, disseminate knowledge and contribute to the 'scientification' of the discipline.

# BIBLIOGRAPHY


Aalto University (2009a). *Research Assessment Exercise 2009, Panel reports*

Aalto University (2009b). *Research Assessment Exercise 2009, Terms of reference for research assessment panels*

Aalto University (2009c). *Bibliometric report 2003-2007*

DCT Disciplineoverlegorgaan Construerende Techniek (2000). *Criteria voor de beoordeling van het ontwerpen in de construerende disciplines*

ERiC publicatie 1001 (2010). *Handreiking Evaluatie van maatschappelijke relevantie van wetenschappelijk onderzoek*

ERiC publicatie 1002 (2010). *Pilot Study at Faculty of Architecture TU Delft, Final Report (Confidential)*

ERiC publicatie 1003 (2010) *Evaluating Research in Context Pilot Study at the Faculty of Electrical Engineering – TU/e (Confidential)*

ERiC publicatie 1004 (2010). *Maatschappelijke relevantie van Werktuigbouwkundig onderzoek. Eindrapport ERiC project CTW Universiteit Twente (vertrouwelijk)*

Hee, Van K, Overveld, K van, (2010). *Criteria voor het beoordelen van een technologisch ontwerp*

HEFCE (2009). *Research Excellence Framework, Second consultation on the assessment and funding of research*

KNAW-commissie kwaliteitszorg (2008). *Kwaliteitszorg in de wetenschap, Van SEP naar KEP: Balans tussen rechtvaardigheid en eenvoud*

Royal Academy of Engineering (2008). *Research Excellence Framework*

Royal Academy of Engineering (2009). *Research Excellence Framework, second consultation. Response to the Higher Education Funding Council for England*

Schouten, M.J.W. (2009). *Kwaliteitsbeoordeling Ontwerp- en Constructiedisciplines*. Technische Universiteit Eindhoven

Stokes, D.E. (1997). *Pasteur's Quadrant;* Basic Science and Technological Innovation, Brookings Institution Press, Washington

VSNU, KNAW and NWO (2009), *Standard Evaluation Protocol 2009-2015*, only available at www.knaw.nl/sep

# APPENDIX 1
# LIST OF INTERVIEWEES

Prof. P.M.G. Apers, University of Twente, Information science
Prof. J.A. Battjes, Delft University, Civil engineering
Prof. A. Beukers, Delft University, Aerospace engineering
Prof. C.J.P.M. de Bont, Delft University, Industrial design
Prof. A.C.J.M. Eekhout, Delft University, Architecture
Prof. K.M. van Hee, Eindhoven University, Mathematics and information science
Prof. J.J. Heijnen, Delft University, Bioprocess engineering
Dr F.D. van der Hoeven, Delft University, Architecture
Prof. F.J.A.M. van Houten, University of Twente, Mechanical engineering/ Industrial design
Prof. J.T.F. Keurentjes, Eindhoven University, Chemical engineering
Prof. H. Leegwater, Eindhoven University, Chemical engineering
Prof. M.C.M. van Loosdrecht, Delft University, Bioprocess engineering
Prof. J.B.O.S. Martens, Eindhoven University, Industrial design
Prof. B. Nauta, University of Twente, Electrical engineering
Prof. C.J. Overbeeke, Eindhoven University, Industrial design
Prof. W.A. Poelman, University of Twente, Industrial design
Prof. J.M. Post, Eindhoven University, Architecture
Prof. A.F.J. van Raan, Leiden University, Quantitative analysis
Prof. M.J.W. Schouten, Eindhoven University, Industrial design
Prof. M. Steinbuch , Eindhoven University, Mechanical engineering
Dr S. Silvester, Delft University, Industrial design
Dr P.E. Vermaas, Delft University, Philosophy
Dr J. Visschers, NIKHEF, Instrumentation design
Prof. H.J. de Vriend, Deltares, Civil engineering
Prof. P.J.V.V. van Wesemael, Eindhoven University, Architecture
Prof. J.J. van Wijk, Eindhoven University, Information science
Prof. H.W. Zandbergen, Delft University, Applied physics

# APPENDIX 2
# ERIC PROJECT INDICATORS

The ERiC project involved three pilots at the Faculty of Electrical Engineering at Eindhoven University of Technology, the Faculty of Architecture at Delft University of Technology, and the Faculty of Mechanical Engineering at the University of Twente. The purpose of the pilots was to develop a method for assessing the societal relevance of the research conducted within these faculties. The indicators that were proposed after the Architecture and Electrical Engineering pilots are listed below.

*Table A2.1 Indicators of societal relevance, Architectural Research [ERiC, 2010]*

| Aspect of societal relevance | Indicators |
|---|---|
| Knowledge dissemination | Professional publications, non-scientific publications, exhibitions, etc.<br>Dissemination of technology, artefacts, standards<br>Advisory and consultancy work<br>Popularisation, education and contribution to public debate<br>Training of professionals, mobility of graduates<br>Master's theses and graduation projects that tackle issues arising in actual practice |
| Stakeholder interest | Number of researchers with relevant practical experience in the sector/sectors to which the research programme relates<br>Public funding related to societal issues<br>Contract financing by potential users<br>Collaboration with stakeholders in research, testing and evaluation<br>Consortiums with non-academic organisations |
| Impact and use of results | Income generated by using results<br>Visibility in public debate/public media rankings |

*Table A2.2 Indicators of societal relevance, Electrical Engineering [ERiC, 2010]*

| Aspect van maatschappelijke relevantie | Indicatoren |
|---|---|
| Knowledge dissemination | PhDs in industry<br>MAs in industry<br>Proof of concept<br>Presentations at specialist conferences |
| Stakeholder interest | Joint roadmaps<br>Presentations by invitation<br>Valorisation grants<br>Industrial financing<br>Staff exchanges (exchange programmes such as Casimir, knowledge transfer schemes)<br>Part-time professors from/working in industry<br>Consortiums with industry |
| Impact and use of results | Market introduction and new projects in industry<br>Spin-offs with industrial contacts<br>Patents |

# Appendix 3 Quality Assessment Abroad

## A3.1 United Kingdom: Research Excellence Framework

In the United Kingdom (UK), current discussion of research assessment methods focuses mainly on the Research Excellence Framework (REF), a new process for assessing the quality of scientific research and meant to succeed the Research Assessment Exercise (RAE). The most recent RAE was conducted in 2008. One important objection to this latest round is that the RAE had become an unwieldy and labour-intensive operation. The purpose of the REF is to identify excellence in research and to evaluate its impact. The assessment will focus on coherent research units, and not assess individuals or whole institutions. The first REF assessments are scheduled to take place in 2013 and will determine the distribution of funding in 2014. The precise way in which the REF will be conducted is still subject to debate. This section describes the broad outlines of the REF as announced in late 2009 [HEFCE, 2009]. It also looks at the key points of debate, with an emphasis on the criticism levelled by the engineering sciences.

### Assessment method

The REF proposal accords a key role to peer review. All research offered for assessment is to be judged by expert panels on three criteria:

1. *Output quality:* the main aim of the REF is to identify research that can be considered excellent within an international context. This covers many different categories of output.
2. *Impact:* this criterion assesses the extent to which scientists have made a demonstrable contribution to the economy, society, public policy, culture and the quality of life. Assessment of impact will be by means of case studies.
3. *Environment:* this criterion will take the quality of the research environment into account. This refers to the research strategy, staff development and the training of researchers.

Not all of the criteria are equally important. The proposal is to weight them as follows for the overall assessment: output quality 60 percent, impact 25 percent and environment 15 percent. The overall assessment will consist of a system of stars, starting with ‘unclassified’ (below standard quality or work does not meet the definition of research) and rising to four stars (‘exceptional’).

Research is to be divided into various review panels, but the details are still subject to debate. A two-tier structure is envisaged, with four main panels and thirty sub-panels. In this proposal, most of the technical sciences will be grouped and assessed in a single large engineering panel covering all research in electrical engineering, chemical engineering, mechanical engineering, aerospace engineering, mining, civil engineering and materials science. Coming under same main panel but a different sub-panel will be computer science and informatics, mathematics, physics, chemistry and earth systems. Architecture has been categorised under the panel for architecture, the built environment, town and country planning, which itself resides under another main panel.

## Topics of discussion

### *Use of bibliometrics*

The usefulness of citation and other quantitative indicators has generated a great deal of discussion. The question is whether such indicators could replace the expert review in the more ‘science-based’ disciplines. Ultimately, the conclusion was that such indicators were not robust enough and would not be acceptable to the sector. Nevertheless, the aim is to let these indicators play a greater role than they do in the RAE. The current idea is for the panels to use the indicators to inform their reviews, in any event in the medical, health, biological and physical sciences, psychology, engineering and computer science. For now, it is up to the panels to decide how they are going to use these indicators.

Much of the discussion until now has focused on the impact criterion because it is very unclear precisely what is being assessed and to whom certain impacts should be attributed. This is also a new criterion compared to the RAE. The units that are being assessed must submit one or more case studies describing their impact. Impact is regarded as the impact on the economy, society, public policy, culture or the quality of life. The following indicators have been proposed:

- indicators of research income generated from users (businesses, government, research charities);
- indicators of collaboration with users;
- indicators specific to a particular area of research. The indicators would be selected from a kind of menu. A proposal for this type of indicator can be found in HEFCE, 2009.

### *Criticism from the engineering sciences*

In a major consultation round, various stakeholders were asked to comment on the new framework. The commentary of the Royal Academy of Engineering is the most pertinent within the context of this advisory report [Royal Academy of Engineering, 2009]. The Academy’s key objections are the following.

- The Academy agrees that citation analysis is not robust enough to be used to assess research results and doubts whether it ever will be, given the breadth of the discipline and the various types of output it produces. It also points out that the publication traditions vary considerably from one sub discipline to the next.
- The Academy applauds the use of citation information to inform the panels, but points out that available information from the citation databases is often insufficient for the engineering disciplines, leading to major differences in the information submitted by the various sub disciplines.
- The Academy agrees that the impact of research should be considered in the assessment. Given the many different categories of engineering output, however, this will be a difficult process and it will take time to build up the necessary experience using it. At this stage, the proposed weight factor of 25 percent is therefore considered too much for this criterion.
- In the *Research Assessment Exercise*, industrial members were full members of the expert panels. In the current REF proposal, they are only involved in assessing impact. The Academy is in favour of also involving members from industry in the engineering panel when assessing all the criteria.
- The Academy seriously questions the proposal to have all engineering research reside under a single panel. That is first of all because a single panel will be obliged to review an enormous amount of research (produced by approximately 4400 FTE researchers). Secondly, it expects that competition between the engineering disciplines will increase as a result, ultimately resulting in less money for the engineering sciences.

## A3.2 Australia: Excellence in Research for Australia

The Australian Minister for Innovation, Industry, Science and Research, Senator Kim Carr, launched a new initiative in 2008 to develop an ambitious method for evaluating scientific research. The Excellence in Research for Australia (ERA) initiative is being developed by the Australian Research Council *(*ARC) together with the Department of Innovation, Industry, Science and Research and is based on a combination of metrics and expert review. A trial ERA was run in 2009 in two clusters, physical, chemical and earth sciences and humanities and creative arts. In 2010, the ERA assessment method will be introduced for all disciplines.

The exercise is intended to collect information on the quality of research, identify national strengths and weaknesses, excellence and emerging areas where there are opportunities for development, and to position Australia within the international research arena.

### *Assessment method*

For evaluation purposes, research is divided into clusters of more or less similar disciplines:

1. Physical, chemical and earth sciences
2. Humanities and creative arts
3. Engineering and environmental sciences
4. Social, behavioural and economic sciences
5. Mathematical information and computing sciences
6. Biological and biotechnological sciences
7. Biomedical and clinical health sciences
8. Public and allied health sciences

These disciplines are themselves divided into a large number of subdisciplines. Engineering, for example, consists of the subdisciplines aerospace engineering, automotive engineering, biomedical engineering, etc. The institutions involved are basically evaluated on the basis of these subdisciplines. The higher-level discipline is only assessed if the relevant subdiscipline's output is too meagre for a proper evaluation. Assessment therefore does not cover whole faculties or individual researchers. The ERA method uses a range of indicators to measure research quality (see Table A3.1).

*Table A3.1 Assessment criteria in Excellence in Research for Australia*

| Category | Indicators |
|---|---|
| Ranked outlets | Publication in peer-reviewed journals<br>Publication in reviewed conference proceedings |
| Citation analyses | Relative citation impact (against world and Australian HEP benchmarks)<br>Distribution of papers by world centile thresholds, profiled against Australian HEP average |
| Research volume and activity | Distribution of research output by outlet types and distribution of FTE and headcounts by HEDSC levels |
| HERDC Research income | Number of grants<br>Average dollar per grant<br>Total income for research<br>Total income per FTE<br>Ratio of total income per FTE compared with discipline-specific benchmark |
| Esteem | Editorship at A and A+ journals<br>Contribution to a prestigious work of reference<br>Curatorial role<br>Fellowship of a Learned Academy<br>Recipient of nationally competitive research fellowship<br>Recipient of prestigious prizes and distinctions |
| Applied | Patents<br>Plant breeder's rights<br>Registered designs<br>Research commercialisation income<br>NHMRC-endorsed guidelines (health care) |

The units of evaluation are assessed and rated by Research Evaluation Committees (RECs), made up of internationally recognised experts with expertise in research evaluation and broad discipline expertise. There are no guidelines for selecting the REC members. The REC chair asks the other members to evaluate the unit based on the data provided. Later, the REC meets to discuss the results. Only in the relevant disciplines/subdisciplines do the REC members (assisted by external experts where necessary) conduct a peer review for certain categories of research output. A very detailed discipline matrix shows what output is eligible for this in each discipline.

The institutions are asked to deliver data on at discipline/subdiscipline level. The data must be entered into the System to Evaluate the Excellence of Research (SEER); they can then be manipulated electronically. The indicator pattern and quantity of output per indicator are then determined, based on the data provided. The indicators are only passed on to the REC if a particular discipline-specific threshold value has been exceeded.

### *Output not limited to scientific publications*

It is possible to include 'non-traditional output types' in submissions in certain research disciplines. This option is limited in the technical sciences to the disciplines built environment and design (including architecture). It is not permitted in any of the other engineering disciplines. Non-traditional output includes architectural designs, creative 3D work, design, games, computer software, and exhibitions.

If an output is submitted for evaluation as non-traditional output, the researchers must explicitly identify the research component. This can be done by submitting the Research Statement for ERA Peer Review of Non-Traditional Research Outputs, which addresses the following categories:

Research background:
- field
- context
- research question

Research contribution:
- innovation
- new knowledge

Research significance:
- evidence of excellence

## A3.3 Finland: Aalto University evaluation

The university system in Finland was completely overhauled in late 2009. The universities gained more financial and administrative autonomy. They were categorised either as public institutions or private foundations. Starting in late 2009, the foundations were no longer part of the state budgetary system. At the moment, a new approach to funding is being developed that places more emphasis on strategic aspects than on indicators of educational/teaching tasks (no. of students, promotions, etc.). This process also involves a search for new quality indicators.

In order to understand how the Finnish are tackling the quality assessment of scientific research, we looked at a recent evaluation of Aalto University in Helsinki. Although the evaluation only concerned a single university, we decided to include it in this report because it was a recent and large-scale evaluation of a university at which the design and engineering sciences are well represented.

Aalto University was founded in 2009 when the Helsinki School of Economics merged with the Helsinki University of Technology and the University of Art and Design Helsinki. That same year, a peer review committee (PRC) consisting of 62 members from 20 countries evaluated the university over the 2003-2008 period. The purpose of the evaluation was to:

- determine the research quality and societal relevance of the new university;
- identify the groups that had the potential to excel internationally;
- analyse the evaluation process itself.

Aalto University wishes to use the recommendations to define and develop research practices, including the strategic allocation of research funding.

### *Assessment method*

The evaluation was conducted by a PRC divided into nine panels:

1. Chemical technology and materials
2. Electronics and electrical engineering
3. Mathematics and physics
4. Computer science and information technology
5. Mechanical engineering
6. Civil engineering and urban and regional studies
7. Business technology, economics and finance
8. Marketing, management and applied business research
9. Architecture , design, media and art research

This meant nine separate PRCs of five to seven members each. Each panel is meant to evaluate a discipline by means of a self-evaluation, a bibliometric analysis and a site visit. Prior to the site visit, the disciplines themselves provided detailed information.

Only active researchers associated with the university were permitted to submit results. There was also a bibliometric analysis. The PRCs were asked to pay particular attention to the university's international research position. International research hence was consistently used as a benchmark.

Based on the site visit and the information provided (self-evaluation and bibliometric analysis), the PRC awarded marks to each discipline on the following criteria:

1. Research quality
2. Research impact
3. Societal impact
4. Research environment

5. Future research potential

The marks ranged from 1 to 5, with 5 standing for excellent international visibility and 1 for internationally emergent.

The PRC expressly included the results of the bibliometric analysis – specifically the NCSf (citation score standardised by research field) in its assessment of the various disciplines. This helped determine the university's visibility in the national and international research arena. The discipline of architecture, design, media and art research is an exception to this; the bibliometric analyses were not considered in this case because there were too few ISI publications.

***Output not limited to scientific publications***

The PRC indicates that it was easy enough to evaluate disciplines 1 to 8 using the current science-based criteria. It also indicates that, on paper, Aalto University leaves enough leeway to include all categories of research output (i.e. artefacts in addition to publications), but that the university administrators in fact strongly suggested that science-related output be provided. As a result, the units operating in the discipline of architecture, design, media and art research submitted much smaller quantities of output (i.e. only their articles) than the site visit actually turned up. The relevant panel believes that the administrators had seriously undervalued these units' research and its societal relevance.

For more information about this evaluation, see [Aalto University, 2009a, b, c].

## A3.4 United States, National Science Foundation

The committee looked at the National Science Foundation (NSF) in the United States. The NSF is an independent federal agency set up: 'to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense...'. With an annual budget of USD 6.9 billion, the NSF subsidises approximately 20 percent of federally funded basic research at American universities and colleges. Unlike the assessment system described in the previous section, the NSF concentrates mainly on the ex-ante evaluation of research proposals. Besides the NSF, there are many private, mission-based and industrial funding bodies active in the United States. They usually do not rely as heavily on clearly defined criteria.

Peer review is crucial to the NSF's evaluation of research proposals. Such proposals are evaluated mainly according to two criteria:

1. What is the scientific merit of the proposed research activity?
2. What is the broader impact of the proposed research activity?

Additional criteria may also be imposed, depending on the programme.

### *Assessment method*
The proposal is submitted to a program officer, who passes it on to a review panel. The panel, consisting of thee to ten members, evaluates the proposal on the above criteria. The program officer then draws up an analysis and a recommendation based on the panel's views. The NSF division director uses this recommendation to inform his or her final funding decision.

Those submitting research proposals are asked to suggest suitable referees. They may also indicate who would not be a suitable referee.

## A3.5 Norway

In Norway, research is usually the province of a general or specialist university, university colleges and research institutes. The Ministry of Education is the main coordinating body and the other relevant ministries are responsible for promoting and financing research in their fields. Part of the budget is provided directly by the Ministry of Education and Research. Before, that was usually on the basis of student numbers. Since 2003, funding is divided into a) basic funding, b) education and c) research. Research funding is based on performance and strategic considerations. Performance is measured by looking at:
- number of Bachelors completed;
- number of Masters completed;
- number of PhDs completed;
- external collaboration;
- staff composition (number of professors);
- number of EU and RCN grants;
- patents and output in the form of scientific and scholarly publications.

The publications are weighted depending on the publication method (article, book) and channel (journal, website, publisher if book).

Universities must enter all the data in a national documentation system because funding depends in part on the documentation in this system.

In addition to the basic funding described here, the Research Council Norway (RCN) also awards competitive grants. Proposals are evaluated by means of peer review based on a huge number of criteria.

The RCN also conducts national evaluations:
- of RCN-funded projects
- at institutional level
- at discipline level.

The purpose of the evaluations is to improve research quality, relevance and efficiency and to identify strengths and weaknesses at national and institutional levels. The RCN also uses the outcomes when drawing up the national research strategy. The outcomes are also helpful when allocating research funding. The RCN is working on a follow-up model for these evaluations.

## A3.6 Sweden

In Sweden, research is conducted at 14 universities and 25 institutions of higher education. Most publicly funded research takes place within the latter. Some of that research is funded directly by government. The public authorities also fund the Research Councils and the Sectoral Research Agencies.

The largest funding body is the Swedish Research Council (SRC), which has an annual budget of SEK 4 billion. The SRC awards grants to research projects in six different areas, including the natural and engineering sciences. Research proposals are submitted to forty panels where they are subjected to peer reviews.
The second largest funding body is the Swedish Agency for Innovation Systems (VINNOVA), with a budget of SEK 2.15 billion. This agency makes funding available in six research areas, including natural sciences and technology.

Sweden does not have a national ex-post research evaluation system.

## A3.7 France

The French system of higher education consists of research institutes, universities and *grandes ecoles*. The institutes are the main centres of research. They decide on and fund their own research and evaluate it themselves. At the moment, the public authorities would like to give the universities more autonomy and turn the research institutes into agencies that distribute research funding on a project-driven basis; see the National Research and Innovation Strategy, SNRI. According to this document, research should be shifted over to the universities. Institutes can also apply for funding through the National Research Agency (ANR).

France does not have a national ex-post evaluation system for research at universities, *grandes ecoles* and institutes.

## A3.8 Germany

Public research institutions in Germany are divided into four networks:
- the Max Planck Society (MPG)
- the Fraunhofer Society (FhG)
- the Helmholtz Association of German Research Centres (HGF)
- the Wilhelm-Gottfried-Leibniz Scientific Community (WGL)

The universities receive their basic funding from the relevant federal states. The largest federal funding body is the German Research Foundation (DFG), which has an annual budget of EUR 1.3 billion. The DFG funds all the disciplines,  from the humanities to engineering, and gets its money from both the federal and the state authorities according to a fixed distribution key. The DFG comes under the auspices of the Federal Ministry of Education and Research (BMBF). Other major funding bodies include the private Alexander von Humboldt Foundation and the German Academic Exchange Service (DAAD), which finances international exchange and other programmes and projects.

Although research is never evaluated at national level, the DFG does produce a ranking (the Förder  ranking), based on the following criteria:
- number of positive funding awards by the DFG;
- composition of research group;
- funding acquired via contract research;
- position in DFG-subsidised programmes;
- number of DFG reviewers;
- number of Alexander von Humboldt visiting researchers;
- number of DAAD researchers and students;
- participation in EU framework programmes;
- articles published in international journals.

In addition, the federal states of North Rhine-Westphalia (based on external funding and number of PhDs) and Lower Saxony (has its own Academic Advisory Council) have developed their own evaluation systems. (The applied research institutes in Baden-Württemberg were evaluated in 2008 at the request of the Ministry of Economic Affairs.) None of these evaluations is linked directly to funding. These evaluation systems are not trendsetting internationally.

The aim of the Excellence Initiative [*Exzellenzinitiative*], a DFG funding programme, is to identify research excellence and to promote and improve the quality of higher education and research in the broader sense. Germany hopes that this initiative will improve its international competitiveness. Research proposals submitted by the entire Germany scientific community are subject to peer review. The proposals that are ultimately selected are then showcased.

## A3.9 Austria

Research in Austria is funded by:
- the Austrian Research Promotion Agency (FFG)
- the Austrian Science Fund (FWF)
- the Austria Wirtschaftsservice (AWS)

These funding bodies are part of the Ministry of Science and Research (BMWF), the Ministry of Transport, Innovation and Technology (BMVIT) and the Ministry of Economy, Family and Youth (BMWFJ), but the division of strategic responsibilities between the ministries and the funding bodies is not entirely clear and the responsibilities of the various funding bodies overlap.

In 2003, the National Foundation for Research, Technology and Development was set up in order to guarantee the major funding bodies an ongoing financial buffer and to ensure the continuity of large-scale, long-term projects. The money comes from the Austria National Bank (OeNB) and the European Recovery Fund (ERP).

Austria does not have a national ex-post evaluation system for research. The Research & Technology Evaluation platform (Fteval) is attempting to set up a transparent evaluation system, part of a push to optimise strategic research, technology and development (RTD) policy planning. Fteval also aims to promote an evaluation culture in Austria in cooperation with Technology and Research policy-makers.