

**KWALITEITSBEOORDELING IN DE
ONTWERPENDE EN CONSTRUERENDE DISCIPLINES**



2011 Koninklijke Nederlandse Akademie van Wetenschappen (KNAW)

© Sommige rechten zijn voorbehouden / Some rights reserved

Voor deze uitgave zijn gebruiksrechten van toepassing zoals vastgelegd in de Creative Commons licentie. [Naamsvermelding 3.0 Nederland]. Voor de volledige tekst van deze licentie zie <http://www.creativecommons.org/licenses/by/3.0/nl/>

Koninklijke Nederlandse Akademie van Wetenschappen

Postbus 19121, 1000 GC Amsterdam

Telefoon + 31 20 551 0700

Fax + 31 20 620 4941

knaw@bureau.knaw.nl

www.knaw.nl

Eerste druk: november 2010

Tweede druk: april 2011

pdf beschikbaar op www.knaw.nl


Basisvormgeving: edenspiekermann, Amsterdam

Opmaak en bureauredactie: Ellen Bouma, Alkmaar

Druk: Bejo druk & print, Alkmaar

Foto cover: Nationale beeldbank/Taco Gooiker

ISBN: 978-90-6984-619-4

Het papier van deze uitgave voldoet aan  iso-norm 9706 (1994) voor permanent houdbaar papier.



Dit advies is gemaakt van FSC-papier en gecertificeerd onder nummer CU-COC-804134-N.

KWALITEITSBEOORDELING IN DE ONTWERPENDE EN CONSTRUERENDE DISCIPLINES

EEN SYSTEMATISCH KADER

Koninklijke Nederlandse Akademie van Wetenschappen
Advies Raad voor Technische Wetenschappen, Wiskunde en Informatica,
Natuur- en Sterrenkunde en Scheikunde van de KNAW

VOORWOORD

De grote waarde van de wetenschap is haar natuurlijke diversiteit. Het is een diamant met vele facetten. Deze verscheidenheid uit zich in de keuze van onderwerp, de methode van onderzoek en de uiteindelijke wijze van verslaggeving. Zij is in haar diepste zin een reflectie van de rijkdom aan verschijnselen in de wereld en in onze geest.

De wetenschap verdient het dat haar grote waarde zichtbaar gemaakt wordt op een manier die zo objectief mogelijk is en tegelijkertijd deze diversiteit respecteert, zo niet celebreert. Een optimale kwaliteitsborging is niet alleen nodig om in de buitenwereld draagvlak voor onderzoek te verwerven. Ook de beoefenaars zelf zijn gebaat bij een heldere evaluatie van hun onderzoek. De uitdaging hierbij is de meetmethoden zo dicht mogelijk te laten aansluiten bij de intrinsieke waarde van het onderzoeksveld. Hier valt nog veel werk te verrichten.

Het in kaart brengen van wetenschappelijke publicaties en hun citaties is daarbij een belangrijke eerste stap. Volgens dit criterium blijkt het Nederlands wetenschappelijk onderzoek van wereldklasse te zijn. Het rapport *Wetenschaps- en Technologie-Indicatoren* dat tweejaarlijks in opdracht van het ministerie van OCW wordt opgesteld, heeft dit onlangs nog eens bevestigd. Nederland behoort in termen van citatie-impact tot de top van de wereldranglijst, samen met de Verenigde Staten, Zwitserland en Denemarken. Bovendien zijn wij bijzonder productief.

Natuurlijk is het goed nieuws dat Nederlandse wetenschappers langs deze as gemeten tot de wereldtop behoren. Het is echter belangrijk dat wij ons realiseren dat wetenschappelijke kwaliteit zich niet *alleen* laat meten aan de hand van wetenschappelijke publicaties en citatie-impact. Voor vele wetenschapsdomeinen zijn citaties slechts het halve verhaal, als het niet nog minder is. De standaardmeetmethoden missen daarmee grote delen van belangrijke onderzoeksvelden. Zo zijn de producten van de ontwerpende en construerende wetenschappen niet alleen *peer-reviewed*

tijdschriftartikelen, maar ook *conference proceedings*, ontwerpen, software en constructies. In de geesteswetenschappen verschijnen veel publicaties in boekvorm of in andere talen dan Engels waardoor ze vrijwel onzichtbaar zijn in de citatie-impact. Dergelijke disciplines kunnen daarom minder uit de voeten met de gebruikelijke manier van kwaliteitsbeoordeling. Het is dan ook niet toevallig dat de KNAW verzocht is te adviseren over hanteerbare beoordelingscriteria voor juist deze disciplines.

De KNAW heeft in 2008 in het advies *Kwaliteitszorg in de wetenschap; van SEP naar KEP* al bepleit dat een nieuw protocol voor de kwaliteitsbeoordeling voldoende flexibiliteit moet hebben om de verschillen tussen vakgebieden te accommoderen. Onderzoekers die bijvoorbeeld de maatschappelijke en culturele relevantie van hun werk of de economische waarde naar voren willen brengen, moeten daartoe in de gelegenheid worden gesteld. De KNAW heeft daarom met genoegen deze adviesvragen opgepakt. Het nu voorliggende advies geeft voor de ontwerpende en construerende wetenschappen invulling aan de roep om een flexibeler beoordeling. Ik ben verheugd dat de commissie tot de conclusie komt dat een specifieke set van criteria voor deze disciplines niet nodig is. Wetenschappelijke kwaliteit en maatschappelijke relevantie volstaan. De disciplinespecifieke invulling zit in de indicatoren waarmee deze twee criteria beoordeeld moeten worden. Dit is een belangrijk leidend principe dat tegelijkertijd de universaliteit en de diversiteit van de wetenschap benadrukt.

De commissie geeft aan bewust geen definitieve selectie of onderlinge weging van de indicatoren te maken. Dit is een stap die voor de verschillende beoordelingsituaties door de wetenschappers zelf in nauwe samenspraak met de universitaire bestuurders moet worden gemaakt. Ik hoop dat de 3TU-Federatie en de ontwerpende en construerende wetenschappers deze handschoen voortvarend zullen oppakken, zodat de Nederlandse wetenschap voortaan in al haar verscheidenheid kan flonkeren.

Robbert Dijkgraaf
President KNAW

Bij de tweede druk

In de tweede druk zijn enkele tekstuele correcties aangebracht. Daarnaast is aanbeveling 4 op pagina 45 verplaatst naar de tekst onder de kop 'Onderzoeksfinanciers'. Verder is in tabel 3.1 en in de begeleidende tekst de kop 'overige output' vervangen door 'ontworpen artefacten'.

INHOUD

VOORWOORD 5

SAMENVATTING 9

1. INLEIDING 13
 - 1.1 Achtergrond 13
 - 1.2 Samenstelling en opdracht van de commissie 13
 - 1.3 Werkwijze 14
 - 1.4 Relatie met andere initiatieven 15
 - 1.5 Afbakening 16

2. KNELPUNTEN BIJ DE KWALITEITSBEOORDELING 19
 - 2.1 Inleiding 19
 - 2.2 Ervaren knelpunten 21
 - 2.3 Evaluatie ervaren knelpunten en conclusies 23

3. BEOORDELINGSCRITERIA 27
 - 3.1 Inleiding 27
 - 3.2 Beoordelingscriteria 29
 - 3.3 Toelichting op de indicatoren 30
 - 3.4 Het relatieve belang van de criteria en de indicatoren 33
 - 3.5 Profielen 34
 - 3.6 Selectie van *peers* 35
 - 3.7 Conclusies 35

4. INVENTARISATIE VAN BUITENLANDSE ERVARINGEN EN INZICHTEN 37
 - 4.1 Inleiding 37
 - 4.2 Samenvatting ervaringen in het buitenland en conclusies 37

5. CONCLUSIES EN AANBEVELINGEN 41

LITERATUUR 46

BIJLAGEN

1 Geïnterviewden 47

2 Indicatoren ERIC-project 48

3 Kwaliteitsbeoordeling in het buitenland 49

SAMENVATTING

Achtergrond

De 3TU-Federatie heeft aan de KNAW gevraagd te adviseren over de criteria die gebruikt kunnen worden bij de ex ante- en ex postbeoordeling van technisch-wetenschappelijk werk in de ontwerpende en construerende disciplines. Ontwerpende en construerende wetenschappers ervaren namelijk regelmatig problemen bij de beoordeling van de kwaliteit van hun wetenschappelijk werk bij visitaties, benoemingen en onderzoeksaanvragen. De indicatoren voor kwaliteit die daarbij gebruikt worden zijn ontleend aan de meer fundamentele wetenschappen (publicaties in ISI-tijdschriften, impactfactoren, citaties, Hirsch-index) en zijn in de ogen van deze wetenschappers ontoereikend. Ten eerste omdat hun producten niet alleen *peer-reviewed* tijdschriftartikelen zijn, maar ook *conference proceedings*, ontwerpen en constructies. Ten tweede, omdat hun onderzoek in het algemeen meer contextspecifiek en multidisciplinair van aard is dan dat van de meer fundamentele wetenschappen. Tijdschriften in deze gebieden hebben daardoor een lagere impactfactor en worden systematisch lager gewaardeerd.

De KNAW heeft hiertoe een commissie ingesteld onder voorzitterschap van prof. dr. ir. A.W.M. Meijers (TU Eindhoven) met als opdracht een advies op te stellen over de criteria voor de ex ante- en ex postbeoordeling van:

- ontwerpende en construerende activiteiten in de technische disciplines die als wetenschappelijk kunnen worden aangemerkt
- wetenschappelijk onderzoek in de ontwerpende en construerende disciplines

De criteria moeten voldoen aan de volgende vereisten:

- van nut voor organisaties die wetenschappelijke activiteiten financieren (NWO, ed.)
- van nut voor universiteiten bij de beoordeling van wetenschappelijk personeel (onderzoekers / ontwerpers)
- geloofwaardig in internationaal verband

In het voorliggende advies doet de commissie een voorstel voor deze criteria en doet zij verslag van haar bevindingen. Het advies is mede gebaseerd op interviews met 27 wetenschappers uit verschillende disciplines en een inventarisatie van de buitenlandse inzichten en ervaringen.

Algemene criteria met disciplinespecifieke indicatoren

Een belangrijk uitgangspunt voor de commissie is dat de beoordelingsmaatstaven voor een bepaalde discipline niet volledig verschillend kunnen zijn van die van een andere discipline. Dat zou de maatstaven arbitrair maken en opportunistisch. De commissie heeft daarom gestreefd naar een beoordelingskader met twee elementen:

- algemeen geldende kwaliteitscriteria
- disciplinespecifieke indicatoren voor deze criteria. Deze indicatoren geven empirische informatie over de mate waarin een persoon, groep of onderzoeksvorstel voldoet aan de kwaliteitscriteria. Dit kan kwantitatieve maar ook kwalitatieve informatie te zijn

De commissie concludeert dat – in tegenstelling tot wat bij de start van het adviestraject werd verwacht – de kwaliteit van het wetenschappelijke werk in de ontwerpende en construerende disciplines kan worden beoordeeld aan de hand van de twee criteria die ook voor andere wetenschapsgebieden gelden: 1) wetenschappelijke kwaliteit en 2) maatschappelijke relevantie. Er zijn dus geen aanvullende criteria nodig voor ontwerpen en construeren. Uit de internationale inventarisatie blijkt dat deze twee criteria ook internationaal gezien geloofwaardig zijn voor deze disciplines.

Dat er geen andere kwaliteitscriteria nodig zijn, betekent echter niet dat alle wetenschapsgebieden op dezelfde manier kunnen worden beoordeeld. Beoordelen van kwaliteit behoort maatwerk te zijn vanwege de verschillen tussen disciplines (waaronder publicatiecultuur), typen wetenschappelijke activiteiten (ontwerpen, onderzoeken) en beoordelingssituaties (visitatie, benoeming, onderzoeksvorstel). Dit maatwerk komt tot uiting in de keuze en weging van de disciplinespecifieke indicatoren voor de criteria wetenschappelijke kwaliteit en maatschappelijke relevantie. Tabel 3.1 op pagina 29 geeft een overzicht van de indicatoren die de commissie relevant acht voor de ontwerpende en construerende disciplines.

Vervolgstap nodig

De commissie heeft er nadrukkelijk *niet* voor gekozen om een nadere selectie of onderlinge weging van de indicatoren voor de ontwerpende en construerende wetenschappen te maken. Dit is een vervolgstap die door de wetenschappers zelf (*peers*) in nauwe samenspraak met de universitaire bestuurders voor de verschillende beoordelingssituaties moet worden gemaakt. De commissie adviseert het bestuur van de 3TU's om ruimte te scheppen voor een disciplinespecifieke kwaliteitsbeoordeling in de technische wetenschappen en aan de ontwerpende en construerende disciplines te vragen indicatoren en hun relatieve gewicht vast te stellen voor de beoordeling van kwaliteit in die disciplines.

Vergelijking tussen vakgebieden problematisch

Tijdens de interviews in het veld bleek dat de beoordeling van kwaliteit van het wetenschappelijk werk door internationaal gerespecteerde *peers* vrijwel nergens als probleem wordt ervaren. De problemen ontstaan bij vergelijking *tussen* vakgebieden of als de gebieden te breed worden. De commissie vindt dat kwaliteit alleen *binnen* disciplines zinvol kan worden vergeleken. Een *one-size-fits-all*-benadering bij de beoordeling van kwaliteit, hoe gewenst ook vanuit bestuurlijk oogpunt vanwege zijn eenvoud, doet geen recht aan significante disciplinaire verschillen en zal dus altijd bepaalde vakgebieden ten onrechte bevoordelen. De commissie adviseert onderzoeksfinanciers zoals NWO en STW om meer in te zetten op programma's voor disciplines die niet goed passen bij de bestaande gebiedsindeling en de bestaande manier van kwaliteitsbeoordeling. Dat geldt in het bijzonder voor de ontwerpende en construerende wetenschappen. Daarvoor moeten kwaliteitsindicatoren worden gehanteerd die recht doen aan deze disciplines.

Peer review centraal

De commissie is van mening dat *peer review* het uitgangspunt moet zijn voor een goede kwaliteitsbeoordeling en juicht het daarom toe dat de meeste onderzoeksfinanciers gebruik maken van *peer review*. De keuze van deze *peers* is wel cruciaal en heeft invloed op de uitkomst van de beoordeling. Het zou goed zijn om de keuze van deze *peers* meer te objectiveren en transparanter te maken, binnen NWO, visitatiecommissies en benoemingsadviescommissies. De *peers* die betrokken zijn bij de beoordeling van kwaliteit in de ontwerpende en construerende wetenschappen moeten zelf ook worden beoordeeld aan de hand van het brede spectrum van indicatoren en niet alleen op publicatiegedrag.

Belang van publicaties

In dit advies wordt voorgesteld voor kwaliteit ook andere indicatoren dan *peer-reviewed* publicaties te gebruiken. De commissie benadrukt echter dat *peer-reviewed* publicaties ook in de ontwerpende en construerende disciplines belangrijk zijn en blijven. Ze dragen bij aan onderlinge kwaliteitstoetsing, aan verspreiding van kennis en aan verdere verwetenschappelijking. Publiceren is in deze disciplines of onderdelen daarvan niet altijd vanzelfsprekend. De commissie adviseert wetenschappers in de ontwerpende en construerende disciplines waar nodig krachtig een cultuur van *peer-reviewed* publicaties te bevorderen. Het is daarbij van belang te zoeken naar publicatievormen die passen bij de (deel)discipline.

1. INLEIDING

1.1 Achtergrond

Voor ontwerpende en construerende wetenschappers is de beoordeling van de kwaliteit van hun wetenschappelijk werk bij visitaties, benoemingen en onderzoeksaanvragen sinds jaar en dag een probleem. De indicatoren voor kwaliteit die daarbij gebruikt worden, zijn ontleend aan de meer fundamentele wetenschappen (publicaties in ISI-tijdschriften, impactfactoren, citaties, Hirsch-index) en zijn in de ogen van deze wetenschappers ontoereikend. Ten eerste omdat hun producten niet alleen *peer-reviewed* internationale publicaties zijn, maar ook ontwerpen en constructies. Ten tweede omdat hun onderzoek in het algemeen meer contextspecifiek en multidisciplinair van aard is dan dat van de meer fundamentele wetenschappen. Tijdschriften in deze gebieden hebben daardoor een lagere *impactfactor* en worden systematisch lager gewaardeerd. Als een gevolg hiervan ervaren onderzoekers in de ontwerpende en construerende disciplines een gebrek aan wetenschappelijke erkenning en problemen bij het verwerken van externe onderzoeksmiddelen, onderzoeksvisitatie en benoemingen.

1.2 Samenstelling en opdracht van de commissie

Met het oog op de hierboven geschetste problematiek heeft de 3TU-Federatie aan de KNAW gevraagd te adviseren over de criteria die gebruikt kunnen worden bij de ex ante- en ex postbeoordeling van technisch-wetenschappelijk werk in de ontwerpende en construerende disciplines. De KNAW-Raad voor Technische Wetenschappen, Wetenschap, Informatica, Natuur- en Sterrenkunde en Scheikunde (TWINS) heeft daartoe een commissie ingesteld bestaande uit:

- Prof. dr. ir. A. van den Berg, Universiteit Twente
- Prof. dr. ir. R. de Borst, Technische Universiteit Eindhoven

- Prof. dr. P.P.M. Hekkert, Technische Universiteit Delft
- Prof. dr. ir. A.W.M. Meijers (voorzitter), Technische Universiteit Eindhoven
- Prof. dr. R.A. van Santen, Technische Universiteit Eindhoven

Dr. E.E.W. Bruins (STW) was toehoorder. Ir. A. Korbijn (KNAW) trad op als secretaris, dr. J.B. Spaapen (KNAW) als adviseur. Drs. C.S. Tan (KNAW) ondersteunde de commissie bij de interviews en de internationale inventarisatie.

Opdracht

De commissie had als opdracht om een advies op te stellen over criteria voor de ex ante- en ex postbeoordeling van:

- ontwerpende en construerende activiteiten in de technische disciplines die als wetenschappelijk kunnen worden aangemerkt
- wetenschappelijk onderzoek in de ontwerpende en construerende disciplines

De criteria moeten voldoen aan de volgende vereisten:

- van nut voor organisaties die wetenschappelijke activiteiten financieren (NWO e.d.)
- van nut voor universiteiten bij de beoordeling van wetenschappelijk personeel (onderzoekers / ontwerpers)
- geloofwaardig in internationaal verband

In het voorliggende advies doet de commissie een voorstel voor deze criteria en doet zij verslag van haar bevindingen.

1.3 Werkwijze

De commissie heeft allereerst interviews gehouden met 27 representanten uit verschillende disciplines van de technische wetenschappen. Het doel van deze interviews was om na te gaan welke problemen men ervaart met de huidige manier van beoordelen en om na te gaan welke aspecten volgens de belanghebbenden in een nieuwe set beoordelingscriteria meegenomen moeten worden. Een inventarisatie van de ervaren knelpunten is opgenomen in hoofdstuk 2. Een lijst van de geïnterviewden en hun vakgebieden is weergegeven in bijlage 1.

Om een beeld te krijgen van de manier waarop in het buitenland tegen deze problematiek wordt aangekeken en wat we van deze ervaringen kunnen leren, is informatie verzameld over de situatie in het Verenigd Koninkrijk, Australië, Finland, de Verenigde Staten, Noorwegen, Zweden, Frankrijk, Duitsland en Oostenrijk. De resultaten hiervan zijn samengevat in hoofdstuk 4; meer gedetailleerde informatie is weergegeven in bijlage 3.

Op basis van de verzamelde informatie heeft de commissie een analyse gemaakt van de knelpunten bij de beoordeling van kwaliteit in de ontwerpende en

construerende wetenschappen en een voorstel geformuleerd voor geschikte beoordelingscriteria. De bevindingen op hoofdlijnen en een eerste opzet van de criteria zijn in een schriftelijke commentaarrronde aan de geïnterviewden voorgelegd. Het uiteindelijke voorstel voor de criteria is beschreven en toegelicht in hoofdstuk 3.

1.4 Relatie met andere initiatieven

1.4.1 *Evaluating Research in Context*

De 3TU's nemen deel in een drietal *pilots* die in het kader van het ERiC-project (*Evaluating Research in Context*) zijn uitgevoerd. ERiC is voortgekomen uit een project van de Commissie Overleg Sectorraden (COS) over het meten van de maatschappelijke kwaliteit van onderzoek. Tijdens dit project werd een meetmethode ontwikkeld, de *sci_Quest*-methode. In 2006 werd besloten deze methode verder te ontwikkelen om de praktische toepasbaarheid te onderzoeken en te vergroten. Hiertoe is het platform ERiC opgericht waarin participeren: de HBO-raad, KNAW, NWO, VSNU en de afdeling *Science System Assessment* van het Rathenau Instituut. Het ministerie van Onderwijs, Cultuur en Wetenschap is als waarnemer betrokken. In het kader van dit project zijn drie *pilots* uitgevoerd bij de faculteiten elektrotechniek van de TU Eindhoven, bouwkunde van de TU Delft en werktuigbouwkunde van de Universiteit Twente. De *pilots* hadden tot doel een evaluatiemethode te ontwikkelen om de maatschappelijke relevantie van het onderzoek aan deze faculteiten te kunnen beoordelen. De voorgestelde indicatoren uit de *pilots* bouwkunde en *electrical engineering* zijn weergegeven in bijlage 2. Bij het opstellen van het onderhavige advies is kennisgenomen van de uitkomsten van de ERiC-*pilots*.

Het voorliggende advies is zowel smaller als breder in *scope* dan het ERiC-project. Ten eerste richt dit advies zich op een beperkter aantal disciplines, namelijk de ontwerpende en construerende wetenschappen. Ten tweede gaat het advies niet alleen over de beoordeling van de maatschappelijke relevantie van het wetenschappelijke werk, maar ook de beoordeling van de wetenschappelijke kwaliteit van dat werk.

1.4.2 *Standaard Evaluatie Protocol*

De KNAW, NWO en de VSNU hebben voor de beoordeling van wetenschappelijk onderzoek het *Standard Evaluation Protocol 2009 – 2015* (SEP) vastgesteld [VSNU, KNAW en NWO 2009]. De beoordeling volgens het SEP 2009 – 2015 bestaat uit een zesjaarlijkse externe evaluatie via een zelfevaluatierapport en een *site visit*, en een interne *midterm review*, die tussen twee externe beoordelingen plaatsvindt. Het SEP kent vier beoordelingscriteria en twee niveaus van beoordeling, het instituut als geheel (of faculteit of onderzoekschool) en de onderliggende groepen of programma's. De vier beoordelingscriteria zijn:

1. Kwaliteit (*Quality*)
2. Productiviteit (*Productivity*)
3. Maatschappelijke relevantie (*Societal relevance*)
4. Vitaliteit en haalbaarheid (*Vitality & Feasibility*)

De beoordeling van het wetenschappelijke onderzoek via het SEP dient zowel de verantwoording over de geleverde prestaties als het verbeteren van de kwaliteit in de toekomst.

Een onderscheid tussen dit adviestraject en het SEP is dat het SEP alleen betrekking heeft op complete instituten of de onderliggende groepen of programma's. In dit adviestraject wordt ook gekeken naar de beoordeling van onderzoeksvoorstellen en individuele personen. Gezien het belang van het SEP in Nederland is in het onderhavige advies wel gezocht naar criteria die goed aansluiten bij de SEP criteria.

1.4.3 Beoordelingscriteria geesteswetenschappen

Ook in de geesteswetenschappen levert de huidige manier van kwaliteitsbeoordeling problemen op. In citatie-indexen wegen bijvoorbeeld alleen artikelen en geen boeken mee, en van de artikelen alleen de publicaties in het Engels. Een bijdrage aan het maatschappelijk debat wordt niet in de kwaliteitsbepaling meegenomen. Op verzoek van de commissie Nationaal Plan Toekomst Geesteswetenschappen (commissie Cohen) is de KNAW daarom bezig om in overleg met onderzoekers in het veld en met gebruikers van gegevens over onderzoek in de geesteswetenschappen (bijv. NWO), een eenvoudig en adequaat systeem van kwaliteitsindicatoren te ontwerpen. Het advies over de geesteswetenschappen is op moment van schrijven van dit advies nog niet gereed.

1.5 Afbakening

De opdracht van de commissie was om te adviseren over criteria voor de beoordeling van kwaliteit van (1) ontwerpende en construerende activiteiten die als wetenschappelijk kunnen worden aangemerkt en (2) onderzoek in de ontwerpende en construerende wetenschappen. Voor deze tweeledige opdracht is gekozen omdat er ook binnen meer 'science-achtige' vakgebieden ontwerpende en construerende activiteiten plaatsvinden die met de huidige systematiek niet goed te beoordelen lijken te zijn (bijvoorbeeld het ontwerpen van instrumentatie). Daarnaast vindt er in de ontwerpende en construerende disciplines onderzoek plaats dat multidisciplinair van aard is en dat een eigen signatuur heeft.

In het kader op de volgende pagina wordt globaal geschetst wat verstaan wordt onder ontwerpende en construerende activiteiten. Ontwerpende en construerende disciplines worden door de commissie dan opgevat als die disciplines waarin dit soort activiteiten een belangrijke rol spelen. Voorbeelden zijn industrieel ontwerpen, bouwkunde,

informatica, werktuigbouwkunde, scheikundige technologie, biotechnologie, civiele techniek, maritieme techniek, elektrotechniek en lucht- en ruimtevaarttechniek. De commissie is van mening dat de in hoofdstuk 3 voorgestelde criteria en indicatoren in ieder geval op deze disciplines goed van toepassing zijn.

TECHNOLOGISCH ONTWERPEN

Bij ontwerpen gaat het niet om het verwerven van een beter begrip van de bestaande werkelijkheid (ware of betrouwbare uitspraken daarover) maar om het creëren van iets nieuws. Ontwerpers voegen iets toe aan de werkelijkheid dat eerder niet bestond en dat waarde of functionaliteit heeft voor gebruikers.

Ontwerpen is een generieke activiteit die in veel domeinen voorkomt: van het ontwerpen van een kunstwerk tot het ontwerpen van een nieuwe wet of van een chemische installatie. In de ontwerpende en construerende disciplines gaat het om het ontwerpen van nieuwe technologische artefacten. Dat kunnen producten of processen zijn die fysisch gerealiseerd kunnen worden (gebouwen, telefoons, elektriciteitscentrales, moleculen, chips), maar ook abstracte algoritmes (software). Daarnaast kunnen ook diensten worden ontworpen (internettransacties), levende organismen (genetisch gemodificeerde gewassen), of menselijke ervaringen (*virtual reality*).

Er zijn vele manieren om ontwerpprocessen te karakteriseren. Sommige leggen het accent op het rationeel nemen van beslissingen op basis van wetenschappelijke modellen, waarbij de oplossingsruimte steeds verder verkleind wordt; anderen proberen ook niet-rationele, creatieve of *art-like* aspecten van het ontwerpproces te verdisconteren. De Amerikaanse *Accreditation Board for Engineering and Technology* karakteriseert ontwerpen als volgt: *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)*. De opties waaruit gekozen kan worden in dit proces zijn overigens niet voorgegeven maar worden daarin juist ontwikkeld.

Een ontwerpproces resulteert in een product: het ontwerp. Dit is een representatie van het beoogde artefact in een of andere vorm (prototype, computermodel, schaalmodel, tekening). Daaraan kan van alles onderzocht worden (prestatie, gebruikersaspecten, produceerbaarheid, kosten, etc). In veel construerende disciplines wordt een ontwerp daarom gezien als een platform voor onderzoek. Sommigen karakteriseren een ontwerp als het analogon in de construerende wetenschappen van een experiment in de natuurwetenschappen [Hee, van en Van Overveld, 2010].

2. KNELPUNTEN BIJ DE KWALITEITSBEOORDELING

2.1 Inleiding

Wetenschappelijk onderzoek kent een grote verscheidenheid. Deze verscheidenheid komt voort uit de verschillende onderzoekstradities en stijlen in de alfa-, bèta- en gammadisciplines. Binnen elk van deze disciplinaire clusters is er opnieuw sprake van verschillende typen van onderzoek. Deze verschillen worden vaak beschreven in termen van het doel van het onderzoek: het streven naar fundamenteel begrip of het oplossen van praktische problemen. Hiermee hangt samen de indeling van het onderzoek in termen van fundamenteel en toegepast onderzoek. De discussie in de literatuur laat zien dat dit geen elkaar uitsluitende categorieën zijn. Een vaak gebruikte categorisering van wetenschappelijk onderzoek is weergegeven in tabel 2.1.

Tabel 2.1 Categorisering van wetenschappelijk onderzoek, ontleend aan [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

Binnen de technische universiteiten is het spectrum van wetenschappelijke activiteiten nog groter en complexer omdat ook ontwerpen en construeren onderdeel uitmaken van het wetenschappelijke portfolio. Voor technische universiteiten zijn deze activiteiten essentieel. Het is immers onderdeel van hun missie om bij te dragen aan

de oplossing van maatschappelijke problemen of aan het creëren van economische kansen en met name bij het ontwerpen en construeren wordt de verbinding met de praktijk gemaakt.

De hierboven genoemde verscheidenheid in wetenschappelijke activiteiten roept de vraag op of het mogelijk is om met één meetlat de kwaliteit van al deze activiteiten te beoordelen. Kunnen criteria die bijvoorbeeld ontleend zijn aan het fundamentele (natuurkundige) onderzoek probleemloos worden gebruikt bij de beoordeling van de kwaliteit van het toegepaste onderzoek, of van construerende en ontwerpende activiteiten? Deze vraag wordt door de beoefenaars van de ontwerpende en construerende disciplines meestal ontkennend beantwoord en vormt de kern van de controverse over de beoordeling van kwaliteit in deze disciplines.

De discussie hierover speelt al vele jaren. In 2000 bracht het Discipline-overlegorgaan Construerende Techniek (DCT) het advies uit: *Criteria voor de Beoordeling van het Ontwerpen in de Construerende Disciplines* [DCT, 2000]. In 2009 verscheen het advies *Kwaliteitsbeoordeling Ontwerp- en Constructiedisciplines* [Schouten, 2009]. In deze adviezen lag de nadruk vooral op de beoordeling van ontwerpen als wetenschappelijke activiteit. Het onderhavige KNAW advies richt zich niet alleen op ontwerpende en construerende activiteiten maar ook op het daaraan gerelateerde wetenschappelijke onderzoek.

Toenemende urgentie

Het belang van een goed beoordelingskader voor kwaliteit in de ontwerpende en construerende disciplines is in de loop van de jaren alleen maar groter geworden. Ten eerste worden onderzoeksprojecten en tijdelijk wetenschappelijk personeel in toenemende mate alleen nog extern gefinancierd. De ontwerpende en construerende disciplines zijn daarbij in competitie met andere disciplines en moeten het opnemen tegen natuurwetenschappelijke disciplines die beter passen bij de bestaande kwaliteitscriteria. Ten tweede is kwaliteitszorg aan de universiteiten steeds belangrijker geworden (externe visitaties, interne *midterm reviews*) en daaraan worden consequenties verbonden door bestuurders. De criteria voor wetenschappelijke kwaliteit die daarbij gehanteerd worden zijn dan cruciaal. Dit probleem speelt overigens niet alleen bij de ontwerpende en construerende disciplines. In het algemeen geldt dat toegepast multidisciplinair onderzoek minder goed scoort op traditionele kwaliteitscriteria. Het gaat vaak om kleinere deelgebieden waarin multidisciplinaire tijdschriften met lagere impactfactoren een belangrijke rol spelen, waarbij het citatiegedrag meestal asymmetrisch is: er wordt in de multidisciplinaire tijdschriften veel verwezen naar artikelen in monodisciplinaire tijdschriften, maar omgekeerd veel minder.

In de volgende paragraaf wordt aangegeven welke knelpunten in de praktijk worden ervaren. De informatie is afkomstig uit de interviews die de commissie gehouden heeft met vooraanstaande wetenschappers in het veld (zie bijlage 1).

2.2 Ervaren knelpunten

Vergelijken van kwaliteit tussen vakgebieden problematisch

De meeste geïnterviewden in de technische wetenschappen vinden dat de ontwerpende en construerende wetenschappen niet goed te beoordelen zijn met de traditionele beoordeling op basis van wetenschappelijke publicaties in ISI-tijdschriften met hoge impactfactoren. De mate waarin men daar zelf last van ondervindt, bijvoorbeeld bij het verwerven van onderzoeksmiddelen, hangt af van het gebied waarin men werkt. De beoordeling van kwaliteit binnen een vakgebied door *peers* wordt daarentegen vrijwel nergens als een probleem ervaren. Onderling weet men goed wie de topwetenschappers onder de collega's zijn. De problemen ontstaan bij vergelijking tussen vakgebieden.

Ontbreken van een duidelijk loket voor ontwerpende en construerende disciplines

Het vergelijken van de kwaliteit van wetenschappelijke werk tussen vakgebieden gebeurt onder andere bij het in competitie verwerven van onderzoeksfinanciering. Als vakgebieden teveel van elkaar verschillen is dit problematisch. Onderzoekers in de ontwerpende en construerende wetenschappen ondervinden problemen omdat meer fundamenteel onderzoek meestal als wetenschappelijk hoogstaander wordt gezien. Het ontbreken van een duidelijk loket voor de ontwerpende en construerende wetenschappen wordt door veel geïnterviewden als gemis ervaren. De wetenschappers die aangeven geen problemen te hebben met de huidige manier van kwaliteitsbeoordeling, ontvangen vaak onderzoeksfinanciering uit het bedrijfsleven of uit meer specifieke subsidieprogramma's zoals 'Point-One' of de Innovatiegerichte Onderzoeksprogramma's (IOP).

De recente overheveling van 100 miljoen euro van de eerste naar de tweede geldstroom heeft het ontbreken van een loket bij NWO voor de ontwerpende en construerende wetenschappen extra urgent gemaakt. Op dit moment bestaan er volgens een aantal geïnterviewden geen gelijke kansen voor deze disciplines bij NWO.

Voor de ontwerpende en construerende disciplines is een extra complicatie dat op een aantal gebieden, bijvoorbeeld de weg- en waterbouw of de volkshuisvesting, intermediaire onderzoeksorganisaties zijn opgeheven, minder kennisintensief zijn geworden of geen onderzoek meer financieren. Hierdoor zijn er minder mogelijkheden voor derde geldstroomonderzoek, terwijl de kansen in tweede geldstroomprogramma's waarbij betrokkenheid van maatschappelijke *stakeholders* wordt geëist afnemen. Ook daardoor is de urgentie van een loket bij NWO toegenomen.

Grote verschillen in publicatiecultuur te weinig erkend

Geen van de geïnterviewden betwist het belang van publiceren als methode van kennisverspreiding en toetsing van resultaten door *peer review*. De grote nadruk op *peer reviewed*-artikelen in ISI-tijdschriften doet echter te weinig recht aan de grote verschillen in publicatiecultuur tussen vakgebieden. Binnen de informatica bijvoorbeeld hebben sommige *peer-reviewed conference proceedings* een hogere status dan publicaties in ISI-tijdschriften. Ook het aantal publicaties dat een onderzoeker jaarlijks kan produceren verschilt sterk per discipline. In de construerende disciplines gaat veel tijd zitten in het ontwerp en de constructie voordat daarover gepubliceerd kan worden. Diverse geïnterviewden rapporteren een steeds grotere druk om te publiceren in hoge impacttijdschriften als *Science* en *Nature* terwijl die tijdschriften door vakgenoten niet worden gezien als leidend voor hun vakgebied.

Geïnterviewden uit disciplines als industrieel ontwerpen en bouwkunde geven aan dat publicaties voor hen in toenemende mate van belang zijn als onderdeel van een strategie van verwetenschappelijking van hun vakgebied. Daarbij moet wel rekening worden gehouden met het feit dat de vorm en het medium waarin gepubliceerd wordt passend is voor de betreffende discipline.

Afnemende rol van ontwerpen en construeren in de technische disciplines

Ontwerpen en construeren zijn pijlers van de ingenieurswetenschappen en vormen een brug naar de praktijk. In een ontwerp of constructie wordt kennis uit verschillende disciplines geïntegreerd voor het realiseren van functies van een systeem ten behoeve van *real-life*-situaties. Veel geïnterviewden constateren dat de verwetenschappelijking in sommige technische wetenschappen (bijvoorbeeld civiele techniek) geleid heeft tot een 'monoscience-achtige cultuur' waarin het ontwerpen en construeren van technische systemen minder aandacht krijgen dan nodig is en er vaak een grote kloof met de praktijk is ontstaan. Dat is een ongewenste situatie.

Grotere afstand tussen universitaire bestuur en werkvloer

Veel geïnterviewden ervaren dat kwaliteitsbeoordeling van wetenschappelijk werk aan de universiteiten is verworden tot een bureaucratisch instrument. Vanwege de toegenomen afstand tussen werkvloer en bestuur is het moeilijk om op bestuurlijk niveau goed zicht te hebben op de kwaliteit van het onderzoek in een bepaalde discipline. Er is daarom behoefte aan 'objectieve' en bij voorkeur kwantitatieve informatie over de wetenschappelijke kwaliteit van verschillende groepen. Bibliometrische analyses of andere cijfermatige indicatoren worden dan gebruikt om een vergelijking te maken tussen die groepen. De ontwerp- en construerende wetenschappen zijn hierbij in het nadeel ten opzichte van de disciplines die beter passen bij de bestaande kwaliteitscriteria. Dat nadeel betreft niet alleen de financiering van die vakgebieden, maar zeker ook de wetenschappelijke reputatie en erkenning.

2.3 Evaluatie ervaren knelpunten en conclusies

Verschillen tussen vakgebieden

De commissie onderschrijft de opvatting dat het vergelijken van wetenschappelijke kwaliteit tussen verschillende vakgebieden zinloos is. Zo'n vergelijking veronderstelt dat er kwaliteitsindicatoren bestaan die voor de betreffende disciplines in gelijke mate gelden of zouden moeten gelden. Die aanname is onjuist. De onderzoekstradities, publicatiegewoontes, soorten activiteiten en producten als tijdschriftartikelen, *proceedings*, boeken, artefacten en ontwerpen zijn daarvoor te divers. Uitspraken waarin bijvoorbeeld de wetenschappelijke kwaliteit van personen in verschillende vakgebieden worden vergeleken op basis van ISI-tijdschriftartikelen en impactfactoren hebben daarom geen betekenis. Ook de veelgebruikte Hirsch-index biedt voor een vergelijking geen basis. Hooguit kan een vergelijking worden gemaakt tussen gegevens die voor de disciplines genormeerd zijn. Het is dan mogelijk om te zeggen dat wetenschapper A tot de top 5 procent behoort in discipline X en dat wetenschapper B tot de top 25 procent behoort in discipline Y.

Kwaliteit kan dus alleen binnen disciplines zinvol worden vergeleken. Een *one-size-fits-all*-benadering, hoe gewenst ook vanuit bestuurlijk oogpunt vanwege zijn eenvoud, doet geen recht aan significante disciplinaire verschillen en zal dus altijd bepaalde vakgebieden ten onrechte bevoordelen. Deze opvatting is in lijn met eerdere conclusies van de KNAW commissie kwaliteitszorg [KNAW-commissie kwaliteitszorg, 2008].

Peer review

De beoordeling van wetenschappelijke kwaliteit binnen disciplines gebeurt meestal door *peers*. *Peeroordelen*, hoe zorgvuldig en genuanceerd zij ook tot stand komen, kunnen echter nooit volledig objectief zijn en recht kan doen aan alle verschillen en controverses binnen een discipline. Er is sprake van een niet te vermijden persoonlijke *bias* en er spelen altijd subjectieve elementen mee. Verschillende *peers* zullen daardoor tot gedeeltelijk andere oordelen kunnen komen. Ook bestaat er een risico dat een klein aantal *peers* een discipline domineren en dat eerdere beoordelingen het latere oordeel vertekenen [KNAW, 2008]. Ondanks deze beperkingen vindt de commissie dat *peer review* de beste manier van het beoordelen van kwaliteit is. Voorwaarde is wel dat een vakgebied voldoende homogeen is en voldoende (internationale) omvang heeft om onafhankelijke *peeroordelen* mogelijk te maken. Als hieraan is voldaan leveren kwaliteitsoordelen door *peers* óók binnen de ontwerpende en construerende wetenschappen betrekkelijk weinig problemen op.

Bibliometrisch onderzoek

De beperkingen van *peer review* hebben mede geleid tot het doen van bibliometrisch onderzoek (naast de hierboven genoemde groter wordende afstand tussen bestuur

en werkvloer). Het is echter een illusie om te denken dat een volledig kwantitatieve/ bibliometrische benadering van wetenschappelijke kwaliteit mogelijk is, of *peer review* zou kunnen vervangen. Niet in de laatste plaats omdat *peers* in het vakgebied nodig zijn om de bibliometrische gegevens over dat vakgebied te kunnen beoordelen en duiden. Bibliometrisch onderzoek kan naar de mening van de commissie hooguit aanvullend zijn op *peer review*. Beide benaderingen hebben hun beperkingen, en het is belangrijk om deze beperkingen voor ogen te houden. Kwaliteitsoordelen moeten daarom altijd met voorzichtigheid gehanteerd worden.

Momentopname

Een extra reden voor voorzichtigheid is dat kwaliteitsoordelen altijd momentopnames zijn en dat de kwaliteit van mensen en hun werk in de loop van de tijd kan variëren. Beoordelingscommissies moeten zich hiervan bewust zijn.

Publicaties ook voor ontwerpende en construerende wetenschappen van belang

De commissie hecht veel waarde het publiceren van wetenschappelijke resultaten in *peer-reviewed* tijdschriften, ook voor vakgebieden als industrieel ontwerpen en bouwkunde. Ten eerste omdat voorafgaande aan een publicatie toetsing door *peers* plaatsvindt van de resultaten van wetenschappelijk werk. Ten tweede omdat door publicaties de kennis die opgebouwd is wordt ontsloten en verspreid. De publicatievormen moeten wel passen bij het vakgebied. Een eenzijdige focus op publicaties in monodisciplinaire tijdschriften met een hoge impactfactor is niet wenselijk, zeker niet voor een technische universiteit waar onderzoek met een multidisciplinair en toepassingsgericht karakter ook belangrijk is. De commissie heeft met instemming kennisgenomen van het toenemende belang dat wetenschappers in vakgebieden als industrieel ontwerpen en bouwkunde toekennen aan *peer-reviewed* publicaties als onderdeel van de wetenschappelijke output.

Criteria sturend bij richting onderzoek

De criteria waarmee de kwaliteit van wetenschappelijk werk wordt beoordeeld spelen zelf weer een rol bij beslissingen over welke typen van activiteiten worden ontplooid aan de universiteiten. Zij werpen hun schaduw vooruit omdat onderzoekers anticiperen op de criteria waarmee hun wetenschappelijk werk beoordeeld zal gaan worden en op basis waarvan ook de financiering van hun werk zal plaatsvinden. Kwaliteitscriteria zijn daarom niet neutraal maar normerend voor een vakgebied. Naast hun rol in het ex postbeoordelen van wetenschappelijk werk spelen ze ook een ex antero rol bij de sturing van het onderzoek. In dit verband onderschrijft de commissie de zorg van een aantal geïnterviewden dat een eenzijdige focus op criteria die passen bij

natuurwetenschappelijk onderzoek in sommige technische disciplines ertoe geleid heeft dat ontwerpen en construeren en het daaraan gerelateerde multidisciplinaire onderzoek onder druk zijn komen te staan. Het is gezien de missie van een technische universiteit van groot belang om een goed evenwicht te bereiken tussen de verschillende soorten onderzoek die eerder onderscheiden zijn (paragraaf 2.1, tabel 2.1) en de ontwerpende en construerende activiteiten. Dat kan alleen maar lukken als de kwaliteitscriteria die gebruikt worden daarop zijn toegesneden.

Onderzoeksfinanciers

Ook voor onderzoeksfinanciers zoals NWO en STW geldt dat een *one-size-fits-all*-benadering geen recht doet aan relevante disciplinaire verschillen en dus altijd bepaalde vakgebieden ten onrechte zal bevoordelen. Verschillende geïnterviewden vinden dat er bij NWO geen gelijke kansen bestaan voor de ontwerpende en construerende disciplines. Meer in het algemeen geldt dat binnen de structuur van NWO multi- of interdisciplinair onderzoek niet vanzelfsprekend past en moet concurreren met disciplinair onderzoek dat veel beter uit de voeten kan met de gehanteerde kwaliteitscriteria. Ook zijn er gevestigde belangen in het geding waartegen nieuwe vakgebieden het moeten opnemen. De commissie is van mening dat NWO het probleem onderkend heeft met het initiëren van gebiedsoverstijgende programma's. Dit is een lijn die verder versterkt moet worden, bijvoorbeeld door ook programma's te starten tussen gebieden van NWO en STW. Wat STW betreft hebben verschillende geïnterviewden aangegeven dat de maatschappelijke relevantie van het onderzoek te beperkt wordt geïdentificeerd met kortetermijninteresse van bedrijven in gebruikerscommissies. Het zou volgens de commissie goed zijn ook hier een meer gedifferentieerde benadering te kiezen en ook onderzoeksprojecten mogelijk te maken die een relevantie op de langere termijn hebben, ook als daarvoor nog geen bedrijven kunnen worden gevonden. Dat is ook voor de ontwerpende en construerende disciplines van belang.

Transparante doelstelling van belang

Ten slotte stelt de commissie vast dat kwaliteitscriteria moeten zijn afgestemd op het doel dat een kwaliteitsoordeel dient. De beoordeling van een onderzoeksvoorstel is een andere beoordeling dan die van de *past performance* van een onderzoeksgroep of van het functioneren van een medewerker bij bevordering of vaste aanstelling.

CONCLUSIE 2.1

Wetenschappelijke kwaliteit kan alleen binnen disciplines zinvol worden vergeleken. Een *one-size-fits-all*-benadering bij de beoordeling van kwaliteit, hoe gewenst ook vanuit bestuurlijk oogpunt vanwege zijn eenvoud, doet geen recht aan significante disciplinaire verschillen en zal dus altijd bepaalde vakgebieden ten onrechte bevoordelen.

CONCLUSIE 2.2

Binnen de ontwerpende en construerende disciplines zijn er betrekkelijk weinig problemen met het vaststellen van de kwaliteit van wetenschappelijk werk of het aanwijken van excellente wetenschappers door *peers*.

CONCLUSIE 2.3

Voor technische universiteiten is het van belang dat de criteria voor wetenschappelijke kwaliteit toegesneden zijn op de verschillende typen onderzoek en activiteiten die daar plaatsvinden. Een eenzijdige focus op criteria die ontleend zijn aan de natuurwetenschappen bemoeilijkt een goed evenwicht tussen fundamenteel onderzoek, toegepast onderzoek en ontwerpende en construerende activiteiten.

CONCLUSIE 2.4

Het publiceren van wetenschappelijke resultaten in *peer-reviewed* tijdschriften, boeken en *proceedings*, is ook voor de ontwerpende en construerende disciplines van groot belang. De publicatievorm moet wel passen bij de aard van het vakgebied.

3. BEOORDELINGSCRITERIA

3.1 Inleiding

In dit hoofdstuk worden criteria en indicatoren geformuleerd die volgens de commissie passend zijn voor de beoordeling van kwaliteit in de ontwerpende en construerende wetenschappen.

Terminologische opmerking

Kwaliteit is een waarderend begrip waarmee wordt aangegeven hoe goed iets is. Het is een generiek begrip dat van toepassing is op objecten, processen en personen. Voor die verschillende situaties gelden verschillende beoordelingscriteria. De kwaliteit van een publicatie bijvoorbeeld wordt anders bepaald dan de kwaliteit van de computer waarop deze is gemaakt. Uitspraken over kwaliteit zeggen niet zoveel als daarbij niet wordt aangegeven op welke criteria deze kwaliteitsoordelen gebaseerd zijn. Hoe goed iets voldoet aan een beoordelingscriterium is meestal niet direct waarneembaar. Daarom wordt gewerkt met indicatoren. Deze geven empirische informatie over de mate waarin een object, proces of persoon voldoet aan een bepaald criterium. Dat hoeft niet per se kwantitatieve informatie te zijn, zoals bijvoorbeeld het aantal publicaties, het kan ook gaan om meer kwalitatieve informatie, zoals bijvoorbeeld het oordeel van collega's over de kwaliteit van het tijdschrift waarin die publicaties staan.

Uitgangspunten bij de beoordeling

Rekening houden met disciplinaire verschillen

Zoals eerder besproken zijn er significante verschillen tussen disciplines. Maar ook binnen het relatief beperkte gebied van de ontwerpende en construerende wetenschappen is de variatie groot. In brede kring wordt gedeeld dat de huidige praktijk van

kwaliteitsbeoordeling deze diversiteit onvoldoende in acht neemt. Met deze diversiteit moet naar mening van de commissie rekening worden gehouden in de beoordelings-systematiek.

Differentiatie noodzakelijk...

Gegeven de verschillen tussen disciplines en gegeven de verschillen tussen beoordelings-situaties moet volgens de commissie differentiatie een uitgangspunt zijn bij het beoordelen van wetenschappelijke kwaliteit. Dit betreft:

- differentiatie naar (deel)vakgebied
- differentiatie naar type activiteit binnen een vakgebied: van onderzoeken tot ontwerpen en construeren
- differentiatie naar het object van beoordeling: een medewerker, een onderzoeksvoorstel, of een onderzoeksgroep

...maar overkoepelend raamwerk nodig

Aan de andere kant meent de commissie dat de beoordelingsmaatstaven voor een bepaalde (deel)discipline niet volledig verschillend kunnen zijn van die van andere disciplines. Dat zou deze maatstaven arbitrair maken en opportunistisch, terwijl ook de indruk zou kunnen ontstaan dat aan de ontwerpende en construerende disciplines 'lichtere' eisen worden gesteld dan aan andere disciplines. Een overkoepelend beoordelingskader is daarom gewenst waarbinnen de genoemde diversiteit een plaats moet krijgen. De commissie heeft daarom gestreefd naar overkoepelende criteria voor de beoordeling van kwaliteit die ook breder van toepassing zijn, met indicatoren die per discipline kunnen verschillen of een ander belang kunnen hebben.

Peer review centraal

In het vorige hoofdstuk is al aan de orde gekomen dat volgens de commissie *peer review* het uitgangspunt moet zijn bij de beoordeling van de kwaliteit van publicaties, onderzoekers, onderzoeksgroepen en onderzoeksvorstellen. Voorwaarde hiervoor is dat een discipline voldoende homogeen is, voldoende (internationale) omvang heeft en dat er voldoende goede *peers* beschikbaar zijn.

Bibliometrie ondersteunend en niet vervangend

Bibliometrische gegevens zijn naar het oordeel van de commissie ondersteunend voor kwaliteitsoordelen van *peers* en niet vervangend. Dit kwam eerder aan de orde. Voor het bepalen van de waarde van bibliometrische indicatoren (citatiescores, impactfactoren, H-index, etc.) is de definitie van wat als een vakgebied wordt beschouwd cruciaal en geen triviale aangelegenheid.

Transparante doelstellingen

Kwaliteitsbeoordelingen zoals onderzoeksvisitaties worden in praktijk vaak gebruikt als *ranking*- en pr-instrument, niet als instrument om mogelijke verbeteringen te

identificeren. Het doel van kwaliteitsbeoordelingen moet voorafgaand aan de beoordeling duidelijk worden gemaakt en de beoordeling moet ook in het licht van deze doelstelling worden gezien.

3.2 Beoordelingscriteria

Van vijf naar slechts twee criteria

In eerste instantie heeft de commissie vijf beoordelingscriteria overwogen, aansluitend bij de benadering die de Royal Academy of Engineering in het Verenigd Koninkrijk heeft voorgesteld voor de beoordeling van kwaliteit in de technische disciplines. Deze criteria waren: 1) publicaties, 2) impact, 3) innovativiteit, 4) betrokkenheid van externe *stakeholders* en 5) reputatie van de betrokken wetenschappers. Voor deze criteria was tijdens de consultatieronde van de commissie met het veld in Nederland veel steun. Bij nadere uitwerking bleken sommige van deze criteria echter indicatoren te zijn die beter bij andere criteria konden worden ondergebracht. Publicaties in gerenommeerde tijdschriften vormen bijvoorbeeld geen zelfstandig criterium maar zijn een indicator voor het criterium wetenschappelijke kwaliteit. Ook de betrokkenheid van externe *stakeholders* bij het onderzoek is geen criterium maar een indicator voor de maatschappelijke relevantie van het betreffende onderzoek. Over innovativiteit als zelfstandig criterium heeft de commissie lang gearzeld. Uiteindelijk is besloten dat innovativiteit twee aspecten combineert en daarom niet zelf een criterium is: wetenschappelijke vernieuwing en maatschappelijke toepasbaarheid. Het eerste aspect komt in wetenschappelijke publicaties tot uitdrukking, het tweede aspect kan door indicatoren voor maatschappelijke relevantie worden weergegeven. De reputatie van wetenschappers ten slotte is naar de mening van de commissie een afgeleide grootte die zowel wetenschappelijke als maatschappelijke aspecten omvat.

De commissie heeft uiteindelijk geconcludeerd dat er in de kern van de zaak maar twee beoordelingscriteria zijn voor de kwaliteit van activiteiten in de ontwerpende en construerende wetenschappen:

1. wetenschappelijke kwaliteit
2. maatschappelijke relevantie

Deze conclusie is opmerkelijk omdat de voorgestelde criteria niet afwijken van de reeds bestaande criteria voor andere wetenschapsgebieden. Anders dan in eerste instantie gedacht, wordt er dus géén aparte set van criteria voor kwaliteit in de ontwerpende en construerende wetenschappen voorgesteld. De bovenstaande twee criteria worden breed gedragen in het wetenschappelijke veld en komen goed overeen met wat internationaal gezien gebruikelijk is (zie hoofdstuk 4). Zij lijken een overkoepevend raamwerk te bieden voor de beoordeling van kwaliteit in alle wetenschapsgebieden.

Gegeven dit overkoepelende raamwerk pleit de commissie vervolgens voor differentiatie ten aanzien van de indicatoren die gebruikt worden om te bepalen hoe goed iets aan de twee beoordelingscriteria voldoet:

1. differentiatie naar discipline: de aard en het relatieve belang van de gebruikte indicatoren zal in de verschillende wetenschapsgebieden verschillend zijn; de ontwerpende en construerende wetenschappen kennen bijvoorbeeld ook andere indicatoren voor wetenschappelijke kwaliteit dan aantallen en citaties in *peer-reviewed* tijdschriften.
2. differentiatie naar het type activiteit in een bepaalde discipline: onderzoeken, ontwerpen of construeren.
3. differentiatie naar het object van beoordeling: werk of persoon. Bij de beoordeling van een subsidieaanvraag voor een onderzoeksproject zijn vaak andere indicatoren relevant dan bij de beoordeling van een persoon voor een benoeming of promotie.

In tabel 3.1 zijn de twee beoordelingscriteria en de daarbij horende indicatoren weergegeven die volgens de commissie geschikt zijn om kwaliteit in de ontwerpende en construerende wetenschappen te beoordelen. Zoals gezegd zijn niet alle indicatoren even relevant of hebben ze hetzelfde relatieve gewicht voor elke (deel)discipline, activiteit of beoordelingssituatie. De tabel beoogt de belangrijkste typen van indicatoren weer te geven. Indicatoren voor de kwaliteit van wetenschappelijk onderwijs zijn niet in de tabel opgenomen en vallen buiten de opdracht van dit advies.

3.3 Toelichting op de indicatoren

In deze paragraaf worden de indicatoren in de tabel voor zover nodig nader toegelicht.

3.3.1 Criterium wetenschappelijke kwaliteit

Wetenschappelijke publicaties

Zoals eerder vermeld zijn wetenschappelijke publicaties ook voor de ontwerpende en construerende wetenschappen een belangrijke indicator voor wetenschappelijke kwaliteit. De vorm van deze publicaties kan echter verschillen voor de verschillende vakgebieden (bijv. tijdschriften, *proceedings*, elektronische publicaties).

Ontworpen artefacten

De commissie vindt dat bij de beoordeling van de kwaliteit van het wetenschappelijke werk in de construerende en ontwerpende wetenschappen ook ontworpen artefacten (producten, processen, software) een rol moeten kunnen spelen. Hierbij moeten wel enkele kanttekeningen worden geplaatst. Het gaat hierbij alleen om ontworpen artefacten die een wetenschappelijke bijdrage leveren aan het vakgebied, niet om artefacten die ontworpen zijn in het kader van het onderwijs. Essentieel voor een wetenschappelijk ontwerp is dat het nieuwe kennis oplevert die generiek toepasbaar is.

Tabel 3.1 kwaliteitsindicatoren voor de technische wetenschappen

	INDICATOREN VOOR WERK	INDICATOREN VOOR PERSOON
WETENSCHAPPELIJKE KWALITEIT	<p>Wetenschappelijke publicaties Artikelen in <i>peer-reviewed</i> tijdschriften (aantal en soort tijdschrift) Artikelen in <i>peer-reviewed conference proceedings</i> (aantal en soort proceedings) Wetenschappelijke boeken bij gerenommeerde uitgevers, of significante bijdragen daaraan (aantal en soort) Citaties van individuele artikelen Impactfactoren van tijdschriften waarin artikelen zijn gepubliceerd</p> <p>Ontworpen artefacten Peer-reviewed artefact (ontwerp) + documentatie. Hieronder valt ook het ontwerp van software</p> <p>Wetenschappelijke impact (ex post) Gebruik van wetenschappelijke producten door andere wetenschappers (artefacten, methoden, meetinstrumenten, tools, standaarden en protocollen)</p> <p>Potentiële wetenschappelijke impact Mogelijke bijdrage aan theorie- en modelvorming, methodeontwikkeling, werkingsprincipes of ontwerpconcepten</p>	<p>Ontvangen blijken van wetenschappelijke erkenning Ontvangen wetenschappelijke prijzen Lidmaatschappen van prestigieuze organisaties zoals academies van wetenschappen Prestigieuze subsidies zoals Veni-, Vidi-, Vici-beurzen, of ERC grants Eredocoraten Visiting professorships</p> <p>Redacteurschappen Chief/full editorschips internationaal wetenschappelijk tijdschrift/boek/conference proceedings</p> <p>Mate waarin men door peers wordt benaderd als expert Adviseurschappen in wetenschappelijke kring (NWO, visitaties, etc.) Keynote lezingen op wetenschappelijke congressen Lidmaatschappen programmacommissies Deelname internationale beoordelingscommissies van wetenschappelijke programma's/instituten of wetenschappelijke adviesraden/instituten</p> <p>Wetenschappelijke impact over gehele loopbaan Citatiescore van persoon Bijdrage aan wetenschappelijke schoolvorming</p>
MAATSCHAPPELIJKE RELEVANTIE	<p>Gebruik van resultaten door externe stakeholders (ex postimpact) Bijdrage aan oplossing maatschappelijke problemen Marktintroducties en nieuwe projecten in industrie Inkomsten door gebruik van resultaten Spin-offs met industrie Gebruikte patenten Gebruikte artefacten (ontwerpen, software)</p> <p>Gebruik van resultaten door professe (ex post impact) Gebruik van artefacten, methoden, meetinstrumenten, tools, standaarden en protocollen</p> <p>Betrokkenheid van externe stakeholders bij wetenschappelijk werk (potentiële maatschappelijke impact) Bedrijven of maatschappelijke organisaties betrokken bij de begeleiding van onderzoeksprojecten (bijvoorbeeld in gebruikerscommissies) Derde geldstroomfinanciering door mogelijke gebruikers (bijv. industrie) Publieke financiering gerelateerd aan maatschappelijke kwesties Valorisation grants</p> <p>Bijdrage aan verspreiding van kennis Professionele publicaties en voordrachten, niet-wetenschappelijke publicaties, tentoonstelling etc. over onderzoeksresultaten</p>	<p>Mate waarin men door externe stakeholders wordt gezien als expert Advies- en consultancywerkzaamheden (gericht op gebruikers) Leidende positie in industrieel onderzoek (bijv. directeur R&D-afdeling)</p> <p>Mate waarin men binnen de professie gezien wordt als expert Oeuvreprijzen (bijvoorbeeld architecten) Overzichtstentoonstellingen</p> <p>Bijdrage aan verspreiding van kennis Activiteiten gericht op popularisering wetenschap, onderwijs en bijdragen aan het publieke debat Training van professionals PhD's met eerste baan in relevante praktijk</p>

De documentatie bij het artefact moet hierover inzicht geven. Als een ontwerp nieuwe kennis oplevert dan is deze kennis meestal ook te publiceren in wetenschappelijke tijdschriften, waardoor kwaliteitsborging via *peer review* gegarandeerd wordt. Dat deze kennis in beginsel publiceerbaar is, betekent nog niet dat hij ook altijd gepubliceerd is. Daarvoor kunnen allerlei goede redenen zijn, die bijvoorbeeld te maken kunnen hebben met de bedrijfsmatige omgeving waarin deze kennis tot stand is gekomen en waarin publiceren niet gebruikelijk of zelfs niet toegestaan is.

Artefacten kunnen als indicator voor wetenschappelijke kwaliteit een rol spelen bij de benoeming van hoogleraren die een tijd in het bedrijfsleven hebben gewerkt. Gezien het belang van dit soort kandidaten voor de technische universiteiten moeten deze wetenschappelijke prestaties mee kunnen wegen. Het is echter essentieel dat in zo'n geval een oordeel van *peers* wordt gevraagd over de wetenschappelijke merites van een (beperkt) aantal goed gedocumenteerde ontworpen artefacten.

Wetenschappelijke impact

Onder wetenschappelijke impact wordt verstaan de mate waarin het wetenschappelijke werk gebruikt wordt door andere wetenschappers. Een van de meest voorkomende vormen van gebruik is het citeren van wetenschappelijk werk. Een andere indicator is het gebruik van nieuw ontwikkelde wetenschappelijke producten door collega wetenschappers. Met wetenschappelijk producten wordt hier bedoeld artefacten, methoden, meetinstrumenten, hulpmiddelen, standaarden, protocollen etc. Het gebruik van wetenschappelijke resultaten door niet-wetenschappers valt niet onder dit criterium maar onder het criterium maatschappelijke relevantie.

Potentiële wetenschappelijke impact

Deze indicator speelt een rol bij de beoordeling van de kwaliteit van onderzoeksvoorstellen. Deze worden mede beoordeeld op hun mogelijke bijdrage aan het vakgebied.

Ontvangen blijken van wetenschappelijke erkenning

Blijken van erkenning zijn een belangrijke graadmeter voor de wetenschappelijke kwaliteit van een persoon. Voorwaarde hiervoor is dat het blijken van erkenning zijn die door *peer review* tot stand zijn gekomen. Voorbeelden zijn Spinozapremies, eredoctoraten, lidmaatschappen van prestigieuze organisaties als academies van wetenschappen en prestigieuze subsidies zoals *Vici's* of ERC Grants.

Mate waarin men door peers wordt benaderd als expert

De mate waarin een wetenschapper door collega's wordt gezien als expert is een belangrijke indicator voor kwaliteit omdat binnen een vakgebied meestal goed duidelijk is wie de topwetenschappers zijn. Dit kan bijvoorbeeld blijken uit *keynote* lezingen op wetenschappelijke congressen, lidmaatschappen van programmacommissies en uit adviseurschappen in wetenschappelijke kring.

Wetenschappelijke impact over gehele loopbaan

Hieronder wordt verstaan de wetenschappelijke *impact* die een persoon heeft gehad gedurende zijn of haar loopbaan. Dit kan bijvoorbeeld blijken uit de citatiescore van een persoon. Ook het oordeel van *peers* over de mate waarin een wetenschapper heeft bijgedragen aan wetenschappelijke schoolvorming kan hier worden meegewogen.

3.3.2 Criterium maatschappelijke relevantie

Gebruik van resultaten

Hieronder vallen de indicatoren die een maat zijn voor het gebruik van de wetenschappelijke kennis of producten buiten het wetenschappelijke veld. Het daadwerkelijke gebruik is meestal alleen te beoordelen na afronding van het onderzoek. Dat gebruik kan vele vormen aannemen: van de bijdrage aan een maatschappelijk probleem tot het commerciële gebruik in een bedrijf. Een aparte categorie betreft het gebruik van resultaten door de beoefenaars van een professie in de maatschappij.

Potentiële impact

De maatschappelijke relevantie van onderzoek kan op verschillende tijdschalen beoordeeld worden. Toegepast onderzoek heeft een kortere tijdshorizon dan fundamenteel onderzoek. Voor het langere termijnonderzoek is het echter notoir moeilijk om de maatschappelijke relevantie vast te stellen. Hier kan gewerkt worden met indicatoren die de mate van interesse van maatschappelijke *stakeholders* in het onderzoek uitdrukken. Die *stakeholders* zijn overigens niet alleen bedrijven, maar ook ministeries of internationale organisaties. Voor het kortere termijnonderzoek is de betrokkenheid van *stakeholders* door co-financiering van het onderzoek een indicator. Deze betrokkenheid kan bovendien de kans op daadwerkelijk gebruik van de resultaten sterk vergroten.

Bijdrage aan verspreiding van kennis

Zowel bij de beoordeling van wetenschappelijk werk als bij de beoordeling van een persoon moet de bijdrage aan de verspreiding van kennis worden meegewogen als onderdeel van maatschappelijke relevantie. Kennisverspreiding draagt immers bij aan de oplossing van maatschappelijke problemen of het creëren van economische kansen. Dit is ook een onderdeel van de missie van een technische universiteit.

3.4 Het relatieve belang van de criteria en de indicatoren

Beoordelen is maatwerk en voor elke beoordelingssituatie moet voorafgaand aan de beoordeling worden bepaald welke van de twee criteria en welke van de indicatoren daarvoor het belangrijkste zijn. Voor een voltijds hoogleraar die een wetenschappelijke groep moet leiden, ligt dit anders dan voor een parttime hoogleraar die vooral de brug naar de praktijk moet slaan. Voor een persoonsgebonden Vidi-beurs ligt dit

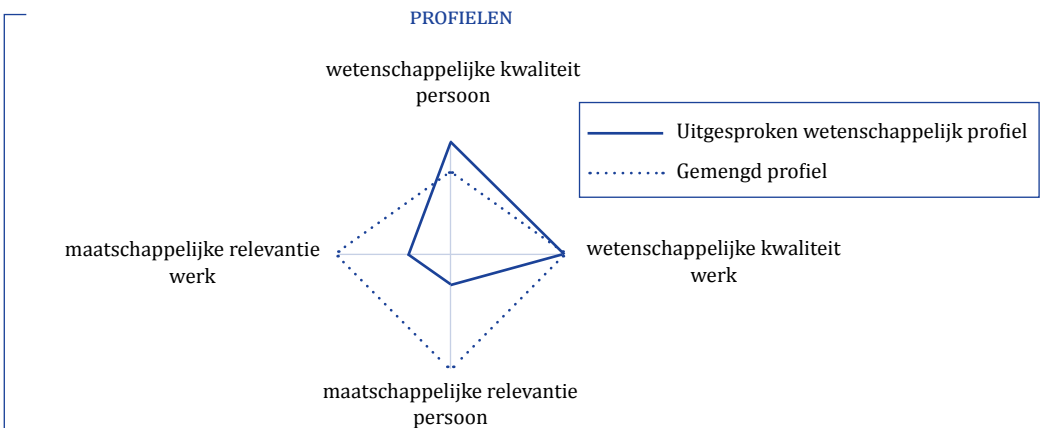
weer anders dan voor een aanvraag in de open competitie van NWO- of voor een STW-aanvraag. Om nog maar niet te spreken van de verschillen per vakgebied. Kortom, er bestaat niet één antwoord op de vraag hoe kwaliteitsbeoordeling moet plaats vinden.

De bijdrage van dit advies is dat er een systematisch raamwerk wordt voorgesteld voor het beoordelen van kwaliteit in de ontwerpende en construerende wetenschappen dat recht doet aan disciplinaire verschillen en ontworpen is voor een veelheid van beoordelingssituaties.

3.5 Profielen

De beoordeling van een persoon of onderzoeksgroep op basis van de bovenstaande tabel mag niet resulteren in een enkel cijfer – dat zou immers ontkennen dat er verschillende kwaliteitscriteria en indicatoren zijn. Het resultaat is een profiel. Zo'n profiel kan grafisch worden weergegeven. Het meest voor de hand liggend is een weergave waarbij een totaal oordeel wordt gegeven voor elk van de kwadranten van de tabel (in plaats van een oordeel voor elk van de indicatoren in dat kwadrant). Dat kan bijvoorbeeld gebeuren op een schaal van laag naar hoog (cijfers suggereren in dit geval een schijn-nauwkeurigheid). Maar een andere weergave kan ook wenselijk zijn.

Het gebruik van profielen maakt een discussie mogelijk over welke profielen wenselijk zijn. Dat kan gebeuren op verschillende aggregatieniveaus (personen, groepen). Die profielen kunnen per discipline en per onderzoeksgroep verschillen. Er kunnen ook verschillen zijn binnen een groep. Een parttime hoogleraar uit de industrie zal een ander profiel hebben dan een *fulltime* hoogleraar die leiding geeft aan de vakgroep. De bedoeling is wel dat een groep als geheel het gewenste profiel heeft. In figuur 3.1 is een voorbeeld opgenomen van twee profielen: een uitgesproken wetenschappelijk profiel en een gemengd profiel.



Figuur 3.1 Voorbeeld van een uitgesproken wetenschappelijk en een gemengd profiel.

3.6 Selectie van *peers*

Een kernprobleem bij *peer review* is het vaststellen wie de geschikte *peers* zijn. Dat is een probleem bij visitaties, maar ook bij de beoordeling van onderzoeksvoorstellen of het aanstellen van wetenschappelijk personeel. Met name in kleine landen en kleine vakgebieden waar iedereen elkaar kent, bestaat het risico dat *peer review* leidt tot vriendjespolitiek. Het is daarom wenselijk om op een meer objectieve manier vast te stellen wie geschikte *peers* zijn in een bepaalde situatie. Hierbij kan de voorgestelde tabel met criteria en indicatoren nuttig zijn. *Peers* kunnen namelijk worden gedefinieerd als die vakgenoten die goed scoren op de indicatoren die kenmerkend zijn voor een bepaalde discipline. Kandidaten voor beoordelingscommissies kunnen dan worden beoordeeld op basis van deze score. Daarnaast kan de onafhankelijkheid van *peers* worden bevorderd door ook internationale *peers* in commissies op te nemen.

3.7 Conclusies

CONCLUSIE 3.1

In de ontwerpde en construerende disciplines kan de kwaliteit van het wetenschappelijke werk worden beoordeeld aan de hand van twee criteria: 1) wetenschappelijke kwaliteit en 2) maatschappelijke relevantie. Een aparte set van criteria is voor deze disciplines niet nodig. Voor elk van de twee criteria bestaan er indicatoren die aangeven hoe goed men op de criteria scoort.

CONCLUSIE 3.2

Het beoordelen van kwaliteit is maatwerk vanwege de verschillen tussen disciplines, typen wetenschappelijke activiteiten (ontwerpen, onderzoeken) en beoordelings-situaties (visitatie, benoeming, onderzoeksvoorstel). Het voorliggende advies biedt een systematisch beoordelingskader dat recht kan doen aan deze verschillen door aan de indicatoren voor wetenschappelijke kwaliteit en maatschappelijke relevantie een bepaald belang of gewicht toe te kennen (tabel 3.1).

4. INVENTARISATIE VAN BUITENLANDSE ERVARINGEN EN INZICHTEN

4.1 Inleiding

Gezien het internationale karakter van wetenschappelijk onderzoek is het van belang dat de in het vorige hoofdstuk voorgestelde criteria in internationaal perspectief geloofwaardig zijn. Bovendien heeft de commissie gebruik willen maken van de ervaringen in het buitenland. Daarom is voor een aantal landen onderzocht hoe men tegen de kwaliteitsbeoordeling van wetenschappelijk onderzoek aankijkt, welke criteria men gebruikt en wat hiermee de ervaringen zijn. Waar mogelijk is specifiek gekeken naar de ontwerpende en construerende wetenschappen. In het Verenigd Koninkrijk (VK) en Australië zijn zeer uitgebreide systemen opgezet voor de landelijke onderzoeksbeoordeling. Daar is veel discussie geweest over geschikte criteria en is ook veel achtergrondonderzoek gedaan. De situatie in deze landen is daarom in detail bekeken. In Finland is gekeken naar de evaluatie van de Aalto Universiteit vanwege de degelijke aanpak en de grote aandacht voor de ontwerpende en construerende wetenschappen in deze evaluatie. Tot slot is van een aantal voor Nederland belangrijke landen kort nagegaan hoe die met kwaliteitsbeoordeling omgaan. De informatie over de individuele landen is te vinden in bijlage 3. In paragraaf 4.2 zijn de belangrijkste bevindingen en conclusies weergegeven.

4.2 Samenvatting ervaringen in het buitenland en conclusies

De meeste buitenlandse beoordelingssystemen hebben betrekking op alle wetenschapsgebieden en gaan meestal slechts zijdelings in op de specifieke problemen bij het beoordelen van de ontwerpende en construerende wetenschappen. Er zijn wel een

paar opvallende observaties die relevant zijn voor dit advies over de construerende en ontwerpende wetenschappen.

Gebruik van bibliometrische gegevens

In de beoordelingsmethodes van het VK, de Finse Aalto Universiteit en Australië wordt veel waarde gehecht aan bibliometrische analyses. Duidelijk is dat er veel discussie is in de betreffende landen over de toepasbaarheid van deze analyses, zeker bij de ontwerpende en construerende wetenschappen. Bij het Britse Research Excellence Framework lijkt men in eerste instantie zwaar op het gebruik van bibliometrische methoden te hebben ingezet maar is het belang ervan uiteindelijk afgezwakt. Op basis van de reacties uit de consultatierondes is men tot de conclusie gekomen dat deze indicatoren niet robuust genoeg zijn en onvoldoende draagvlak in het veld hebben. Ze zijn echter niet uit het beoordelingssysteem verdwenen, de analyses dienen nu als achtergrondinformatie voor *review*commissies. In het Australische systeem is nog wel steeds een prominente plek voor bibliometrie weggelegd. Wel heeft men daar differentiatie toegepast naar vakgebied door gedetailleerd vast te leggen voor welk gebied dit van toepassing is. Voor de ontwerpende en construerende wetenschappen is dit voor alle disciplines behalve voor de bouwkundige disciplines het geval. De *engineering* disciplines worden op vergelijkbare manier als de natuurkunde en chemie beoordeeld. In de Finse evaluatie geeft de evaluatiecommissie aan dat de bibliometrische analyses in het panel *Architecture, design, media en art research* niet bruikbaar waren omdat er te weinig was gepubliceerd in ISI-tijdschriften.

Mogelijkheid om niet-traditionele informatie mee te nemen

Zowel de evaluatie van de Finse Aalto Universiteit als de Australische ERA (*Excellence in Research for Australia*) bieden de mogelijkheid om wat men noemt 'niet traditionele output' in de beoordeling mee te nemen. Hieronder worden bijvoorbeeld verstaan films, websites, tentoonstellingen en creatief 3D-werk (architectuur, design, games, computerprogramma's). Ook hier lijkt deze uitzondering vooralsnog vooral betrekking te hebben op de disciplines architectuur/bouwkunde en industrieel ontwerpen. De beoordeling hiervan vindt plaats door *peer review*commissies die een aantal door de indieners zelf voorgelegde cases beoordelen. In het Australische systeem is daarbij expliciet beschreven dat de indieners zelf moeten aangeven waaruit de wetenschappelijke component van het werk bestaat. De ervaring met de evaluatie van de Aalto University in Finland leert dat hier nog wel een gewenningsproces nodig is. De evaluatiecommissie wees erop dat het bestuur weliswaar deze mogelijkheid heeft gecreëerd, maar er in de praktijk op aangedrongen is vooral wetenschappelijke artikelen aan te leveren.

Maatschappelijk *impact* internationaal gezien in de belangstelling

Naast de wetenschappelijke kwaliteit van de *output* laten alle besproken evaluatiesystemen maatschappelijke *impact* meewegen in de beoordeling. De manier waarop dat moet gebeuren is nog sterk in discussie.

Redelijk vergelijkbare criteria waarop wordt beoordeeld

Er is redelijke overeenstemming over de criteria waarop de kwaliteit van de wetenschappelijke *output* moet worden beoordeeld. Hoewel de exacte omschrijving verschilt, worden in essentie vrijwel altijd de criteria: Wetenschappelijke kwaliteit, Maatschappelijke relevantie en in wat mindere mate Productiviteit gebruikt.

Manier waarop *peer review*commissies worden samengesteld nergens beschreven

*Peer review*commissies spelen in de verschillende beoordelingsprocessen een cruciale rol. Dit betekent dat de kwaliteit van het beoordelingsproces in hoge mate afhangt van de kwaliteit van de *peer review*commissies. Zeker gezien de enorme breedte van het onderzoeksterrein waar sommige panels zich over moeten buigen, is het de vraag of de leden van de *peer review*commissies wel voldoende '*peers*' zijn voor het betreffende onderzoek. Opvallend is dat de meeste van de besproken beoordelingsprocedures niet ingaan op de manier waarop de *peer review*commissies samengesteld worden of de manier waarop de leden worden geselecteerd.

CONCLUSIE 4.1

In de beoordelingssystemen van toonaangevende landen zoals het Verenigd Koninkrijk, Australië, Verenigde Staten en Finland zijn wetenschappelijke kwaliteit en maatschappelijke relevantie de belangrijkste criteria.

CONCLUSIE 4.2

In verschillende buitenlandse beoordelingssystemen is het voor de ontwerpde en construerende wetenschappen mogelijk om resultaten zoals creatief werk, websites, producten in de beoordeling mee te laten wegen. Deze mogelijkheid is vaak wel beperkt tot de meer artistieke richtingen.

CONCLUSIE 4.3

In vrijwel alle beoordelingssystemen wordt veel aandacht besteed aan bibliometrische informatie. De waarde die eraan wordt gehecht verschilt.

CONCLUSIE 4.4

Het in dit advies voorgestelde beoordelingskader is geloofwaardig in internationaal verband.

5. CONCLUSIES EN AANBEVELINGEN

In hoofdstuk 2 is aangegeven welke knelpunten er zijn bij de huidige beoordeling van de kwaliteit van wetenschappelijk werk in de ontwerpende en construerende wetenschappen. In hoofdstuk 3 is een nieuw beoordelingskader geschetst dat tegemoet komt aan deze knelpunten. Dat kader is in hoofdstuk 4 geplaatst in een internationale context. In dit laatste hoofdstuk wordt de reikwijdte van het advies besproken en de gevolgen voor de bestaande praktijk. Tevens wordt aangegeven welke vervolgstappen nodig zijn. Ten slotte worden enkele aanbevelingen geformuleerd. Eerst worden nogmaals de belangrijkste conclusies hieronder weergegeven.

CONCLUSIE 2.1

Wetenschappelijke kwaliteit kan alleen binnen disciplines zinvol worden vergeleken. Een *one-size-fits-all*-benadering bij de beoordeling van kwaliteit, hoe gewenst ook vanuit bestuurlijk oogpunt vanwege zijn eenvoud, doet geen recht aan significante disciplineverschillen en zal dus altijd bepaalde vakgebieden ten onrechte bevoordelen.

CONCLUSIE 3.1

In de ontwerpende en construerende disciplines kan de kwaliteit van het wetenschappelijke werk worden beoordeeld aan de hand van twee criteria: 1) wetenschappelijke kwaliteit en 2) maatschappelijke relevantie. Een aparte set van criteria is voor deze disciplines niet nodig. Voor elk van de twee criteria bestaan er indicatoren die aangeven hoe goed men op de criteria scoort.

CONCLUSIE 3.2

Het beoordelen van kwaliteit is maatwerk vanwege de verschillen tussen disciplines, typen wetenschappelijke activiteiten (ontwerpen, onderzoeken) en beoordelings-situaties (visitatie, benoeming, onderzoeksvoorstel). Het voorliggende advies biedt een systematisch beoordelingskader dat recht kan doen aan deze verschillen door aan de indicatoren voor wetenschappelijke kwaliteit en maatschappelijke relevantie een bepaald belang of gewicht toe te kennen (tabel 3.1).

CONCLUSIE 4.4

Het in dit advies voorgestelde beoordelingskader is geloofwaardig in internationaal verband.

Reikwijdte advies

Conform de adviesaanvraag van de drie technische universiteiten aan de KNAW is het beoordelingskader in de eerste plaats bedoeld voor de ontwerpende en construerende wetenschappen. De commissie verwacht echter dat het ontwikkelde beoordelingskader ook bruikbaar is voor andere (technische) disciplines. Ten eerste omdat de twee criteria die onderscheiden worden (wetenschappelijke kwaliteit en maatschappelijke relevantie) goed aansluiten bij de bestaande praktijk van kwaliteitsbeoordeling (SEP, beoordeling van onderzoeksvoorstellen, etc.). Ten tweede omdat het voorgestelde kader zodanig flexibel is dat er recht kan worden gedaan aan relevante verschillen tussen disciplines. Die flexibiliteit komt van de indicatoren die gebruikt worden om te bepalen hoe goed wetenschappelijk werk aan de twee kwaliteitscriteria voldoet. Deze indicatoren hoeven niet voor elk wetenschapsgebied identiek of even belangrijk te zijn. Er kunnen zelfs discipline-specifieke indicatoren worden toegevoegd. Het voorgestelde kader is daarom niet alleen bruikbaar voor de ontwerpende en construerende disciplines maar bijvoorbeeld ook voor meer fundamentele (natuur)wetenschappen. De daarvoor veel gebruikte kwaliteitsindicatoren (hoge impacttijdschriftpublicaties, etc.) vormen immers een deelverzameling van de indicatoren in tabel 3.1. Het ontwikkelde kader is dus generiek, maar om operationele waarde te krijgen moet het discipline-specifiek worden gemaakt door de keuze van de kwaliteitsindicatoren en hun relatieve belang. Wetenschappers uit de meer fundamentele disciplines kunnen bijvoorbeeld aangeven dat de kwaliteit van hun werk onder andere met de indicatoren hoge impacttijdschriftpublicaties bepaald kan worden.

Inpasbaarheid in bestaande praktijk

Er zijn verschillende situaties waarbij uitspraken moeten worden gedaan over kwaliteit: de beoordeling van een onderzoeksvoorstel, de beoordeling van een onderzoeksgroep en de beoordeling van een medewerker bij bevordering of aanstelling. De commissie is van mening dat het voorgestelde advies betrekkelijk eenvoudig in te passen

is in de bestaande beoordelingspraktijk. Ook het SEP-protocol dat gebruikt wordt bij visitaties is compatibel met het hier voorgestelde beoordelingskader.

De afwijzing van een *one-size-fits-all*-benadering bij de beoordeling van kwaliteit heeft als consequentie dat in de bestaande praktijk beleidsoordelen waarschijnlijk een explicietere rol gaan spelen bij het maken van keuzes in situaties waar een vergelijking tussen disciplines moet worden gemaakt. Bijvoorbeeld bij het in competitie verdelen van onderzoeksmiddelen door NWO in gebieden die heterogeen zijn, of bij de toekenning van wetenschappelijke prijzen (NWO-Spinozapremie). Deze keuzes kunnen niet louter op basis van een kwaliteitsoordeel gemaakt en gemotiveerd worden. Het vergelijken van kwaliteit tussen verschillende disciplines is immers zinloos omdat er appels met peren worden vergeleken (persoon p_1 of onderzoeksvoorstel o_1 scoort waarde w_1 op indicatorenset i_1 en persoon p_2 of onderzoeksvoorstel o_2 scoort waarde w_2 op indicatorenset i_2). Hooguit kunnen scores die genormeerd zijn voor de betreffende disciplines met elkaar worden vergeleken (persoon p_1 behoort tot de top-5 procent van scorende wetenschappers op indicatorenset i_1 in discipline d_1 terwijl persoon p_2 behoort tot de top-25 procent scorende wetenschappers op indicatorenset i_2 in discipline d_2). Maar zelfs dan is de betekenis hiervan niet evident.

Adviesaanvraag 3TU

De commissie juicht het toe dat de Colleges van Bestuur van de 3TU's de positie van ontwerpende en construerende activiteiten en het daaraan gerelateerde onderzoek aan de orde hebben gesteld. Deze activiteiten staan onder druk door een beoordelingskader voor kwaliteit dat daarop niet voldoende is toegesneden. Hierdoor dreigt in sommige gebieden het evenwicht tussen fundamenteel onderzoek, toegepast onderzoek en ontwerpende en construerende activiteiten verstoord te worden. Dit is gezien de missie van de technische universiteiten een ongewenste situatie. Als de drie Colleges van Bestuur het voorliggende advies overnemen en overbrengen aan hun wetenschappelijke staf dan wordt daarmee ook het signaal afgegeven dat voor technische universiteiten kwalitatief hoogwaardige ontwerpende en construerende activiteiten van belang zijn en een onlosmakelijk deel van de technische wetenschappen uitmaken.

De eerste stap voor de implementatie van het advies is dat de ontwerpende en construerende wetenschappen zelf bepalen welke kwaliteitsindicatoren uit de tabel op hen van toepassing zijn en wat het relatieve belang daarvan is. Deze indicatoren zullen in internationaal verband binnen de discipline geloofwaardig moeten zijn. De bepaling hiervan kan alleen door *peers* in die disciplines gebeuren. Het ligt voor de hand om de bestaande cycli van kwaliteitszorg hiervoor te gebruiken. De besturen van de 3TU's kunnen bijvoorbeeld aan een discipline vragen om – voorafgaande aan een visitatie – kwaliteitsindicatoren vast te stellen die de visitatiecommissie kan gebruiken bij de beoordeling van het onderzoek. Zo'n commissie, die uit internationale *peers* bestaat, kan ook worden gevraagd om over deze indicatoren een oordeel uit te spreken.

AANBEVELING 1

De commissie adviseert de besturen van de 3TU's om ruimte te scheppen voor een disciplinespecifieke kwaliteitsbeoordeling in de technische wetenschappen en aan de ontwerpende en construerende disciplines te vragen indicatoren en hun relatieve gewicht vast te stellen voor de beoordeling van kwaliteit in die disciplines. Deze indicatoren moeten geloofwaardig zijn in internationaal verband.

Onderzoeksfinanciers

De commissie is van mening dat de manier van kwaliteitsbeoordeling niet *a priori* bepaalde vakgebieden mag bevoor- of benadelen. Voor NWO/STW zou dit een kernpunt moeten zijn van de beoordelingssystematiek. Het in dit advies ontwikkelde beoordelingskader kan hulpmiddel zijn bij het vaststellen van kwaliteitsindicatoren die recht doen aan disciplinaire verschillen.

Met de introductie van gebiedsoverstijgende programma's toont NWO zich in ieder geval bewust van het belang van programma's die niet binnen de bestaande kaders passen en waarvoor ook andere kwaliteitsindicatoren kunnen gelden. Deze lijn moet volgens de commissie verder worden versterkt, bijvoorbeeld door ook programma's tussen gebieden van NWO en STW te starten. Dat is voor de ontwerpende en construerende disciplines, die een loket missen bij NWO, van groot belang. De commissie raadt STW aan om een meer gedifferentieerde benadering te kiezen ten aanzien van de maatschappelijke relevantie van onderzoeksprojecten. Ook projecten waarvoor bedrijven vanwege de langere termijn nog geen directe interesse tonen zouden voor financiering in aanmerking moeten kunnen komen. De ontwerpende en construerende disciplines hebben hier belang bij omdat op een aantal gebieden intermediaire onderzoeksorganisaties zijn opgeheven (zie hoofdstuk 2).

De commissie juicht het toe dat de meeste onderzoeksfinanciers gebruik maken van *peer review*. De keuze wie deze *peers* zijn is wel cruciaal en heeft grote invloed op de uitkomst van de beoordeling. De speelruimte van beoordelingscommissies om af te wijken van *refereerapporten* is namelijk klein. Het zou goed zijn om de keuze van *peers* meer te objectiveren en transparanter te maken. Dat komt de kwaliteit van het *peer review* proces ten goede. Tabel 3.1 kan daarbij van nut zijn. De *peers* die betrokken zijn bij de beoordeling van kwaliteit in de ontwerpende en construerende wetenschappen moeten zelf ook worden beoordeeld aan de hand van het brede spectrum van indicatoren en niet alleen op publicatiegedrag.

AANBEVELING 2

De commissie adviseert NWO en STW om meer in te zetten op programma's voor disciplines die niet goed passen bij de bestaande gebiedsindeling en de bestaande manier van kwaliteitsbeoordeling. Dat geldt in het bijzonder de ontwerpende en construerende wetenschappen. Daarvoor moeten kwaliteitsindicatoren worden gehanteerd die recht doen aan deze disciplines. De commissie adviseert STW om ook projecten mogelijk te maken met een maatschappelijke relevantie op de langere termijn waarvoor bedrijven nog geen concrete belangstelling hebben.

AANBEVELING 3

De commissie adviseert de universiteiten en NWO/STW om de keuze van *peers* transparanter en objectiever te maken. Dat komt de kwaliteit van het *peer review* proces ten goede. Het in dit advies ontwikkelde beoordelingskader met discipline-specifieke kwaliteitsindicatoren kan daarbij een rol spelen. De keuze van de juiste *peers* is van belang voor de samenstelling van visitatiecommissies en beoordelingscommissies, alsook voor het *referee* proces van onderzoeksvorstellen.

Wetenschappers in de ontwerpende en construerende wetenschappen

Een op de ontwerpende en construerende wetenschappen toegesneden wijze van kwaliteitsbeoordeling vraagt om een actieve inzet van de wetenschappers in deze disciplines. Die is nodig voor het vaststellen, wegen en operationaliseren van kwaliteitsindicatoren. Daarnaast zullen zij actief binnen NWO en STW moeten participeren als *peers* bij de beoordeling van onderzoeksvorstellen. Het is de commissie opgevallen dat dit maar op beperkte schaal gebeurt.

Het in dit advies geschetste beoordelingskader maakt het gebruik van andere indicatoren voor wetenschappelijke kwaliteit dan *peer-reviewed* publicaties mogelijk. Dit neemt niet weg dat de commissie en veel geïnterviewden vinden dat *peer-reviewed* publicaties ook in de ontwerpende en construerende disciplines belangrijk zijn en blijven. Ze dragen bij aan onderlinge kwaliteitstoetsing en aan verspreiding van kennis. Publiceren is in deze disciplines of onderdelen daarvan niet vanzelfsprekend. Het is van belang om daar te zoeken naar publicatievormen die passend zijn bij de (deel) discipline. Daarbij kunnen nieuwe media een rol spelen.

AANBEVELING 4

De commissie adviseert wetenschappers in de ontwerpende en construerende disciplines om daar waar nodig krachtig een cultuur te bevorderen van *peer-reviewed* publicaties. Die dragen bij aan de toetsing van resultaten, de verspreiding van inzichten en de verdere verwetenschappelijking van de discipline.

LITERATUUR

- 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, K. van, 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 en NWO (2009), *Standard Evaluation Protocol 2009-2015,* beschikbaar op www.knaw.nl/sep.

BIJLAGE 1

GEÏNTERVIEWDEN

Prof. dr. P.M.G. Apers, Universiteit Twente, informatica
Prof. dr. ir. J.A. Battjes, TU Delft, civiele techniek
Prof. ir. A. Beukers, TU Delft, lucht-en ruimtevaartechiek
Prof. dr. C.J.P.M. de Bont, TU Delft, industrieel ontwerpen
Prof. dr. ir. A.C.J.M. Eekhout, TU Delft, bouwkunde
Prof. dr. K.M. van Hee, TU Eindhoven, wiskunde en informatica
Prof. dr. ir. J.J. Heijnen, TU Delft, bioprocestechnologie
Dr. ir. F.D. van der Hoeven, TU Delft, bouwkunde
Prof. dr. ir. F.J.A.M. van Houten, Universiteit Twente, werktuigbouwkunde /
industrieel ontwerpen
Prof. dr. ir. J.T.F. Keurentjes, TU Eindhoven, scheikundige technologie
Prof. ir. H. Leegwater, TU Eindhoven, scheikundige technologie
Prof. dr. ir. M.C.M. van Loosdrecht, TU Delft, bioprocestechnologie
Prof. dr. ir. J.B.O.S. Martens, TU Eindhoven, industrieel ontwerpen
Prof. dr. ir. B. Nauta, Universiteit Twente, elektrotechniek
Prof. dr. C.J. Overbeeke, TU Eindhoven, industrieel ontwerpen
Prof. dr. ir. W.A. Poelman, Universiteit Twente, industrieel ontwerpen
Prof. ir. J.M. Post, TU Eindhoven, bouwkunde
Prof. dr. A.F.J. van Raan, Universiteit Leiden, kwantitatieve analyse
Prof. dr. ir. M.J.W. Schouten, TU Eindhoven, industrieel ontwerpen
Prof. dr. ir. M. Steinbuch, TU Eindhoven, werktuigbouwkunde
Dr. ir. S. Silvester, TU Delft, industrieel ontwerpen
Dr. P.E. Vermaas, TU Delft, filosofie
Dr. J. Visschers, NIKHEF, instrumentatieontwerp
Prof. dr. ir. H.J. de Vriend, Deltares, civiele techniek
Prof. dr. ir. P.J.V.V. van Wesemael, TU Eindhoven, bouwkunde
Prof. dr. ir. J.J. van Wijk, TU Eindhoven, informatica
Prof. dr. H.W. Zandbergen, TU Delft, technische natuurkunde

BIJLAGE 2

INDICATOREN ERiC-PROJECT

In het kader van het ERiC-project zijn een drietal *pilots* gehouden bij de faculteiten *electrical engineering* van de TU Eindhoven, bouwkunde van de TU Delft en werktuigbouwkunde van de Universiteit Twente. Deze *pilots* hadden tot doel een evaluatiemethode te ontwikkelen die gebruikt kan worden om de maatschappelijke relevantie van het onderzoek aan deze faculteiten te beoordelen. De voorgestelde indicatoren uit de *pilots* bouwkunde en *electrical engineering* zijn hieronder weergegeven.

Tabel BL.2.1 Indicatoren voor maatschappelijke relevantie van bouwkundig onderzoek [ERiC, 2010]

Aspect van maatschappelijke relevantie	Indicatoren
Verspreiding van kennis	Professionele publicaties, niet-wetenschappelijke publicaties, tentoonstellingen etc. Verspreiding van technologie, artefacten, standaarden Advies- en consultancywerkzaamheden Popularisering, onderwijs en bijdragen aan het publieke debat Training van professionals, mobiliteit van opgeleiden Masterscripties en afstudeeropdrachten die vragen uit de praktijk beantwoorden
Interesse van stakeholders	Aantal onderzoekers met relevante praktijkervaring in de sector(en) waar het onderzoeksprogramma zich op richt Publieke financiering gerelateerd aan maatschappelijke kwesties Derde geldstroomfinanciering door mogelijke gebruikers Samenwerking met maatschappelijke stakeholders in onderzoek, tests en evaluaties Consortia met niet-academische organisaties
Impact en gebruik van resultaten	Inkomsten door gebruik van resultaten Zichtbaarheid in publiek debat / publieke <i>mediarankings</i>

Tabel BL 2.2 Indicatoren voor maatschappelijke relevantie voor electrical engineering [ERiC, 2010]

Aspect van maatschappelijke relevantie	Indicatoren
Verspreiding van kennis	PhD's in industrie MA's in industrie <i>Proofs of concept</i> Presentaties op gespecialiseerde conferences
Interesse van <i>stakeholders</i>	Gezamenlijke <i>roadmaps</i> Presentaties op uitnodiging Valorisatiegrants Industriële financiering Uitwisseling van staf (uitwisselingsprogramma's zoals Casimir, kennisregeling) Parttime professors vanuit / in industrie Consortia met industrie
Impact en gebruik van resultaten	Marktintroductie en nieuwe projecten in industrie <i>Spin offs</i> met industriecontacten Patenten en octrooien

BIJLAGE 3

KWALITEITSBEOORDELING IN HET BUITENLAND

BL3.1 Verenigd Koninkrijk, *Research Excellence Framework*

De discussie in het Verenigd Koninkrijk (VK) over de kwaliteitsbeoordeling van wetenschappelijk onderzoek staat momenteel vooral in het teken van het *Research Excellence Framework* (REF). Het REF is een nieuw proces voor de beoordeling van wetenschappelijk onderzoek in het VK en is bedoeld als opvolger van de *Research Assessment Exercise* (RAE). De laatste RAE is gehouden in 2008. Een belangrijk bezwaar van de laatste RAE is dat het een veel te omvangrijke en arbeidsintensieve operatie is geworden. Het doel van de REF is om excellent onderzoek te identificeren en de *impact* daarvan te beoordelen. De beoordeling vindt plaats op het niveau van samenhangende groepen onderzoekers. Er worden geen individuele onderzoekers of complete instituten beoordeeld. De eerste beoordelingen volgens het REF zijn gepland in 2013 en zullen dan de verdeling van de financiering in 2014 bepalen. De manier waarop de REF exact zal worden uitgevoerd is nog onderwerp van discussie. In deze paragraaf worden de hoofdlijnen van de REF beschreven zoals die eind 2009 bekend waren [HEFCE, 2009]. Tevens wordt ingegaan op de belangrijkste discussiepunten waarbij gezien de focus van dit advies de nadruk wordt gelegd op de kritiek vanuit ingenieurswetenschappen.

Manier van beoordelen

In het voorstel voor de REF staat *peer review* centraal. Al het onderzoek dat ter beoordeling wordt voorgedragen wordt door *expert panels* beoordeeld op drie criteria:

1. *Output quality*. Het belangrijkste doel van de REF is om onderzoek te identificeren dat internationaal gezien als excellent kan worden betiteld. Hierbij mogen allerlei vormen van *output* ter beoordeling worden aangeboden.
2. *Impact*. Met dit criterium wordt beoordeeld in hoeverre wetenschappers aantoonbare bijdragen leveren aan de economie, maatschappij, beleidsontwikkeling, cultuur en kwaliteit van leven. Dit wil men gaan beoordelen door middel van *case-study's*.
3. *Environment*. Dit criterium is bedoeld om de kwaliteit van de onderzoeksomgeving mee te wegen. Hieronder wordt o.a. de *researchstrategie*, de ontwikkeling van de staf en training van onderzoekers verstaan.

Niet alle criteria wegen even zwaar. Voor het totaaloordeel denkt met aan de weging: *output quality*: 60 procent, *impact*: 25 procent en *environment*: 15 procent. Het totaaloordeel bestaat uit een sterrenstelsel oplopend van *unclassified* (minder dan goed of niet te beoordelen) tot *four stars (exceptional)*.

De indeling van het onderzoek in verschillende *review panels* is nog onderwerp van discussie. Men voorziet een tweetrapsstructuur met vier hoofdpanels met daaronder dertig subpanels. Het grootste deel van de technische wetenschappen zal in dit voorstel worden beoordeeld in één groot *engineeringpanel* waarin alle onderzoek zit op het terrein van elektrotechniek, chemische technologie, werktuigbouw en lucht- en ruimtevaart, mijnbouw, civiele techniek, materiaalkunde. Onder hetzelfde hoofdpanel maar onder een ander subpanel vallen dan: *computer science and informatics, mathematics, physics, chemistry en earth systems*. Opvallend is dat architectuur in het panel *Architecture, the built environment, town and countryplanning* is ondergebracht dat zelfs onder een ander hoofdpanel valt.

Discussiepunten

Gebruik bibliometrie

Er is veel discussie geweest over de bruikbaarheid van citatie-indicatoren en andere kwantitatieve indicatoren. Gekeken is of in de meer *science* georiënteerde gebieden deze indicatoren de *expert review* zouden kunnen vervangen. Uiteindelijk is men tot de conclusie gekomen dat deze indicatoren niet robuust genoeg zijn en ook onvoldoende draagvlak hebben in het wetenschappelijke veld. Desalniettemin wil men deze indicatoren een grotere rol laten spelen dan voorheen in de RAE. De bedoeling is nu dat de indicatoren worden gebruikt door de panels als hulpmiddel om hun oordeel op te baseren. Dit zou in ieder geval moeten gebeuren voor de panels in de gebieden medisch, biologie, de meer fysisch georiënteerde natuurwetenschappen, psychologie, engineering en informatica. Hoe dit precies moet gaan gebeuren wordt vooralsnog aan de panels overgelaten.

Een groot deel van de discussie spitst zich nog toe op het criterium *impact*, omdat erg onduidelijk is welke *impact* er precies meegenomen moet worden en wie nu de waardering krijgt voor bepaalde ontwikkelingen. Dit criterium is ook nieuw t.o.v. de RAE. Bij de beoordeling moeten de te beoordelen *units* één of meerdere *cases* indienen aan de hand waarvan ze hun *impact* beschrijven. *Impact* wordt daarbij gezien als invloed op economische, sociale, *public policy*, culturele of kwaliteit-van-leven-aspecten. Als indicatoren voorziet men indicatoren die:

- aangeven hoeveel *research*inkomsten er zijn gegenereerd uit gebruikers zoals bedrijven, overheid, en liefdadigheidsinstellingen.
- een maat geven voor de mate van samenwerking met gebruikers.
- specifiek zijn voor een bepaald onderzoeksgebied. Deze zouden geselecteerd moeten worden uit een soort menu. Een voorstel voor dit type indicatoren is te vinden in [HEFCE (2009)].

Kritiek vanuit de ingenieurswetenschappen

In een grootschalige consultatieronde zijn verschillende belanghebbenden gevraagd commentaar te geven op de beoogde opzet. Voor dit advies is het commentaar van de Royal Academy of Engineering het meest interessant [Royal Academy of Engineering, 2009]. Hun belangrijkste bedenkingen zijn:

- Men onderschrijft dat het gebruik van citatie-analyse niet robuust genoeg is om gebruikt te worden bij de beoordeling van onderzoeksresultaten en betwijfelt of dit ooit het geval zal worden gezien de breedte van het vakgebied en de verschillende typen van *output*. Bovendien wijst men erop dat de verschillen in publicatiegewoontes tussen de verschillende subdisciplines bijzonder groot zijn.
- Het gebruik van citatie-informatie als informatie voor de panels wordt toegejuicht maar men wijst er wel op dat beschikbare informatie uit citatiedatabases vaak ontoereikend is voor de *engineering*disciplines waardoor er grote verschillen kunnen ontstaan in de aangeleverde informatie in de verschillende subdisciplines.
- Men onderschrijft dat de *impact* van onderzoek meegenomen moet worden in de beoordeling. Dit is echter gezien de breedte van de verschillende soorten *engineering output* een lastig proces. Het vergt tijd om hier ervaring mee op te doen. De voorgestelde weegfactor van 25 procent voor dit criterium wordt daarom in dit stadium te hoog geacht.
- In de *Research Assessment Exercise* waren industriële leden volwaardig lid van de *expert panels*. In het huidige voorstel van de REF zijn zij alleen betrokken bij de beoordeling van de *impact*. De Royal Academy of Engineers pleit ervoor dat industriële leden bij het *engineering panel* betrokken zijn bij de beoordeling van alle criteria.
- De Royal Academy of Engineering heeft sterke bedenkingen bij het onderbrengen van alle *engineering*onderzoek in één panel. Ten eerste omdat hierdoor een enorme hoeveelheid onderzoek (omvang ongeveer 4.400 fte onderzoekers) door één panel moet worden beoordeeld. Ten tweede omdat men verwacht dat hierdoor de concurrentie tussen de *engineering*disciplines onderling toeneemt waardoor er uiteindelijk minder geld voor de *engineering*wetenschappen komt.

BL3.2 Australië, Excellence in Research for Australia

Op initiatief van de Australische minister voor innovatie, industrie en wetenschap, senator Kim Carr, wordt sinds 2008 een ambitieuze methode ontwikkeld voor de evaluatie van wetenschappelijk onderzoek. Dit zogenaamde *Excellence in Research for Australia* (ERA)-initiatief, wordt ontwikkeld door de Australian Research Council (ARC) samen met het Department of Innovation, Industry, Science and Research en is gebaseerd op een combinatie van *metrics* en *expert review*. In 2009 is ervaring opgedaan met een beperkte proef in twee clusters: *physical, chemical and earth sciences* en *humanities and creative arts*. In 2010 gaat de beoordeling voor alle vakgebieden volgens deze methode van start.

Met deze exercitie wil men informatie verzamelen over de kwaliteit van het onderzoek, nationale sterktes en zwaktes identificeren, excellentie, opkomende gebieden en kansen opsporen en positionering in de internationale onderzoeksarena bepalen.

Manier van beoordelen

Voor deze evaluatie wordt het onderzoek in acht clusters met min of meer vergelijkbare disciplines ingedeeld:

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*

Deze disciplines zijn weer onderverdeeld in een groot aantal subdisciplines. De discipline *engineering* bestaat bijvoorbeeld uit de subdisciplines *aerospace engineering*, *automotive engineering*, *biomedical engineering*, etc. De betrokken instellingen worden in principe beoordeeld op het niveau van deze subdisciplines. Alleen als zo'n subdiscipline te weinig *output* heeft om een goede beoordeling te maken, vindt beoordeling op niveau van de overkoepelende discipline plaats. De beoordeling heeft dus geen betrekking op complete faculteiten of op individuele onderzoekers. ERA gebruikt een reeks van indicatoren om de kwaliteit van het onderzoek te meten (zie tabel BL 3.1).

De eenheden worden geëvalueerd door *Research Evaluation Committees* (RECs), bestaande uit internationaal erkende experts op het gebied van onderzoeksevaluaties en de betreffende disciplines. Er zijn geen richtlijnen gegeven voor de manier waarop de REC-leden worden geselecteerd. De voorzitter van de REC vraagt de overige leden ieder individueel de eenheid te evalueren op grond van de aangeleverde data. In een later stadium komt de REC samen om de resultaten te bespreken. Alleen in relevante (sub)disciplines voeren de REC-leden, zonedig aangevuld met externen, een *peer review* uit voor bepaalde vormen van *research output*. In een zeer uitgebreide disciplinematrix staat per discipline aangegeven welke *output* hiervoor in aanmerking komt.

De instellingen worden verzocht op (sub)disciplineniveau gegevens aan te leveren. Deze gegevens moet men invoeren in het *System to Evaluate the Excellence of Research* (SEER) waarna ze elektronisch worden verwerkt. Op grond van de aangeleverde gegevens wordt het indicatorenpatroon en de hoeveelheid *output* per indicator vastgesteld. De indicatoren worden alleen aan de REC doorgegeven als een bepaalde disciplinespecifieke drempelwaarde is overschreden.

Tabel BL3.1. Beoordelingscriteria in Excellence in Research for Australia

Categorie	Indicatoren
<i>Ranked outlets</i>	Publicaties in wetenschappelijke tijdschriften Publicaties in <i>reviewed-conference proceedings</i>
<i>Citation analyses</i>	Relatieve citatie-impact (tegen wereldgemiddelde en Australisch gemiddelde voor het betreffende veld) Verdeling van publicaties in vergelijking met de rest van de wereld en in vergelijking met het betreffende veld
<i>Research volume and activity</i>	Totaal aan wetenschappelijk <i>output</i> voor verschillende resultaten, aantal fte van betrokken onderzoekers
<i>HERDC Research income</i>	Aantal subsidies Inkomsten per subsidie Totale inkomsten voor onderzoek Inkomsten voor onderzoek per fte Verhouding <i>research</i> inkomsten per fte in vergelijking met discipline specifieke <i>benchmark</i>
<i>Esteem</i>	Redacteur bij A- en A*-tijdschriften Bijdrage aan een prestigieus <i>work of reference</i> <i>Curatorial role</i> Gekozen <i>fellowship</i> van <i>learned academies</i> Nationale competitieve <i>research-fellowships</i> Prestigieuze prijzen en erkenningen
<i>Applied</i>	Toegekende patenten Kwekersrechten Geregistreerde ontwerpen Inkomen via commercialisatie wetenschap NHMRC-(gezondheidszorg)handleidingen

Output niet alleen wetenschappelijke publicaties

Voor bepaalde onderzoeksdisciplines is het mogelijk om wat men noemt niet-traditionele *output* ter beoordeling op te voeren. Voor de technische wetenschappen is deze mogelijkheid echter beperkt tot de disciplines *built environment* en *design* (w.o. architectuur). In geen van de overige *engineering*disciplines is dit toegestaan. Onder niet-traditionele *output* wordt dan bijvoorbeeld verstaan architectonische ontwerpen, creatief 3D-werk, design, games, computerprogramma's, tentoonstellingen.

Indien een werk wordt opgevoerd ter beoordeling als niet-traditionele *output* moeten de onderzoekers nadrukkelijk aangeven waaruit de wetenschappelijke component bestaat. Dit gebeurt in een verklaring waarin de volgende aspecten worden geadresseerd:

Onderzoeksachtergrond:

- veld
- context
- onderzoeksvraag

Bijdrage aan wetenschap:

- innovatieve component
- genereren nieuwe kennis

Significantie wetenschap:

- bewijs van excellentie

BL3.3 Finland, evaluatie Aalto University

Eind 2009 is het universitaire systeem in Finland grondig hervormd. De financiële en administratieve autonomie van de universiteiten werd vergroot. Ze werden opgedeeld in instellingen vallend onder *public law* of stichtingen vallend onder *private law*. Vanaf eind 2009 vielen de stichtingen niet meer onder het *state budgetary system*. Momenteel wordt een nieuw vorm van fondsverstrekking uitgewerkt die meer nadruk legt op strategische aspecten dan op indicatoren afgeleid van opleidings-/ onderwijstaken (in het bijzonder aantal studenten, promoties etc.). In het kader hiervan is men nu op zoek naar nieuwe kwaliteitsindicatoren.

Om een beeld te krijgen van de manier waarop in Finland over de kwaliteitsbeoordeling van wetenschappelijk onderzoek wordt gedacht, is gekeken naar de recent uitgevoerde evaluatie van de Aalto University in Helsinki. Hoewel deze evaluatie slechts betrekking heeft op één universiteit wordt deze hier toch besproken, omdat het een recente en grootschalige evaluatie is van een universiteit met een belangrijk aandeel faculteiten uit de ontwerpende en construerende wetenschappen.

De Aalto universiteit is in 2009 ontstaan uit een fusie tussen de Helsinki School of Economics, de Helsinki University of Technology en de University of Art en Design Helsinki. In datzelfde jaar heeft een *peer review*commissie (PRC), bestaande uit 62 leden afkomstig uit twintig landen, de universiteit geëvalueerd over de periode 2003-2008. De evaluatie had als doel om:

- de wetenschappelijke kwaliteit en de maatschappelijke *impact* van de nieuwe universiteit te bepalen;
- de groepen te identificeren die potentie hebben om internationaal te excelleren;
- het evaluatieproces zelf te analyseren.

De Aalto universiteit wil de aanbevelingen gebruiken om de onderzoeksstrategie te definiëren en *research practices* te ontwikkelen, inclusief het strategisch toewijzen van onderzoeksgelden.

Manier van beoordelen

De evaluatie is uitgevoerd door een PRC, verdeeld over 9 panels

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

Het zijn negen losse *peer review*commissies met ieder vijf tot zeven leden. Ieder panel heeft de opdracht een discipline te evalueren aan de hand van een zelfevaluatie, een bibliometrische analyse en een bezoek ter plaatse. Voorafgaand aan de visitatie hebben de disciplines zelf uitgebreide informatie aangeleverd.

Alleen actieve, aan de universiteit verbonden onderzoekers mochten resultaten aanleveren. Daarnaast is een bibliometrische analyse uitgevoerd. De PRC werd verzocht speciale aandacht te geven aan de internationale onderzoekspositie. Het internationale onderzoeksveld heeft dan ook steeds als ijkpunt gefungeerd.

Aan de hand van een bezoek ter plaatse en de aangeleverde informatie (zelfevaluatie en bibliometrische analyse) heeft de PRC per discipline cijfers toegekend aan de criteria:

1. Wetenschappelijk kwaliteit
2. Wetenschappelijke invloed (*science impact*)
3. Maatschappelijke invloed (*societal impact*)
4. Onderzoeksomgeving (*research environment*)
5. Toekomstig wetenschappelijke potentie (*future research potential*)

Deze cijfers varieerden van 1 tot 5 waarbij 5 staat voor uitmuntende internationale zichtbaarheid en 1 voor internationaal opkomend.

De PRC heeft de uitkomsten van de bibliometrische analyse, met name de NCSf (citatiescore genormaliseerd naar onderzoeksveld) expliciet meegenomen in het oordeel van de verschillende disciplines. Mede hiermee is zichtbaarheid in de (inter)nationale onderzoeksarena bepaald. De discipline *architecture, design, media* en *art research* vormt hierop een uitzondering; hier zijn de bibliometrische analyses niet meegenomen omdat er te weinig ISI-publicaties zijn.

Output niet alleen wetenschappelijke publicaties

De PRC geeft aan dat de disciplines 1 t/m 8 goed te evalueren zijn met de huidige, op *science* gerichte criteria. De PRC geeft aan dat de Aalto universiteit op papier de ruimte schept om alle vormen van wetenschappelijke *output* (vooral naast publicaties ook artefacten) mee te nemen, maar dat er in praktijk vanuit het bestuur sterk aangestuurd is op het leveren van *science* gerelateerde *output*. Het gevolg was dat vooral de eenheden uit de discipline *architecture, design, media* en *art research* veel geringere hoeveelheden *output* (alleen hun artikelen) hebben aangeleverd dan ze bij de bezoeken van de PRC bleken te hebben. Het betreffende panel was van mening dat het bestuur hierdoor een ernstige onderwaardering van de waarde van het wetenschappelijke onderzoek en de maatschappelijke relevantie heeft gecreëerd.

Meer informatie over deze evaluatie is te vinden in [Aalto University, 2009a, b, c].

BL3.4 Verenigde Staten, National Science Foundation

In de Verenigde Staten is gekeken naar de werkwijze van de National Science Foundation (NSF). De NSF is een onafhankelijk federaal agentschap met als doel: *'to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense...'*. Met een jaarlijks budget van 6,9 miljard dollar subsidieert NSF ongeveer 20 procent van het federaal gefinancierde fundamentele onderzoek aan de Amerikaanse universiteiten en colleges. In tegenstelling tot de in de vorige paragraaf beschreven beoordelingssystemen richt NSF zich vooral op ex ante-beoordeling van onderzoeksvoorstellen. Naast de NSF zijn in de VS nog veel private, missiegeoriënteerde en industriële fondsverstrekkers actief. Deze leunen meestal minder op duidelijk vastgelegde criteria.

Centraal in de beoordeling van onderzoeksvoorstellen bij de NSF is *peer review*. Daarbij worden voorstellen voornamelijk beoordeeld op twee criteria:

1. Wat is de wetenschappelijke merites van de voorgestelde onderzoeksactiviteit?
2. Wat is de bredere impact van de voorgestelde onderzoeksactiviteit?

Daarnaast kunnen er afhankelijke van het programma nog aanvullende criteria worden opgelegd.

Manier van beoordelen

Het voorstel komt binnen bij een *program officer* die het doorsluist naar een *review panel*. Dit panel, bestaande uit drie tot tien leden, beoordeelt het voorstel op bovenstaande criteria. Op basis van de oordelen van de panelleden stelt de *program officer* een analyse en een aanbeveling op. De *division director* van NSF neemt op basis van deze aanbeveling een definitief besluit over toekenning.

De indieners van een onderzoeksvoorstel wordt gevraagd om suggesties te doen voor geschikte referenten en men mag ook aangeven wie geen geschikte referenten zijn.

BL3.5 Noorwegen

Wetenschappelijk onderzoek in Noorwegen wordt vooral uitgevoerd door (algemene) universiteiten, gespecialiseerde universiteiten, *university colleges* en onderzoeksinstituten. Het ministerie van onderwijs en onderzoek is hoofdcoördinator en de betreffende vakministeries zijn belast met en verantwoordelijk voor het promoten en financieren van onderzoek dat op hun terrein valt. Een deel van het budget wordt direct door het ministerie verstrekt. Dit geschiedde voorheen met name op grond van studenten-aantallen. Sinds 2003 is financiering ingedeeld in a) basisfinanciering, b) onderwijs en c) onderzoek. De onderzoeksfinanciering is gebaseerd op prestaties en strategische overwegingen. Om prestaties te meten kijkt men naar:

- aantal afgeronde bachelors
- masters
- PhD's
- externe samenwerking
- samenstelling personeel (aantal professoren)
- aantal EU- en RCN-beurzen
- patenten en *output* in de vorm van wetenschappelijke publicaties

Hiervoor kent men aan de publicaties weegfactoren toe afhankelijk van de publicatievorm (artikel, boek) en het -kanaal (tijdschrift, website, uitgever, in geval van boek).

Universiteiten moeten alle gegevens invoeren in een landelijk documentatiesysteem, omdat de financiering mede wordt bepaald op grond van de documentatie in dit systeem.

Naast de hierboven geschetste basisfinanciering kent het Research Council Norway (RCN) competitieve beurzen toe. Ingediende voorstellen worden beoordeeld op grond van een groot aantal criteria aan de hand van *peer review*.

Ook voert het RCN landelijke evaluaties uit bij door RCN gesubsidieerde projecten, op instellingsniveau en op disciplineniveau. Doel van de evaluaties is om de kwaliteit, de relevantie en de efficiëntie van het onderzoek te verhogen en sterktes en zwaktes te identificeren op nationaal en instellingsniveau. Bovendien gebruikt het RCN de uitkomsten bij het opstellen van een nationale onderzoeksstrategie. Daarnaast zijn de uitkomsten een hulpmiddel bij het toewijzen van onderzoeksgelden. De RCN werkt momenteel aan een *follow-up*-model voor deze evaluaties.

BL3.6 Zweden

In Zweden wordt wetenschappelijk onderzoek uitgevoerd aan veertien universiteiten en 25 instellingen voor hogeronderwijs. Het meeste van het met publieke gelden uitgevoerde onderzoek vindt plaats bij de instellingen voor hogeronderwijs. Een deel van het onderzoek wordt direct gefinancierd door de overheid. Bovendien financiert de overheid de *research councils* en de sectorale *research agencies*.

De grootste subsidieverstrekker is het Swedish Research Council (SRC) dat een jaarlijks budget van vier miljard Zweedse kronen ter beschikking stelt. Het SRC kent beurzen toe aan onderzoek verdeeld over zes gebieden waaronder *natural and engineering sciences*. Onderzoeksvorstellen worden voorgelegd aan veertig panels waar ze aan de hand van *peer review* geëvalueerd worden.

De op twee na grootste fondsverstrekker is de Swedish Agency for Innovation Systems (VINNOVA) met een budget van 2,15 miljard Zweedse kronen. Dit agentschap verstrekt fondsen aan zes onderzoeksgebieden waaronder *natural sciences and technology*. Zweden kent geen landelijk ex post-onderzoeksevaluatiesysteem.

BL3.7 Frankrijk

Het Franse hogeronderwijsstelsel kent onderzoeksinstituten, universiteiten en *grandes écoles*. Wetenschappelijk onderzoek wordt hoofdzakelijk aan de instituten uitgevoerd. Hierbij bepalen, financieren en evalueren de instituten zelf hun onderzoek. Momenteel wil de overheid (zie *Stratégie Nationale de Recherche et d'Innovation* (SNRI); de nationale onderzoeks- en innovatiestrategie) de universiteiten meer autonomie geven en de onderzoeksinstituten omvormen tot agentschappen die de onderzoeksmiddelen projectmatig verdelen. Onderzoek zou volgens deze strategienota verplaatst moeten worden naar de universiteiten. Daarnaast kunnen instituten subsidie ontvangen via het Agence Nationale de la Recherche (ANR).

Frankrijk kent geen landelijk ex post-evaluatiesysteem voor onderzoek uitgevoerd aan universiteiten, *grandes écoles* en instituten

BL3.8 Duitsland

In Duitsland zijn de publieke onderzoeksinstituten in vier netwerken georganiseerd:

- Max Planck Gesellschaft (MPG)
- Fraunhofer Gesellschaft (FhG)
- Helmholtz-Gemeinschaft Deutscher Forschungszentren (HGF)
- Wissenschaftsgemeinschaft Wilhelm-Gottfried-Leibniz (WGL)

De universiteiten worden in eerste instantie gefinancierd door de betreffende zestien deelstaten. De grootste federale fondsverstrekker is de Deutsche Forschungsgemeinschaft (DFG) met een jaarlijks budget van 1,3 miljard euro. De DFG verstrekt fondsen aan alle disciplines van humaniora tot *engineering* en krijgt geld van zowel de federale als de deelstaatsoverheden volgens een vaste verdeelsleutel. DFG valt onder auspiciën van het Bundesministerium für Bildung und Forschung (BMBF).

Andere grote fondsverstrekkers zijn de private Alexander von Humboldt Foundation en de Deutscher Akademischer Austausch Dienst (DAAD), die internationale (uitwisselings)programma's en projecten financiert.

Er vindt geen onderzoeksevaluatie op nationaal niveau plaats maar de DFG produceert wel een classificatielijst (*Förder ranking*), gebaseerd op de volgende criteria:

- aantal positieve subsidietoewijzingen DFG
- samenstelling onderzoeksgroep
- fondsverwerving uit derde geldstroom
- positie in door DFG gesubsidieerde programma's
- aantal DFG-*reviewers*
- aantal Alexander von Humboldt *visiting* onderzoekers
- aantal DAAD-wetenschappers en -studenten
- deelname aan EU-*framework*programma's
- publicaties in internationale tijdschriften

Daarnaast hebben de deelstaten Nordrhein-Westfalen (gebaseerd op externe financiering en het aantal PhD's) en Niedersachsen eigen evaluatie systemen ontwikkeld. Niedersachsen heeft bovendien een eigen academisch adviescommissie. In opdracht van het ministerie van Economische zaken zijn in Baden-Württemberg in 2008 de toegepaste onderzoeksinstituten geëvalueerd. Geen van deze evaluaties is direct aan subsidiering gekoppeld. Internationaal gezien lopen deze evaluatiesystemen niet voorop.

Ook het zgn. *Exzellenzinitiative*, een subsidieprogramma van de DFG, heeft als doel heeft om *toplevel* onderzoek te identificeren en de kwaliteit van het hogeronderwijs en wetenschap in de brede zin te bevorderen en te verbeteren. Met dit initiatief wil Duitsland zijn internationale concurrentiepositie verbeteren. De onderzoeksvoorstellen, ingestuurd door de hele Duitse wetenschappelijke gemeenschap, worden aan de hand van *peer review* beoordeeld. De uitgekozen voorstellen worden in een staalkaart gepresenteerd.

BL3.9 Oostenrijk

Wetenschappelijk onderzoek in Oostenrijk wordt gesubsidieerd door:

- Austrian Research Promotion Agency (FFG)
- Austrian Science Fund (FWF)
- Austria Wirtschaftsservice (AWS)

Deze fondsverstrekkers vallen onder het Ministry of Science and Research (BMWF), het Ministry of Transport, Innovation and Technology (BMVIT) en het Ministry of Economy, Family and Youth (BMWFJ) maar strategische verantwoordelijkheden tussen ministeries en fondsverstrekkers zijn niet geheel duidelijk en de verantwoordelijkheden van de verschillende fondsverstrekkers overlappen elkaar.

In 2003 is de National Foundation for Research, Technology and Development opgericht om een constante financiële buffer voor de grote fondsverstekkers te garanderen en de continuïteit van langdurige, grote projecten te waarborgen. Het geld is afkomstig van de Austria National Bank (OeNB) en het European Recovery Fund (ERP).

Oostenrijk kent geen landelijk ex post-evaluatiesysteem voor onderzoek. Wel streeft het platform Research & Technology Evaluation (Fteval) naar het opzetten van een transparant evaluatiesysteem in het kader van optimale strategische planning van RTD-beleid. Daarnaast wil het Fteval een evaluatiecultuur in Oostenrijk bevorderen in samenwerking met de beleidsmakers op het gebied van technology- en research-beleid.