

# **Research Prestatiemeting: Een Internationale Vergelijking**

Ben Jongbloed  
Carlo Salerno  
Jeroen Huisman  
Hans Vossensteyn

Center for Higher Education Policy Studies  
Universiteit Twente  
Postbus 217  
7500 AE Enschede  
T: 053-4893263  
F: 053-4340392  
E: [secr@cheps.utwente.nl](mailto:secr@cheps.utwente.nl)  
[www.utwente.nl/cheps](http://www.utwente.nl/cheps)

Kenmerk: C5BJ169  
13-04-2005

## Inhoud

Voorwoord.....	5
<b>1 Inleiding .....</b>	<b>7</b>
1.1 Achtergrond en probleemstelling .....	7
1.2 Accountability en andere redenen voor prestatiemeting .....	7
1.3 De onderzoeksvragen en selectie van landen .....	9
1.4 De vele gedaantes van onderzoeksprestaties .....	12
<b>2 Meting van onderzoeksprestaties in Vlaanderen .....</b>	<b>13</b>
2.1 De infrastructuur voor onderzoek.....	13
2.2 Recente beleidsontwikkelingen en prestatiemeting.....	15
2.3 De gemeten prestaties.....	16
2.4 Resultaten .....	17
<b>3 Measuring research performance in the United Kingdom.....</b>	<b>21</b>
3.1 Brief overview of national research system .....	21
3.2 How is research output measured? .....	21
3.3 Results from evaluation studies.....	23
3.4 Are measurements used for funding? .....	25
3.5 Recent developments and experiences .....	26
Appendix to chapter 3: Some of the OST indicators.....	29
<b>4 Measuring research performance in Germany .....</b>	<b>31</b>
4.1 Brief overview of national research system .....	31
4.2 How is research output measured? .....	32
4.3 Results from evaluation studies.....	33
4.4 Are measurements used for funding? .....	34
<b>5 Measuring research performance in Norway .....</b>	<b>37</b>
5.1 The Norwegian academic research system.....	37
5.2 Measuring research output .....	39
<b>6 Measuring research performance in Australia.....</b>	<b>45</b>
6.1 The Australian academic research system.....	45
6.2 Measuring Research Output .....	48
6.3 Results, experiences, problems.....	52
<b>7 Measuring research performance in New Zealand .....</b>	<b>57</b>
7.1 The NZ research system and its funding .....	57
7.2 Measures of research performance .....	58
7.3 The peer review process .....	59
7.4 Further information on the PBRF process.....	61
7.5 Evaluation of the PBRF and the Quality Evaluation .....	62
<b>8 Synthesis and Analysis.....</b>	<b>67</b>
8.1 General observations about RPM .....	67
8.2 Rationale for RPM.....	68
8.3 Two dimensions.....	69
8.4 Data used for RPM .....	70
8.5 RPM indicators used in the six countries .....	72
8.6 Trends, and some words of caution .....	74
Appendix to chapter 8. – Summary of 6-country findings .....	75
<b>9 Research Prestatiemeting in Nederland.....</b>	<b>76</b>
9.1 Inleiding.....	76
9.2 RPM in Nederland .....	76
9.3 Verslag expertbijeenkomst RPM.....	79
Appendix bij hoofdstuk 9: Stellingen ten behoeve van de expertbijeenkomst.....	83
Lijst van Contactpersonen .....	85



## Voorwoord

Dit rapport doet verslag van een onderzoek uitgevoerd door CHEPS voor het Ministerie van Onderwijs, Cultuur en Wetenschap, Directie Onderzoek en Wetenschapsbeleid. Doel van het project is het verzamelen van informatie over de wijze waarop in een aantal landen door nationale instanties de onderzoeksprestaties van afzonderlijke universiteiten in beeld worden gebracht. Zijn er indicatorsystemen of andere meetsystemen in gebruik aan de hand waarvan valide en betrouwbare informatie over onderzoeksprestaties wordt verzameld; wat zijn de karakteristieken daarvan en welke ervaringen zijn er mee opgedaan? In hoeverre wordt via prestatiemeting zowel zicht op de wetenschappelijke output en de kwaliteit ervan als de maatschappelijke toepassing en benutting van academisch onderzoek verkregen?

Ten behoeve hiervan is naar zes landen gekeken (Australië, Verenigd Koninkrijk, Duitsland, Vlaanderen, Noorwegen, Nieuw-Zeeland). De buitenlandse ervaringen, samengevat in hoofdstuk 8, vormden een input voor een expertbijeenkomst waaraan een aantal Nederlandse deskundigen heeft deelgenomen op uitnodiging van het Ministerie om van gedachten te wisselen over prestatiemeting. Deze bijeenkomst, waarvan wij verslag doen in hoofdstuk 9, had tot doel om zicht te krijgen op de voors en tegens van een systeem van prestatiemeting in Nederland.

Zes van negen hoofdstukken in dit rapport zijn in het Engels geschreven om terugkoppeling met onze contactpersonen in de niet-Nederlands sprekende landen mogelijk te maken. Deze contactpersonen hebben ons informatie verstrekt over systemen waarmee in hun land de researchprestaties worden gemeten.

Behalve deze contactpersonen danken wij bij deze de deelnemers aan de expertbijeenkomst voor hun bereidheid om in alle openheid hun oordeel te geven over research prestatiemeting. Onze dank gaat in het bijzonder uit naar drs. Jan van Steen, die vanuit het Ministerie dit onderzoek op deskundige wijze heeft begeleid.

Namens de auteurs,  
Ben Jongbloed



# 1 Inleiding

## 1.1 Achtergrond en probleemstelling

De bekostiging van het universitaire onderzoek staat al geruime tijd ter discussie. Niet alleen in Nederland, maar ook in de rest van Europa. Dynamisering, valorisatie, overhevelingsoperaties en prestatiebekostiging zijn termen die vaak worden gebruikt in de discussies over de wijze waarop publieke middelen voor onderzoek aan de Nederlandse universiteiten ter beschikking worden (of zouden moeten worden) gesteld. Twee vraagstukken staan daarbij centraal:

1. Hoe kunnen de universitaire onderzoeksprestaties worden gemeten?
2. Hoe kan deze meting worden ingebracht in de wijze waarop de publieke onderzoeksmiddelen over de universiteiten worden verdeeld?

Bij de eerste vraag draait het om het meten van de uitkomsten en het rendement van het universitaire onderzoek. Het 'rendement' wordt daarbij opgevat in termen van *excellentie* en *relevantie*. Idealiter moet universitair onderzoek enerzijds van een hoog wetenschappelijke kwaliteit zijn en anderzijds dient het een bijdrage te leveren aan de oplossing van maatschappelijke problemen op de korte dan wel langere termijn. Met de laatste doelstelling wordt veelal geduid op de positieve bijdrage die universitair onderzoek kan leveren aan de concurrentiepositie en innovatiekracht van een land.

Bij de tweede vraag is de kwestie aan de orde hoe de onderzoeksbekostiging kan worden aangepast opdat het geïnvesteerde geld het best rendeert. De discussies en plannen die in Nederland momenteel circuleren<sup>1</sup> richten zich sterk op *prestatiebekostiging* als een van de instrumenten die worden genoemd om de wetenschap beter te laten renderen. Het beleidsinstrument figureert in het werkprogramma van het Innovatieplatform en het recente Wetenschapsbudget.

De vraag die ten grondslag lag aan dit rapport dat op verzoek van de Directie Onderzoek en Wetenschapsbeleid van het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW) door CHEPS is geschreven luidt:

“Is het mogelijk om bij de verdeling van publieke onderzoeksmiddelen rekening te houden met de onderzoeksprestaties van universiteiten om op deze wijze te komen tot een verbetering van onderzoeksprestaties en productiviteit, in de zin van zowel wetenschappelijke kwaliteit als maatschappelijke relevantie?”

Het doel van dit rapport nu is een inventarisatie en analyse te maken van modellen voor de meting van onderzoeksprestaties in het buitenland en het trekken van lessen hieruit voor de Nederlandse situatie. Dit zal moeten plaatsvinden aan de hand van een aantal vragen die zijn geformuleerd door het Ministerie van OCW. Deze vragen laten we in de volgende paragraaf de revue passeren. Allereerst maken we echter enkele opmerkingen over de context van prestatiemeting.

## 1.2 Accountability en andere redenen voor prestatiemeting

In vrijwel alle OECD landen wordt in toenemende mate aandacht gegeven aan de verantwoordingsplicht (*accountability*) van publieke organisaties. Zo dienen ook de hoger onderwijsinstellingen die publieke middelen ontvangen voor het uitvoeren van academisch

---

<sup>1</sup> Zie onder andere AWT-advies 61: *Een Vermogen Betalen* (Februari 2005).

onderzoek hun efficiëntie en effectiviteit aan te tonen aan de subsidieverstrekker en de samenleving in het algemeen. Aan de roep om verantwoording van de middelen ingezet voor het doen van academisch onderzoek wordt in Nederland op verschillende manieren gehoor gegeven. De universiteiten produceren jaarverslagen en leveren gegevens aan ministerie, CBS, Inspectie, VSNU, KNAW, NWO, en andere organisaties. Daarnaast bestaan er systemen van *peer review*, dat wil zeggen onderdelen van universiteiten laten hun onderzoek eens in de zes jaar beoordelen door experts (*peers*) die zich een beeld vormen van de kwaliteit en kwantiteit van het onderzoek. In peer review wordt door vakgenoten en andere beoordelaars een gefundeerd oordeel gevormd van de prestaties op basis van een visitatie en/of het aandragen van informatie via een zelfstudie en ander bewijsmateriaal. Peer review brengt echter aanzienlijke kosten met zich mee – zowel voor de beoordeelde organisatie-eenheid (het schrijven van een zelfevaluatie, het aanleveren van publicaties, het organiseren van een site visit voor de beoordelaars), als voor de beoordelaars en de organisatie die de beoordelaars ondersteunt. Om deze reden vindt ze dan ook slechts eens in de zes jaar plaats.<sup>2</sup> Naast peer review is er daarom behoefte aan kwantitatieve indicatoren die meer frequent en op efficiënte wijze de academische onderzoeksprestaties in beeld kunnen brengen.

Het meten van onderzoeksprestaties dient niet alleen de verantwoordingsfunctie, maar levert ook informatie voor het evalueren van beleid en het vaststellen van beleidsprioriteiten voor de toekomst. Met andere woorden, prestatiemeting heeft zowel een ex post evaluatiefunctie als een ex ante evaluatiefunctie. Een voorbeeld van ex ante evaluatie is het identificeren van kwalitatief goede onderzoekers om deze in aanmerking te laten komen voor onderzoekssubsidies. Met andere woorden, prestatiemeting heeft meerdere doelstellingen. We noemen de volgende acht, ontleend aan Behn:<sup>3</sup>

1. Evalueren
2. Controleren
3. Budgetteren
4. Motiveren
5. Overtuigen
6. Vieren
7. Leren
8. Verbeteren

Hoewel de Nederlandse discussie zich momenteel toespitst op prestatiemeting als input voor beslissingen ten aanzien van de budgettering (doelstelling #3), dienen we in het achterhoofd te houden dat in andere landen andere doelstellingen aan de orde kunnen zijn. Er dient derhalve eerst duidelijkheid te zijn over de context waarbinnen prestatiemeting plaatsvindt alvorens men zich een oordeel kan vormen over de wijze van prestatiemeting. Een meting of een meetinstrument dat geschikt is voor een van de acht hier genoemde doelen kan immers ongeschikt zijn voor het nastreven van een van de andere zeven doelen. Kortom, men zal aan de doelstellingen van prestatiemeting aandacht moeten besteden zodra buitenlandse ervaringen met prestatiemeting worden geïnterpreteerd of er lessen uit worden getrokken voor Nederland.

---

<sup>2</sup> Om de administratieve lasten van peer reviews enigszins te beperken en ook om meer maatwerk ten aanzien van de beoordeelde organisatie mogelijk te maken is het Nederlandse visitatiesysteem in 2003 aanzienlijk gestroomlijnd. De opzet en organisatie van de beoordelingen van publiek gefinancierd onderzoek aan universiteiten is vastgelegd in het 'Standard Evaluation Protocol 2003–2009 – For Public Research Organisations'. Deze opzet staat bekend als het Van Bommel protocol, genoemd naar de voorzitter van de werkgroep die de nieuwe beoordelingssystematiek heeft voorgesteld in zijn rapport 'Kwaliteit Verplicht. Naar een nieuw stelsel van kwaliteitszorg voor het wetenschappelijk onderzoek'. Rapport van de werkgroep Kwaliteitszorg Wetenschappelijk Onderzoek en standpuntbepaling KNAW, NWO en VSNU (gepubliceerd in 2001).

<sup>3</sup> Zie Behn, R.D. (2003), Why measure performance? Different purposes require different measures. *Public Administration Review*. Vol. 63, No. 5, pp. 586-606.



De verschillende doelen die aan systemen voor research prestatie meting (RPM) zijn verbonden zijn op hun beurt weer verbonden met de verschillende gebruikers van RPM-systemen. Het ene systeem kan worden gebruikt door een ministerie ter onderbouwing van bekostigingsbeslissingen, terwijl het andere systeem door research councils wordt gehanteerd om de onderzoekers te identificeren die het best zijn gekwalificeerd om een zeker onderzoeksprogramma uit te voeren. Wat betreft de gebruikers van RPM-systemen onderscheiden we (zonder uitpuittend te willen zijn) de volgende partijen:

- beleidsmakers op nationaal niveau (ministeries, parlement)
- subsidieverstrekkers en intermediaire organisaties (research councils, onderzoeksstichtingen)
- internationale organisaties (OECD, EU, ERC)
- de onderzoeksgemeenschap (onderzoekers, researchinstellingen, universiteiten)

Elk niveau kent zijn eigen doelen en zijn eigen keuzes wat betreft meettechniek en/of de indicatoren.

### 1.3 De onderzoeksvragen en selectie van landen

We presenteren in deze paragraaf de onderzoeksvragen. De eerste – naar de selectie van landen – wordt daarbij direct beantwoord. Zoals we hierboven reeds hebben opgemerkt moet onze inventarisatie van systemen van prestatie meting in het licht worden gezien van de plannen om prestatie bekostiging in te voeren voor het universitair onderzoek. In dit kader gaat onze aandacht dan ook vooral uit naar buitenlandse modellen waarin onderzoeksprestaties worden meegewogen in de universitaire bekostiging. We vragen ons derhalve allereerst af

*Welke landen bezitten een systeem van prestatie meting voor universitair onderzoek dat wordt ingezet bij de verdeling van publieke onderzoeksmiddelen over universiteiten?*

We perken deze vraag verder in door ons alleen op de *eerste geldstroom* van universiteiten – de basis bekostiging – te richten. Bovendien focussen we op prestatie informatie die is toegespitst op het *instellingsniveau*. Wat dit laatste betreft, veel landen verzamelen weliswaar een grote hoeveelheid data voor allerlei doeleinden en organisaties (onder meer de OECD), maar aggregeren deze data voor rapportages die op het nationale niveau betrekking hebben. Wij zijn echter vooral geïnteresseerd in systemen die de data voor *afzonderlijke* universiteiten en onderzoeksgroepen (disciplines) vastleggen.

In een aantal landen wordt een vorm van prestatie bekostiging (en dus prestatie meting) toegepast voor de verdeling van publieke onderzoeksmiddelen. We noemen Australië, Noorwegen, het Verenigd Koninkrijk, Nieuw-Zeeland, Hong Kong, en Israël.<sup>4</sup> Van deze landen kiezen we de eerste vier. Daaraan voegen we Duitsland en Vlaanderen toe, omdat dit voor Nederland (buur-)landen zijn waar zich interessante ontwikkelingen in verband met prestatie meting en prestatie bekostiging voordoen. Aldus resulteren de volgende zes landen die we nader willen bezien in deze studie naar prestatie meting:

1. Vlaanderen,
2. Verenigd Koninkrijk
3. Duitsland
4. Noorwegen
5. Australië
6. Nieuw Zeeland

---

<sup>4</sup> Zie: Jongbloed, B. (2002), *Bekostiging universitair onderzoek: Perspectieven op een nieuw sturingsarrangement*. Rapport van de Werkgroep Financiering Master. Subwerkgroep Bekostiging Onderzoek Lange Termijn (BOLT). Enschede: CHEPS.

Deze keuze voor een beperkt, maar tegelijkertijd gevarieerd aantal landen is gemaakt in overleg met de opdrachtgever om de studie voldoende reikwijdte te geven qua landen en de mate van detail per land. Voor de zes genoemde landen schetsen we steeds de context waarin prestatiebekostiging en meting van onderzoeksprestaties plaatsvindt en wat (binnen de eerste geldstroom) de relatieve omvang van de prestatiegerelateerde bekostiging is. De globale organisatie van het universitaire onderwijs en onderzoek, de plaats en bekostiging van het universitaire onderzoek daarbinnen en de omvang en inrichting van de universitaire bekostiging dienen in ogenschouw te worden genomen bij het analyseren van systemen van prestatiemeting.

Voor de zes landen zal worden nagegaan welke prestaties worden gemeten en meetellen in de onderzoeksbekostiging (eerste geldstroom) en op welke wijze dit is vormgegeven. De tweede geldstroom valt enigszins buiten het bestek van onze opdracht, hoewel betoogd zou kunnen worden dat ze een vorm van prestatiebekostiging is. De tweede geldstroom bestaat immers uit subsidies en beurzen die door *research councils* in competitie aan individuele onderzoekers, projecten en programma's worden toegekend. Daarbij speelt de kwaliteit (en of de prijs-kwaliteitsverhouding) van de aanvraag c.q. de aanvrager een belangrijke rol via het track record en de past performance van de aanvrager(s). Ook de derde geldstroom (betalingen voor contractonderzoek) valt buiten onze studie – ook al zijn overheden (nationale, Europese) of non-profit instellingen vaak opdrachtgever en ook al hanteren zij bij de opdrachtverstrekking veelal prestatiecriteria. Hoewel we ons dus richten op de eerste geldstroom, kunnen niettemin toch de tweede en derde geldstromen in beeld komen. Het succes van een universiteit in het verwerven van middelen uit de tweede en derde geldstroom wordt, zoals we hierna zullen zien, namelijk soms gebruikt als prestatie-indicator.

Onze belangrijkste onderzoeksvraag luidt nu als volgt:

*Welke onderzoeksprestaties worden gemeten en op welke manier?*

De vraag is hier naar de gebruikte prestatie-indicatoren en in hoeverre in het buitenland (de zes landen) via prestatiemeting zowel zicht op de wetenschappelijke output en de kwaliteit ervan als de maatschappelijke toepassing en benutting van onderzoek wordt verkregen en in hoeverre hiermee rekening wordt gehouden in de bekostiging.

Ten aanzien van de wetenschappelijke prestaties zijn veelgebruikte indicatoren het aantal publicaties, het aantal citaties en het aantal uitgereikte onderzoeksdiploma's of doctoraten. De (bibliometrische) indicatoren 'publicaties' en 'citaties' zijn weliswaar relatief gemakkelijk beschikbaar en kunnen de kwaliteit van het onderzoek laten meewegen (via de impactfactor van een tijdschrift), maar kennen ook een aantal bezwaren. We komen daarop terug bij de volgende onderzoeksvraag.

*Wordt er rekening gehouden met verschillende vormen van onderzoeksprestaties van verschillende disciplines en zo ja, op welke manier gebeurt dit?*

Meting via een specifieke indicator kan voor de ene discipline zeer wel acceptabel zijn, terwijl ze voor een andere discipline veel minder goed werkt. Ter illustratie kan worden gewezen op verschillen die bestaan tussen de natuurwetenschappen en de maatschappijwetenschappen. In de eerstgenoemde discipline zijn publicatiegewoontes heel anders dan in de andere. Ook kan het zijn dat in sommige disciplines de nadruk niet ligt op publicaties, maar op andere producten van onderzoek. We noemen hier de outputs van disciplines als informatica of 'creatieve' disciplines die niet als publicatie maar in de vorm van computersoftware, ontwerpen, kunstuitingen, et cetera naar buiten worden gebracht. De vraag is dus of en in hoeverre er bij het meten van onderzoeksprestaties rekening wordt gehouden met zaken die specifiek zijn voor een discipline.

De prestatiemeting kan op verschillende wijzen zijn georganiseerd. Nationale organisaties, variërend van ministeries, statistische bureaus en evaluatiecommissies tot bufferorganisaties

kunnen erbij betrokken zijn naast (vertegenwoordigers van) universiteiten en onafhankelijke deskundigen. Een belangrijke vraag is:

*Hoe is het systeem van prestatiemeting georganiseerd en wat zijn de kosten ervan?*

Deze vraag naar de kosten is lastig te beantwoorden omdat prestatiemeting zoals gezegd niet alleen plaatsvindt in het kader van de bekostiging of verantwoording, maar ook vele andere doelen kan dienen (zie paragraaf 1.2). Kortom, de kosten van prestatiemeting zijn veelal onderdeel van de dagelijkse apparaatskosten van instellingen en ministeries en moeilijk apart te kwantificeren. Ondanks dit gegeven zijn voor het Verenigd Koninkrijk schattingen gemaakt van de kosten van de *Research Assessment Exercise* waarmee op grond van prestatiemeting door de publieke autoriteiten onderzoeksmiddelen over de Britse universiteiten worden gealloceerd (zie hoofdstuk 3).

Als landen een systeem van prestatiemeting hanteren en dit gebruiken ter onderbouwing van de universitaire onderzoeksbekostiging is een volgende vraag:

*Welke ervaringen (dan wel evaluatieve studies) zijn er in de betreffende landen met een dergelijk systeem?*

Belangrijk daarbij is het eventuele *gedragseffect* van de prestatiemeting en -bekostiging. Heeft de meting gevolgen voor het gedrag van onderzoekers en universiteiten en, zo ja, moet dit gedrag als gewenst, dan wel ongewenst worden beschouwd? Een keuze voor specifieke prestatieindicatoren (PI's) en daaraan gekoppelde bekostigingsgrondslagen geeft een signaal richting onderzoekers ten aanzien van wat door de partij die de PI's verlangt blijkbaar van waarde wordt geacht en zal om die reden gevolgen kunnen hebben voor het gedrag van onderzoekers en instellingen. PI's die voor de bekostiging worden gebruikt kunnen ongewenst gedrag uitlokken als onderzoekers meer nadruk leggen op het aantal publicaties in plaats van de kwaliteit ervan, of als ze in hun onderzoek bepaald risicomijdend gedrag aan de dag leggen, zich op geijkte paden blijven begeven of gespitst zijn op snelle publicaties in plaats van publicaties die na de meetdatum liggen.

Door naar een selectie van hoger onderwijsystemen te kijken waar van *research prestatie meting* (RPM) gebruik wordt gemaakt in de onderbouwing van het nationale onderzoeksbeleid, hopen we een beter beeld te krijgen van de voors en tegens van het gebruik van PI's en andere systemen van RPM. Gelet op de ervaringen in het buitenland is de vraag in hoeverre de systemen van prestatiemeting en -bekostiging kunnen worden getransplanteerd naar Nederland. Vragen die we stellen zijn:

- *Wat zijn bruikbare systemen voor de Nederlandse situatie?*
- *Wat zijn de mogelijkheden tot het ontwikkelen van een meetsysteem in de Nederlandse situatie en op welke manier zou dit kunnen gebeuren?*
- *Welke informatie is hiervoor nodig en welke informatie is hiervoor reeds beschikbaar?*

Terwijl de eerder genoemde vragen rondom prestatiemeting aan de hand van zes landenstudies (case studies) zullen worden beantwoord, ons daarbij baserend op desk research, zal onze verkenning van de mogelijkheden voor prestatiemeting in Nederland worden uitgevoerd aan de hand van een expertbijeenkomst. In deze bijeenkomst wordt van een aantal daartoe uitgenodigde deskundigen een oordeel gevraagd over de mogelijkheden en haalbaarheid van prestatiemeting in Nederland en van gedachten gewisseld over prestatiebekostiging. Tot de genodigden behoren vertegenwoordigers/afgevaardigden van: universiteiten, VSNU, QANU, AWT, NWO, KNAW, Ministerie van Onderwijs, Cultuur en Wetenschap, Ministerie van Economische Zaken, Centraal Planbureau, CWTS (Universiteit Leiden) en CBS.

De uitkomsten van de zes case studies en de expertbijeenkomst zijn in de navolgende hoofdstukken opgenomen. Omdat enkele teksten ter controle zijn voorgelegd aan buitenlandse deskundigen zijn vijf case studies (behalve de Vlaanderen case) in het Engels geschreven.

#### 1.4 De vele gedaantes van onderzoeksprestaties

In het verloop van de volgende case studies zal veelvuldig het woord *onderzoeksprestaties* worden gebruikt. Het zal echter duidelijk zijn dat deze term een aantal verschillende uitingen van prestaties bevat:

1. *outputs* – welke producten zijn het directe gevolg van de onderzoeksactiviteiten van universiteiten?
2. *uitkomsten* (outcomes) – wat zijn de bruikbare en tastbare resultaten van onderzoek in de zin van uitkomsten van conceptuele aard (bijv. een nieuwe theorie), van praktische aard (een nieuwe analysetechniek), of fysieke aard (een nieuw apparaat, product of ontwerp)?
3. *impact* – wat is de invloed en wat zijn de gevolgen die onderzoeksuitkomsten hebben in de onderzoekswereld zelf, dan wel in de bredere samenleving? (Bijv. wat is de sociaal-economische invloed).

Lopend van 1 naar 3 zijn de prestaties minder concreet en minder direct aan onderzoeksactiviteiten toe te schrijven. We maken kort enkele nadere opmerkingen.

ad 1. Onder de outputs vallen onder andere boeken, wetenschappelijke publicaties, onderzoekspapers, ‘populaire publicaties’ (krantenartikelen), datasets, software, consultancies, aantal uitgereikte onderzoeksdiploma’s of doctoraten.

ad 2. Via peer review wordt veelal een beeld gevormd van de uitkomsten en betekenis van onderzoek – of de onderzoeksoutputs worden benut en hoe de kwaliteit in brede zin moet worden beoordeeld. De keuze van de peers (het panel van beoordelaars) en de criteria die zij aanleggen zijn derhalve van belang voor de vraag wat wordt gemeten en hoe uiteenlopende prestaties worden meegewogen. Peer review kent dus vele varianten.

ad 3. Voor het in beeld brengen van de bredere (sociaal-economische) impact van onderzoek lijken op het eerste gezicht slechts weinig indicatoren beschikbaar te zijn. Soms evalueren gebruikerspanels deze bredere bijdrage van onderzoek via industry/user surveys of vindt een vorm van *modified peer review* plaats. Ook worden de inkomsten uit octrooien en licenties wel gebruikt om de impact van onderzoek te benaderen. Als het om de impact in de wetenschappelijke wereld gaat komen citaties en uitkomsten van bibliometrische analyses naar voren als indicator. Een van de doelen van onze studie is dan ook te inventariseren welke indicatoren in gebruik zijn in de diverse landen voor het meten van de interacties van onderzoekers met industrie en samenleving.

Wat het laatste betreft, het onderzoekssysteem en zijn prestaties zijn meer en meer van belang in een maatschappelijke context waarin kennis en innovatie sleutelbegrippen zijn geworden. De vraag is hoe het nationale onderzoekssysteem presteert en of het bij de wereldtop behoort. Wat krijgt de samenleving terug voor de publieke en private middelen die ze investeert in academisch onderzoek? Dit rendement luidt niet alleen in financieel-economische termen, maar uit zich ook in sociale en culturele effecten.

## 2 Meting van onderzoeksprestaties in Vlaanderen

### 2.1 De infrastructuur voor onderzoek

#### *Actoren*

In Vlaanderen zijn twee departementen verantwoordelijk voor het onderzoeksbeleid: de Administratie Wetenschap en Innovatie (AWI) en de Administratie Hoger Onderwijs en Wetenschappelijk Onderzoek (AHOWO). De laatste richt zich vooral op het beleid ten aanzien van het universitaire onderzoek. Belangrijke (min of meer uitvoerende) – decretaal ingestelde – organisaties zijn het Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO) en het Instituut voor de aanmoediging van Innovatie door Wetenschap en Techniek (IWT). De Vlaamse Raad voor Wetenschapsbeleid (VRWB) is ingesteld om de Vlaamse regering te adviseren (op verzoek of uit eigener beweging) inzake wetenschaps- en technologiebeleid. De raad bestaat uit zes leden uit het academische milieu, zes uit het socio-economische milieu en vier door de regering benoemde leden.

De uitvoerende onderzoeksorganisaties zijn de Vlaamse universiteiten, de Vlaamse onderzoeksinstituten (waarvan de drie grootste zijn: IMEC: Interuniversitair Micro-elektronica Centrum [1985], VITO: Vlaamse Instelling voor Technologisch Onderzoek [1993] en VIB: Vlaams Interuniversitair Instituut voor de Biotechnologie [1995]), collectieve centra (opgezet door bedrijfsfederaties), bedrijven en hogescholen. De hogescholen zijn recent meer betrokken bij het onderzoeksbeleid, vanwege de vorming van associaties met de universiteiten en de introductie van een apart financieringsinstrument, het Projectmatig Wetenschappelijk Onderzoek (PWO).

#### *Typen onderzoek en geldstromen*

In Vlaanderen wordt een onderscheid gemaakt tussen drie typen onderzoek: grensverleggend onderzoek, strategisch basisonderzoek en toegepast onderzoek. Onder grensverleggend onderzoek wordt onderzoek verstaan met als doel om nieuwe kennis te vergaren, zonder daarbij een specifieke toepassing of specifiek gebruik op het oog te hebben. Strategisch basisonderzoek wordt gedefinieerd als kwalitatief hoogwaardig op middellange en lange termijn gericht onderzoek dat het opbouwen van wetenschappelijke of technologische capaciteit beoogt. In – Nederlandse – termen van geldstromen, vallen grensverleggend en strategisch basisonderzoek onder de eerste en tweede geldstroom. Toegepast onderzoek valt onder de derde geldstroom en is gericht op een concrete, praktische toepassing van kennis die op korte termijn te realiseren is.

#### *Instrumenten en bekostiging*

Een belangrijk deel van de universitaire budgetten bestaat uit de operationele middelen van de overheid. Deze budgetten dekken in belangrijke mate (ongeveer 80-85%) de personeelskosten. Daarnaast draagt de overheid bij in de vorm van investeringsbudgetten en sociale voorzieningen. In verhouding tot de operationele middelen zijn dit relatief beperkte budgetten.

Projectmatige overheidsmiddelen worden toegekend aan instellingen via het Bijzonder Onderzoeksfonds (BOF), het Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO) en het Instituut voor de aanmoediging van Innovatie door Wetenschap en Techniek (IWT). Tabel 1 geeft een overzicht van de overheidsbudgetten in het afgelopen decennium.

Tabel 1: Overheidsbudgetten 1993-2003 (in miljoenen Euro, prijsniveau 2004)

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Operationele subsidies	480	484	495	506	518	528	536	541	555	572	580
Investeringsmiddelen	17	17	16	16	16	17	18	23	23	23	23
Sociale voorzieningen	-	13	13	14	14	14	15	14	14	14	14
BOF	23	23	24	32	39	45	52	69	87	93	94
FWO	59	61	64	73	83	90	97	96	84	85	86
IWT (spec. beurzen)	7	9	12	13	14	13	17	16	17	19	18
Totaal	587	606	624	654	686	708	734	760	780	805	816

Bron: Vervliet, G. (2004), *Science, technology and innovation. 2004 Information guide*. Brussel: Ministerie van de Vlaamse Gemeenschap.

Het Bijzonder Onderzoeksfonds heeft als doelstelling het universitaire, instellingsspecifieke onderzoeksbeleid uit te bouwen door middel van intra-universitaire competitie. Middelen worden verdeeld over universiteiten, gebaseerd op de hoogte van de werkingstoelagen (operationele middelen, gewicht 15%), de aantallen tweede cyclusediploma's (over een periode van vier jaar, gewicht 35%) en de aantallen doctoraatsdiploma's (over een periode van vier jaar, met gewichten<sup>5</sup> naar discipline, gewicht 50%). Vanaf 2003 worden ook criteria meegenomen die gebaseerd zijn op wetenschappelijke productie (publicaties en citaties, met hulp van de *Science Citation Index*) in de afgelopen tien jaar. In eerste instantie wordt 10% van het budget hierdoor bepaald, maar dit percentage groeit tot 30% vanaf 2005<sup>6</sup>. De universiteiten zijn autonoom met betrekking tot het intern verdelen van de BOF-middelen.

Het Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO) verdeelt middelen op basis van competitie tussen universiteiten. Het bestuur van het FWO bestaat uit vertegenwoordigers van Vlaamse universiteiten, de Vlaamse overheid en bedrijfsleven. Middelen van het FWO kunnen worden aangewend voor beurzen voor doctoraatsstudenten en postdocs, meerjarige onderzoeksprojecten, internationale netwerken en congressen en symposia. Voorstellen worden geselecteerd op kwaliteit door middel van peer review (met buitenlandse inbreng), maar ook op basis van het aantal en het belang van externe contracten en het aantal promoties.

Het Instituut voor de aanmoediging van Innovatie door Wetenschap en Techniek (IWT) – opgericht in 1991 – is voornamelijk gericht op het toegepast (technologisch) onderzoek in bedrijven. Als Vlaamse overheidsinstelling stimuleert en ondersteunt IWT-Vlaanderen de technologische innovatie in Vlaanderen. De taakstelling van het IWT werd gedefinieerd volgens het innovatiedecreet van 1999. Het IWT draagt zorg voor de uitvoering van het programma Strategisch Basisonderzoek (SBO). Dit programma beoordeelt en financiert onderzoeksprojecten op het grensvlak van wetenschappelijke excellentie en maatschappelijk valorisatiepotentieel. Het programma staat open voor alle betrokkenen in het Vlaamse onderzoekslandschap. Daarnaast speelt het IWT een rol in het toekennen van specialisatiebeurzen voor doctoraatsopleidingen en post-doctorale onderzoekers en ondersteunt het IWT het bedrijfsleven bij deelname aan internationale programma's.

Tot slot is het Industrieel Onderzoeksfonds (IOF) vermeldenswaard. Dit fonds stelt zich ten doel een financieel kader te bieden voor instellingen om het onderzoek dat door externen wordt gefinancierd te ondersteunen. Het is dus bedoeld voor de uitbouw van het instellingsbeleid van de universiteiten inzake strategisch en toegepast onderzoek.

<sup>5</sup> Er zijn drie gewichten (1,2,3) die een weerslag zijn van de kosten van het doen van onderzoek. Natuurkunde heeft een gewicht van 3, economie 1.

<sup>6</sup> Overigens worden de prestaties van de onderzoeksinstellingen (IMEC, enz.) en van afzonderlijke disciplines (bijv. biomedisch en natuurwetenschappelijk onderzoek) ook regelmatig gemeten door middel van bibliometrische indicatoren. De toepassing van bibliometrische indicatoren is in die zin uniek voor de BOF-middelen, omdat de prestatiemeting financiële consequenties heeft.

## 2.2 Recente beleidsontwikkelingen en prestatie­meting

De aandachtspunten in het recente onderzoeks­beleid zijn af te lezen uit de Beleidsnota's 2000-2004 en 2004-2009.<sup>7</sup> Deze nota's vragen vooral aandacht voor het tekort aan gekwalificeerde onderzoekers en het aantrekkelijker maken van een academische loopbaan, het uitbouwen van degelijke basis­infrastructuur voor het onderzoek, het optimaliseren van de beleidsinstrumenten, het bereiken van de Lissabon doelstellingen. Het innovatiepact (een plan van aanpak voor het verbeteren van de infrastructuur voor de gehele kennis- en innovatieketen, getekend in 2003) mag niet onvermeld blijven.

De aandacht voor prestatie­meting moet gezien worden in de context van het stijgende belang van beleids­indicatoren (prestatie-­indicatoren) voor de ontwikkeling van de Vlaamse kennis­economie. De Vlaamse overheid is van mening dat beter dan voorheen de zwakke en sterke punten in de beleids­voorbereiding, -uitvoering en -opvolging in kaart moeten worden gebracht om effecten van een aantal beleids­maatregelen te kunnen opvolgen. Deze ontwikkelingen moeten ook gezien worden in relatie tot het Vlaamse comptabiliteits­decreet, waarin sterke aandacht is voor prestatie­meting, interne controle binnen agentschappen, alsmede aandacht voor beleidseffectenrapportage.<sup>8</sup>

De recente aandacht betekent overigens niet dat er in het recente verleden geen aandacht is geweest voor prestatie­meting (met betrekking tot het onderzoek). Vermeldenswaard zijn de initiatieven van het Vlaams Technologie Observatorium (VTO) van het IWT in het werkprogramma 1996-1999 die sterk aanleunen tegen de werkzaamheden van OESO en Eurostat op het terrein van innovatie-­indicatoren (octrooien, publicaties, innovaties, technologische betalings­balans).<sup>9</sup> Daarnaast moet niet onvermeld blijven dat prestatie­meting reeds in het begin van de jaren negentig een belangrijke impuls kreeg. Twee belangrijke oorzaken waren de stagnerende overheids­bijdrage voor onderzoek (die de universiteiten uitdaagde om externe middelen te verwerven) en de verplichting om evaluaties uit te voeren naar de centrale instellings­activiteiten. Deze ontwikkelingen noopten de instellingen om een coherent onderzoeks­beleid uit te zetten, waarbij meting van bereikte resultaten een belangrijk element geacht werd. Voorbeelden van initiatieven zijn bibliometrische analyses van onderzoeks­prestaties binnen een aantal disciplines van enkele Vlaamse universiteiten, een bibliometrische sterkte-­zwakte analyse van het Vlaamse informatietechnologie-­onderzoek, een bibliometrische analyse en benchmark van het onderzoek binnen IMEC en een analyse van de relatie tussen capaciteit en kwaliteit van fundamenteel onderzoek en de capaciteit voor extern verworven onderzoek.<sup>10</sup>

Specifiek ten behoeve van de verbetering van de onderzoeks­beleids­processen zijn in 2001 en 2002 een aantal Steunpunten beleids­relevant onderzoek opgezet (die onder de

<sup>7</sup> Zie de volgende drie nota's: (1) Moerman, F. (2004), *Beleidsnota 2004-2009*. Brussel: Vlaams Ministerie van Economie, Ondernemen, Wetenschap, Innovatie en Buitenlandse Handel. (2) Van Mechelen, D. (2003), *Wetenschappen en technologische innovatie. Beleidsprioriteiten 2003-2004*. Brussel: Vlaamse Ministerie van Financiën en Begroting, Ruimtelijke Ordening, Wetenschappen en Technologische Innovatie. (3) Vanderpoorten, M. & D. Van Mechelen (2000), *Wetenschaps- en technologiebeleid. Beleidsnota 2000-2004*. Brussel: Vlaamse Regering.

<sup>8</sup> Zie: Bouckaert, G., W. Van Dooren & M. Sterck (2003), *Prestaties meten in de Vlaamse overheid: Een verkennende studie*. Leuven: Bestuurlijke Organisatie Vlaanderen.

<sup>9</sup> Zie: Larosse, J. (1997), *Theoretische en empirische bouwstenen van het Vlaams Innovatie Systeem*. Brussel: IWT-VTO, en: Larosse, J. (1997), *Het Vlaamse Innovatie Systeem. Een nieuw statistisch kader voor het innovatie- en technologiebeleid*. Brussel: IWT-VTO.

<sup>10</sup> Zie: Van den Berghe, H., J.A. Houben, R.E. De Bruin, H.F. Moed, A. Kint, M. Luwel & E.H.J. Spruyt (1998), Bibliometric indicators of university research performance in Flanders, *Journal of the American Society for Information Science* 49(1): pp. 59-67, en: Luwel, M., E.C.M. Noyons & H.F. Moed (1999), Bibliometric assessment of research performance in Flanders: Policy background and implications, *R&D Management* 29(2): pp. 133-141.

verantwoordelijkheid van de bevoegde minister vallen en bestaan uit één of meer onderzoeksgroepen uit de universiteiten). Eén van die steunpunten is het *Steunpunt voor O&O Statistieken* (SOOS). De opdrachten voor dit instituut zijn: het aanleveren van gegevens die nodig zijn om beleidsvragen te beantwoorden en een systeem van indicatoren op te bouwen dat toelaat de omvang en de internationale positie van Vlaams potentieel inzake onderzoek en ontwikkeling in kaart te brengen. Op dit moment wordt binnen SOOS met ongeveer 10 fte aan de dataverzameling, -analyse en rapportage gewerkt. In het licht van de doelstellingen van prestatiemeting die door Behn<sup>11</sup> worden onderscheiden, worden in de Vlaamse context vooral “evalueren”, “leren” en “verbeteren” benadrukt, er is minder aandacht voor enerzijds “controleren” en “budgetteren” en anderzijds “motiveren” en “overtuigen”.

### 2.3 De gemeten prestaties

Het Vlaams Indicatorenboek<sup>12</sup> geeft een overzicht van de recent gemeten onderzoeksprestaties (veelal over het afgelopen decennium) en geeft tevens inzicht in de wijze waarop deze prestaties (en door wie) worden gemeten. Hieronder volgt een overzicht:

- Menselijk potentieel (uit: databanken VLIR en departement onderwijs).
  - uitgereikte diploma’s (hogescholen en universiteiten) naar type, geslacht en studiegebied
- Publicaties en citaties (uit: *Science Citation Index* van het *Institute for Scientific Information* voor verschillende, voornamelijk technische, levens- en natuurwetenschappelijke disciplines)
  - aantallen Vlaamse publicaties;
  - aantallen Vlaamse publicaties in comparatief perspectief;
  - citatie-impact van Vlaamse publicaties voor de bovengenoemde disciplines;
  - citatie-impact van Vlaamse publicaties in comparatief perspectief;
  - internationale co-publicaties.
- Octrooien (uit: databanken zoals *US Patent and Trademark Office* en het *European Patent Office*)
  - aantallen octrooien;
  - aantallen octrooien in comparatief perspectief;
  - aantallen octrooien naar organisatietype (onderneming, universiteit, persoon);
  - samenwerkingspatronen in octrooien (co-uitvinderschap naar nationaliteit);
  - distributie van octrooien over technologiedomeinen;
  - de relaties tussen octrooien en productie in technologiedomeinen.
- Innovaties (uit: *Community Innovation Survey*, CIS van Eurostat)
  - innovatie door bedrijven (aard van de innovaties, type bedrijven, financiering van innovatie, samenwerkingspatronen en resultaten van innovaties).
- Participatie in kaderprogramma’s EU: KP3, KP4 en KP5 (uit: databestanden Vlaamse overheid)
  - deelname: categorieën deelnemers, deelname naar hoofdgebied van het kaderprogramma.
- Spin-off activiteiten van universiteiten (uit: data van de Administratie Wetenschap en Innovatie)

<sup>11</sup> Behn, R.D. (2003), Why measure performance? Different purposes require different measures. *Public Administration Review*. Vol. 63, No. 5, pp. 586-606.

<sup>12</sup> Zie Debackere, K. & R. Veugelers (eds. 2003), *Vlaams indicatorenboek 2003*. Brussel: SOOS, p. 77. Het Indicatorenboek richt zich op beleidsindicatoren in het algemeen, dus niet alleen op prestatiemeting (naast outputs ook aandacht voor inputs).



- spin-outs van universiteiten: omvang interfacediensten, aantallen spin-outs per onderzoeksinstelling, economisch belang van spin-outs en financieringsstromen voor spin-outs (waaronder startkapitaal).

De SOOS speelt een centrale rol in het verzamelen en bewerken van prestatiegegevens. Zoals uit het bovenstaande overzicht blijkt wordt een zeer groot deel van de gegevens ontleend aan bestaande databases (internationaal: SCI, patentbureaus, EUROSTAT; en nationaal: administraties/centrale overheid). Het is van belang hierbij op te merken dat hoewel het SOOS – ingesteld door de minister – vrij autonoom werkt aan het analyseren en presenteren van data, zij input ontvangt van diverse betrokkenen in het onderzoeksveld, waaronder (werkgroepen van) de Vlaamse Interuniversitaire Raad (VLIR). De VLIR buigt zich onder andere – in samenwerking met het SOOS – over de verdere uitbreiding van betrekken van bibliometrische gegevens (e.g. voor cultuur- en gedragswetenschappen; en voor publicaties buiten ISI journals). Een ander punt van aandacht is hoe – naast wetenschappelijke productiviteit en kwaliteit – maatschappelijke relevantie operationeel gemaakt kan worden.

## 2.4 Resultaten

De prestatieingen hebben een aantal ontwikkelingen in het afgelopen decennium zichtbaar en transparant (in internationaal-comparatief perspectief) gemaakt. De analyses van het SOOS (2003) stellen dat de sterke groei van de Vlaamse investeringen in het laatste decennium de prestaties hebben doen groeien in kwalitatieve en kwantitatieve zin. Dit betreft zowel prestaties op het terrein van het aantal afgestudeerden en doctoraten, de publicatie-output, de aantallen octrooien, als aantallen technologiemarkters en universitaire spin-offs. De geconstateerde samenhang tussen stijgende input en stijgende prestaties is overigens ook herkenbaar in een aantal andere Europese landen zoals Spanje en Ierland.<sup>13</sup>

Met betrekking tot prestatieingen in het begin van de jaren negentig, concluderen onderzoekers dat departementen (d.i. vakgroepen) met een hoge internationale status meer geprofiteerd hebben van externe middelen dan groepen met een minder hoge internationale impact. Tegelijkertijd heeft de verwerving van externe middelen geleid tot het aantrekken van een groter aandeel jongere stafleden, wat tot een terugval in de productiviteit heeft geleid.<sup>14</sup> In de context van de relevantie voor beleid merken Luwel c.s. op dat er vooral sprake is van indirecte effecten van prestatieingen.<sup>15</sup> De resultaten hebben bijgedragen aan het debat over prestatieingen, maar niet direct tot specifieke beleidsbeslissingen van de overheid. Van den Berghe c.s. merken (ook) op dat bibliometrische analyses vooral tot veel discussies hebben geleid, maar ook gedragsverandering: veel academici hebben hun publicatiestrategie aangepast, gekenmerkt door het vermijden van publicaties in *low impact* tijdschriften.<sup>16</sup>

### *Positieve effecten en problemen*

Zoals in veel gevallen waarin prestaties zichtbaar worden gemaakt en door stakeholders als belangrijk worden ervaren, stimuleert transparantie en vergelijking met anderen vaak tot het overwegen van maatregelen de prestaties te veranderen/verbeteren. In hoeverre daarvan sprake is in de Vlaamse context is moeilijk te zeggen. Uit onderzoek naar prestatieingen in een iets verder verleden zijn echter wel enkele lessen te trekken (zie hieronder).

<sup>13</sup> Zie: Debackere, K. & R. Veugelers (eds. 2003), *Vlaams indicatorenboek 2003*. Brussel: SOOS, p. 77.

<sup>14</sup> Zie: Luwel, M., E.C.M. Noyons & H.F. Moed (1999), Bibliometric assessment of research performance in Flanders: Policy background and implications, *R&D Management* 29(2): pp. 133-141.

<sup>15</sup> Zie: Luwel op cit.

<sup>16</sup> Zie: Van den Berghe, H., J.A. Houben, R.E. De Bruin, H.F. Moed, A. Kint, M. Luwel & E.H.J. Spruyt (1998), Bibliometric indicators of university research performance in Flanders, *Journal of the American Society for Information Science* 49(1): pp. 59-67.

### *Prestaties gebruikt voor bekostiging?*

Er wordt in het Vlaamse onderzoekstelsel slechts beperkt gebruik gemaakt van prestatiebekostiging. De verdeling van de FWO- middelen (11% van het totale overheidsbudget voor onderzoek 2003) vindt plaats op competitieve basis (kwaliteit van het projectvoorstel, beoordeeld door FWO-commissies met hulp van *peer review*). Bij de verdeling van de BOF-middelen (12% van het totale overheidsbudget voor onderzoek in 2003) tussen universiteiten spelen prestaties (diploma's en wetenschappelijke productie) een rol. In hoeverre deze prestaties een rol spelen bij de universiteitsinterne verdeling van de BOF-middelen is niet bekend (in het nabije verleden werden middelen intern vaak verdeeld op basis van beoordelingen door peers).

Sinds 2003 worden voor de interuniversitaire verdeling van BOF-middelen over de zes Vlaamse universiteiten<sup>17</sup> voor het eerst ook bibliometrische data gebruikt. Dit had nogal wat voeten in de aarde. Zo werd voor deze eerste exercitie besloten om van de drie ISI publicatiedatabases voorlopig alleen de *Science Citation Index* (SCI) te gebruiken. Dit omdat de *Social Science Citation Index* (SSCI) en de *Arts and Humanities Citation Index* (A&HCI) tot teveel controverses leidde. Vervolgens werd besloten om co-publicaties met auteurs van verschillende instellingen niet voor een deel aan de betrokken instellingen toe te rekenen, maar volledig aan elke universiteit. Dezelfde procedure werd gevolgd voor de citaties. Met andere woorden, de gegevens zijn niet meer dan 'ruwe' data – mede omdat niet met de impact factor van tijdschriften rekening werd gehouden. Opschoningen werden vooralsnog niet gemaakt en uitgesteld tot later. De publicatie- en citatiegegevens hebben betrekking op een periode van tien jaar voorafgaand aan het jaar waarin de begroting wordt opgesteld.

Het aandeel van een universiteit in de BOF middelen voor het jaar j+1 zoals berekend in jaar j is een weging van twee onderdelen die in onderstaande formule (in 2003 voor het eerst gebruikt) tussen haakjes zijn geplaatst:

$$w_1 (0,5 \text{ PhD's} + 0,35 \text{ Diploma's} + 0,15 \text{ Operationele subsidies}) + w_2 (0,5 \text{ publicaties} + 0,5 \text{ citaties})$$

De eerste term, die een gewicht  $w_1$  bezit, is de weerslag van de oude verdeelsleutel voor de BOF-middelen. De drie criteria zijn het aandeel van de betreffende universiteit in het aantal afgeleverde doctoraten (50%), het aandeel in het aantal afgeleverde tweede cyclusediploma's (35%) en het aandeel in de operationele subsidies (de 'eerste geldstroom') (15%). De drie criteria worden over een vier jarig tijdsvenster gemiddeld. De tweede term is een gemiddelde van 'ruwe' aantallen publicaties en citaties: het aandeel van elke Vlaamse universiteit in de totale Vlaamse academische publicatie- en citatie-output volgens de gegevens van de Science Citation Index gedurende een voortschrijdend tijdsvenster van tien jaar.

De gewichten van de beide termen veranderen geleidelijk totdat ze in 2005 een 'steady state' hebben bereikt. Onderstaande tabel laat dit zien:

Gewicht	BOF 2003	BOF 2004	BOF 2005	BOF 2006
$W_1$	0,9	0,8	0,7	0,7
$W_2$	0,1	0,2	0,3	0,3

Omdat het systeem nog maar recent is ingevoerd, zijn er nog maar in beperkte mate effecten zichtbaar. Als gevolg van de introductie van de bibliometrische criteria in 2003 bleek één grote universiteit duidelijk minder BOF-middelen te ontvangen en de andere grote universiteit aanmerkelijk minder dan volgens de oorspronkelijke BOF verdeelsleutel. Eén middelgrote

<sup>17</sup> Het gaat om twee grote universiteiten (Leuven, Gent), twee middelgrote (VU Brussel, Antwerpen) en twee kleine (KU Brussel, Limburg).

instelling kreeg – vanwege haar reputatie in de biomedische wetenschappen – een groter aandeel, de andere middelgrote instelling kreeg ongeveer evenveel.

Debackere en Glänzel wagen zich voorzichtig aan een aantal speculaties.<sup>18</sup> Het opnemen van aantallen publicaties kan kwantitatief gericht (in plaats van kwaliteitsgericht) publicatiegedrag uitlokken; de criteria zouden kunnen worden overgenomen in de universiteitsinterne verdeling (wat niet wenselijk is gezien de grote verschillen tussen de disciplines, met het risico van een Mattheus-effect); en de criteria zouden kunnen uitnodigen tot strategische co-publicaties. Het ongewenste gedrag (eerste speculatie) kan overigens beperkt worden door *Journal Impact Factors* mee te nemen in de scores, hoewel de auteurs opmerken dat impactfactoren geen uitdrukking zijn van de kwaliteit van een tijdschrift.

Ter afsluiting melden we dat het Vlaamse systeem rondom R&D beleid en bekostiging een grote nadruk legt op indicatoren en prestatiemeting. Er is hiertoe zelfs een speciaal instituut in het leven geroepen (SOOS). Tegelijkertijd moet echter worden benadrukt dat vooralsnog slechts een klein deel van de middelen (het BOF) op grond van enkele ‘ruwe’, uiterst gebrekkige prestatie maatstaven wordt verdeeld en dat mede hierom de gehanteerde systematiek zich nog in een experimenteel stadium bevindt. De verdeling van deze BOF middelen (ruim 90 miljoen Euro) vindt deels plaats op grond van bibliometrische indicatoren die slechts op de harde disciplines (natuur, techniek, medicijnen) betrekking hebben. Het wordt aan de universiteiten zelf overgelaten om intern andere verdeelsleutels te hanteren.

---

<sup>18</sup> Debackere, K. & W. Glänzel (2004), Using a bibliometric approach to support research policy decisions: The case of the Flemish BOF-key. Leuven: SOOS.



## 3 Measuring research performance in the United Kingdom

### 3.1 Brief overview of national research system

Public funding for academic research comes through two channels. Project-based funding comes largely from the Office of Science and Technology's (OST) science budget (The OST is housed within the Department of Trade and Industry) from which funds are allocated to eight national Research Councils.<sup>19</sup> Such funding is designed to cover both direct and indirect project costs.<sup>20</sup> The second major form of support is funding channelled through the Higher Education Funding Councils and the Department for Education and Skills (DfES). Here funding is distributed in the form of block grants that are largely designed to subsidize research infrastructure costs, indirect costs and various other fixed costs (e.g., staff, equipment, etc.). Finally a small amount of research is funded through several additional outlets such as government units, private industry, charitable foundations and the EU Framework Programs. Comprehensive statistics are not available but in the fiscal year 2003-04, the combined budget of the Research Councils was £1.9 billion (€2.71 billion), of which around 40% funds research projects within universities. In 2004-05, the Higher Education Funding Council for England (HEFCE) alone is slated to spend £1.08 billion (€1.54 billion) on research.

The hallmark of the UK's system for funding academic research is the *Research Assessment Exercise* (RAE). First undertaken in 1986, it has since gone through four additional rounds, the most recent being 2001. The RAE replaced the University Grants Committee systems of selectively allocating research funding through subject-based committees. Its primary objective is to transparently allocate funding to the best researchers through an elaborate quality-oriented evaluation procedure.

### 3.2 How is research output measured?

#### *SET*<sup>21</sup>

The OST produces a variety of statistics related to UK-wide research performance. Their Science, Engineering and Technology (SET) indicators are prepared jointly with the Office of National Statistics and rely on data collected from several public agencies.<sup>22</sup> These aim to:

- Provide a historical analysis of the Government financing of Science, Engineering and Technology (SET) activities in the UK;
- Describe the relationship between the funders and performers of Research and Development (R&D) in the UK (Government, higher education, business enterprise, charities and overseas);
- Report on business enterprise R&D expenditure;

---

<sup>19</sup> Technically there are currently seven research councils and the Arts and Humanities Research Board (AHRB). The AHRB will, however, officially become a research council in April 2005.

<sup>20</sup> The government recently announced that it will gradually move to a system where Research Councils will pay the full economic costs of research they fund.

<sup>21</sup> The list of indicators employed is expansive and grouped into seven (7) broad categories. These are all available on-line at: <http://www.ost.gov.uk/setstats/index.htm>

<sup>22</sup> The main source of the statistics is the annual Office for National Statistics (ONS) survey of R&D in Government. Other sources include ONS's annual survey of R&D in businesses; the OECD's Main Science and Technology Indicators database; the Higher Education Statistics Agency (HESA) and the Universities and Colleges Admissions Service (UCAS).

- Summarize key data on output and employment of science graduates and postgraduates, and other employment data; and
- Show how the UK compares with other G7 countries.

The SET indicators differ from traditional R&D figures in the sense that they exclude information related to financial, technical, and commercial aspects of the R&D process. Research is divided into basic, applied and experimental development categories. Recognizing the blurred boundary between basic and applied research, overlapping areas are categorized as “strategic” research. This, in turn, generates a matrix of pure- and orientated-basic research on one axis and strategic- and specific-applied research on the other.<sup>23</sup>

### ***OST Indicator Basket***

In line with the broader government push towards improving the UK’s international performance in science and engineering research, the OST and Department of Trade and Industry have commissioned a study to define sets or “baskets” of indicators and metrics showing how the UK stands globally and among what it regards as its peer group (of countries). Indicators are grouped into seven categories:<sup>24</sup>

1. Inputs (including expenditure on research)
2. Outputs (including people and publications)
3. Outcomes (research recognition, citations, training and research quality)
4. Productivity – financial (outputs and outcomes related to inputs)
5. Productivity – labour (outputs and outcomes related to other measures)
6. People
7. Business expenditure

Data is collected from multiple publicly-available sources including the OECD, various public agencies in the UK, and SET data. All data is purported to possess three attributes: 1) subject area, 2) time, and 3) location. The three primary categories are inputs, activities and outputs. Secondary indicators are meant to describe the relationships between primary category indicators. Where possible, attention is given to outputs resulting in identifiable outcomes or impacts. An example of the summary tables for the first three indicators in the list immediately above is provided in the Appendix to this chapter.

### ***The RAE***

Unlike the collections of research output information presented above, the RAE is primarily a quality-oriented exercise and this is evident in the type of information it collects and disseminates.<sup>25</sup> All publicly funded higher education institutions in the UK are invited to submit information related to research activity in 68 different assessment units. The information they provide is standardized though it does include a mix of qualitative and quantitative information. Each institution decides which subjects to submit to and, importantly, which staff to include in the submission packet. Each researcher is permitted to submit no more than four outputs. According to the official “Guide to the 2001 Research Assessment Exercise” all forms of research output (e.g., books, journal articles, etc.) are treated equally. The panels reviewing the outputs (note that they do not make site visits) are concerned only with their quality, hence total publication/output counts are not factored in.

<sup>23</sup> Detailed background information on definitions used, information sources and issues affecting the final statistics can be found on-line at: [http://www.ost.gov.uk/setstats/background\\_info.htm#item03](http://www.ost.gov.uk/setstats/background_info.htm#item03)

<sup>24</sup> Source: *PSA Target Metrics for the UK Research Base* (2004). United Kingdom Office of Science and Technology Policy, Department of Trade and Industry. Available [on-line] at: [http://www.ost.gov.uk/research/psa\\_target\\_metrics\\_oct2004.pdf](http://www.ost.gov.uk/research/psa_target_metrics_oct2004.pdf)

<sup>25</sup> All information sent in by academic units at various UK institutions is provided by HERO (Higher Education and Research Opportunities) and is available on-line at: <http://www.hero.ac.uk/rae/submissions/>

Institutions are required to submit information for each researcher and their respective department in the following areas:

1. Staff information that summarizes all academic staff, details research-active staff and lists research support staff and assistants.
2. Research output (up to four per researcher)
3. Text descriptions about the research environment, structure and policies, strategies for research development, qualitative information on research performance and measures of esteem.
4. Related data that focuses on amounts and sources of funding, numbers of research students, number and source of research studentships, numbers of research degrees awarded and indicators of peer esteem.

This information is then used by panels of experts to produce ratings of institutions' individual programs. The rating scale used in the RAE is presented below in Table 1. In addition, all universities submitting to the RAE are requested to fill out an annual *Research Activity Survey*. This form asks for supplementary information on the numbers of research students, research assistants, research fellows and the amount of research income from charities (foundations).

**Table 1. Rating scale used in the 2001 RAE**

Rating	Description
5-star	Quality that equates to attainable levels of international excellence in more than half of the research activity submitted and attainable levels of national excellence in the remainder
5	Quality that equates to attainable levels of international excellence in up to half of the research activity submitted and to attainable levels of national excellence in virtually all of the remainder
4	Quality that equates to attainable levels of national excellence in virtually all of the research activity submitted, showing some evidence of international excellence
3a	Quality that equates to attainable levels of national excellence in over two-thirds of the research activity submitted, possibly showing evidence of international excellence
3b	Quality that equates to attainable levels of national excellence in more than half of the research activity submitted
2	Quality that equates to attainable levels of national excellence in up to half of the research activity submitted
1	Quality that equates to attainable levels of national excellence in none, or virtually none, of the research activity submitted

Source: A Guide to the 2001 Research Assessment Exercise

### 3.3 Results from evaluation studies

In 2003, the UK government released a white paper<sup>26</sup> in which it hinted at rethinking how research funding is managed and distributed. Worried about international competition, the line it took was that in order for the UK to maintain its current levels of international

<sup>26</sup> Department for Education and Skills (2003), *The Future of Higher Education*, Cm 5735, London: TSO. <http://www.dfes.gov.uk/highereducation/hestrategy/>

scientific excellence research would need to be increasingly concentrated in a number of large, highly-rated units. The extent to which such intentions have already been in motion is evident in the follow-up to the 2001 exercise. Though much speculation surrounded whether 4-star rated institutions and departments would receive lower levels of funding, it was eventually decided to maintain the status quo until the 2008 RAE. The ambiguity and uncertainty that lingered during and after the debate underscored the government's commitment to acting on the white paper's major points.

In response, *Universities UK*, the organization speaking on behalf of higher education institutions in the UK (indeed they boast that all universities and a number of further education colleges across the UK are members) commissioned a study examining the further concentration of university research performance and regional capacity.<sup>27</sup> In reality though, the study really was a critical examination of the RAE. Their main finding was the absence of any solid evidence to suggest that better research would come from further concentration. What is more, it showed that many of the science-based grade 4 listed institutions were still performing above the world average.

Following the completion of the 2001 RAE, the House of Commons' Science and Technology Committee has since conducted two separate assessments of it. The first was completed in mid-2002. While it praised the RAE for what it had done for enhancing UK research it also raised a number of serious shortcomings to the process. Among others:<sup>28</sup>

1. *Accuracy of results* – The selective exclusion of researchers was one of the major concerns as the Committee argued that “funding should reflect the actual amount of research and its quality over the whole department and not those deemed active. Universities should have no incentive to omit any researchers (p. ).” Another issue was the practice of moving researchers across assessment units or merging departments could improve ratings without commensurate improvements in quality.
2. *RAE effects* – Some suggested that the RAE was imposing extraordinarily high bureaucracy costs on universities, distorting research practice and driving universities to neglect other core activities. In the same vein, the Committee also showed concern about “collateral damage” (p. 8) to staff careers and that the exercise was distracting universities from their other roles.
3. *Research practice* – Given that some forms of outstanding research may take extended periods of time to complete, some expressed worry that the current 6-year cycle would discourage the production of long-term research.
4. *Morale and careers* – “It is clear that the RAE has had a negative effect on university staff morale (p. 9).”
5. *Neglect of teaching and other university activities* – While the Committee agreed that the RAE is not designed to reward teaching they suggest that it (the RAE) must produce a counter-incentive to promoting good teaching (p. 10). Specifically there was clear concern that some universities would become teaching-only institutions or that further concentration of research would reduce the number of future scientists to be trained.

At the insistence of the four Higher Education Funding Councils an inquiry into the RAE was commissioned and conducted by Sir Gareth Roberts. In early 2004 the Funding Councils proposed a number of substantial changes based on Roberts' findings. At that point the

---

<sup>27</sup> Source: Funding research diversity: A technical report supporting a study of the impact of further concentration on university research performance and regional research capacity.

<sup>28</sup> For a full list of the Committee's recommendations see “The Research Assessment Exercise: Government response to the Committee's Second Report. Available: [on-line] at: <http://www.publications.parliament.uk/pa/cm200102/cmselect/cmsctech/995/995.pdf>



Science and Technology Committee decided to conduct a second review,<sup>29</sup> focusing on the RAE as a mechanism, in order to evaluate whether the proposed revisions would indeed correct the problems identified in the first report. This produced a number of more specific recommendations for the 2008 Exercise, including:

1. Consider establishing panel moderators to improve the recognition procedure for interdisciplinary research and strengthening the use of international panel members.
2. Provide more resources to the review panels or run the risk of individuals producing biased or wrong conclusions about research quality.
3. Panels should give equal weight to pure and applied research and universities should be notified.
4. HEFCE should instruct panels not to rely on where a publication is published as a sole measure of quality.
5. Quality criteria should concentrate more on the impact of research instead of where it is published.
6. Introduce *quality profiles* (see section 3.5, below) as a replacement for the current rating system. This would particularly help institutions/departments on the fringes of different rating points.
7. In light of strategies employed by universities to improve their RAE ratings, HEFCE should publish an analysis of those strategies being used and outline what it considers to be acceptable practice.
8. If universities are allowed to selectively choose which researchers will submit then the names of those individuals should be publicly disclosed in the name of greater transparency.
9. HEFCE should draw up guidelines for universities on how *quality profiles* will be used to calculate funding.

### 3.4 Are measurements used for funding?

The information collected and disseminated by the OST is both comprehensive and detailed. While no direct evidence was found to suggest that such information is used in national funding decisions, given that OST plays such a substantial role in funding research at the national level it seems likely that such information must implicitly shape OST funding decisions.

According to HEFCE's website, approximately 98% of the funding it distributes for research is "allocated selectively according to quality." The extent to which RAE results are used in research funding decisions can be seen in the components underlying HEFCE's funding model, listed in Table 2. Funding rates also differ by unit of analysis (the 68 different categories mentioned above). The variation between those units receiving 4 and a 5-star rating is substantial. In 2004-05, the average funding (across all units) for the former was £9,980 (£14,254) while for the latter it was £31,498 (£44,988).<sup>30</sup>

---

<sup>29</sup> "Research Assessment Exercise: a re-assessment: Government Response to the Committee's Eleventh Report of Session 2003-04." Available: [on-line] at:

<http://www.publications.parliament.uk/pa/cm200405/cmselect/cmsctech/34/34.pdf>

<sup>30</sup> Source: HEFCE (<http://www.hefce.ac.uk/research/funding/QRfunding/2004/rates0405.xls>)

**Table 2. Components in quality-related HEFCE funding model**

<b>Volume measure</b>	<b>Source of data for 2002-03</b>	<b>Source of data for 2003-04</b>	<b>Source of data for 2004-05</b>	<b>Weighting in QR model</b>
(Category A ) Research-active staff from general funds	2001 RAE	2001 RAE	2001 RAE	1
NHS-funded staff from specific funds	2001 RAE	2001 RAE	2001 RAE	1
Research fellows	2001 RAE	2002 Research Activity Survey	2003 Research Activity Survey	0.1
Research assistants	2001 RAE	2002 Research Activity Survey	2003 Research Activity Survey	0.1
Research students (weighted according to year of program)	2001 Research Activity Survey	2002 Research Activity Survey	2003 Research Activity Survey	0.15
Income from charities (average over two years)	1999-2000 income from the 2001 RAE. 2000-01 income from the 2001 Research Activity Survey	2000-01 income from the 2001 Research Activity Survey, and 2001-02 income from the 2002 Research Activity Survey	2001-02 income from the 2002 Research Activity Survey, and 2002-03 income from the 2003 Research Activity Survey	0.177 per £25,000 of income from charities

<b>Research rating</b>	<b>3b</b>	<b>3a</b>	<b>4</b>	<b>5</b>	<b>5*</b>
<b>Quality weighting</b>	0	0	1	2.793	3.362

### 3.5 Recent developments and experiences

The RAE is generally agreed to have had a significant positive impact. The exercise has driven a sustained improvement in the overall quality of the UK research base. It has highlighted the very best research and has encouraged HEIs to take a rigorous approach in developing and implementing their own research strategies. It has enabled the Government and funding bodies to maximise the return from the limited public funds available for basic research. At the same time, the exercise has been subject to some criticism. Concerns have been expressed that the exercise:

- favours established disciplines and approaches over new and interdisciplinary work
- does not deal well with applied and practice-based research in particular
- places an undue administrative burden on the sector

- has a negative impact upon institutional behaviour as HEIs and departments manage their research strategies, and shape their RAE submissions, in order to achieve the highest possible ratings.

After the 2001 exercise there were also concerns that, with over half of all submitted work divided between the top two points on a seven-point scale, the ratings produced by the exercise could no longer provide the degree of discrimination required by a policy of selective funding. The main points from the reviews were that there was overwhelming support for an assessment process built around expert review conducted by discipline-based panels. There also was strong support for the proposal to replace the rating scale by a quality profile. Because there was support for the principle that the assessment process should be designed better to recognise excellence in applied and practice-based research, in new disciplines, and in fields crossing traditional discipline boundaries, the new RAE process will be designed to recognise excellence in applied research, in new disciplines and in fields crossing traditional discipline boundaries.

This new UK-wide RAE is to be completed in 2008. Like previous exercises, it will be based upon expert review by discipline-based panels. However, based on the above-mentioned reviews and parliamentary discussions, a number of significant changes were announced. One of the main points<sup>31</sup> is that the rating scale used in 2001, under which departments were assigned grades (1-5\*), will be replaced by a *quality profile*. This will identify the proportions of work in each submission that falls in each of four defined 'starred' quality levels. Work judged to fall below the lowest level will be 'unclassified'. (See Table 3 below.) There are some problems to be expected with the distinctions between ratings, but these are left to the panels to cope with.

**Table 3. Sample quality profile\***

Unit of assessment A	FTE staff submitted for assessment	Percentage of research activity in the submission judged to meet the standard for:				
		four star	three star	two star	one star	unclassified
University X	50	15	25	40	15	5
University Y	20	0	5	40	45	10

\* The figures are for fictional universities. They do not indicate expected proportions.

#### Definitions of Quality Levels

Rating	Definition
Four star	Quality that is world-leading in terms of originality, significance and rigour.
Three star	Quality that is internationally excellent in terms of originality, significance and rigour but which nonetheless falls short of the highest standards of excellence.
Two star	Quality that is recognised internationally in terms of originality, significance and rigour.
One star	Quality that is recognised nationally in terms of originality, significance and rigour.
Unclassified	Quality that falls below the standard of nationally recognised work. Or work which does not meet the published definition of research for the purpose of this assessment.

<sup>31</sup> See <http://www.rae.ac.uk/pubs/2004/01/rae0401.doc> , published by the joint funding bodies in 2004.

In any case, the use of quality profiles is expected to reduce the tactical element in preparing submissions. The incentive will be for institutions to include all their good researchers rather than aiming for a particular grade. The new method will also benefit institutions with comparatively small pockets of excellence within a larger research unit, as the true scale and strength of their best work will be more visible.

#### *costs*

The issue of the costs of running the RAE has proved to be a contentious one. Calculating the direct additional costs to the higher education funding bodies is comparatively straightforward. The funding councils estimated that costs for the 2001 exercise came to some £5.6 million. The largest element in this total was costs related to panel meetings, including members' fees. However, calculating the cost to HEIs has proved more problematic. It is possible to estimate the time that academic staff engaged in the assessment exercise are not available to carry out their normal academic duties, and to calculate this as an opportunity cost. It is also possible to estimate the amount of academic and administrative staff time devoted to preparing submissions for assessment. But here one runs into difficult questions about how much of this work a well-managed HEI or research department would undertake in any case (for example, keeping departmental and institutional research plans up to date, and maintaining information about the research activity and outputs of individual staff).

A survey of the costs to the sector of the 1996 exercise, based upon a survey among universities, produced an estimate of some £30 million. A later study of costs was carried out in one research-intensive HEI. This produced an estimate (including opportunity costs) of £37.5 million for all universities in England – or 0.8 per cent of the total funds allocated on the basis of the RAE's results. These estimates were argued to reflect the amount of work that universities need to undertake for the exercise, over and above what might otherwise be expected of a well-managed institution.

The results of the next exercise will probably be used to allocate some £8 billion of research funding across a six-year period. If the total cost were £45 million, this would represent around 0.6 per cent of the resources allocated, comparing very favourably with the costs associated with project-based grant allocations by the research councils, using expert review.

### Appendix to chapter 3: Some of the OST indicators

Indicator number	Description of performance indicator	Condition signalling improvement	Level of disaggregation	Primary data sources
<i>THEME 1 INPUTS including expenditure on research</i>				
1.01	GERD relative to GDP (R&D intensity)	Increased proportion of R&D specific spend	System	OECD MSTI 2003-2
1.02	Publicly performed R&D as proportion of GDP	Increased proportion of R&D specific spend	System	OECD MSTI 2003-2
<i>THEME 2 OUTPUTS including people and publications</i>				
2.01	Number and share of OECD PhD awards	Increased count & increased share by comparison with competitors	System	OECD Education Database
2.02	PhDs awarded per head of population	Increased ratio	System	OECD MSTI 2003-2/OECD Education Database
2.03	Number and share of world publications	Increased count & increased share by comparison with competitors	System	ISI National Science Indicators 2003
<i>THEME 3 OUTCOMES including research recognition and citations; training and research quality</i>				
3.01	Number and share of world citations	Increased count & increased share by comparison with competitors	System	ISI National Science Indicators 2003
3.02	Number & share of world citations in 9 main research areas	Increased national count and share	SUoA	ISI National Science Indicators 2003
3.03	Rank of share of world citations by 9 main research fields - frequency in top 3	More frequent presence in top three among fields	System	ISI National Science Indicators 2003



## 4 Measuring research performance in Germany

### 4.1 Brief overview of national research system

Germany has an institutionally divided research system within which decision making is shared between the federal level governments and state governments. The German science system is one of the largest science systems in the OECD and the structure rests on three basic pillars: (1) industry, (2) higher education institutions, and (3) public research institutes. Most academic research is conducted either in organised research institutes or in the universities (Wissenschaftliche Hochschulen). Some research is also carried out in the polytechnics (or Fachhochschulen) which are, however, mainly teaching institutions.

Universities are generally regarded as the central players in the system. There are some 82 universities. Public research institutes are organized into four expansive networks:

- the *Max Planck Gesellschaft* (MPG)
- the *Fraunhofer Gesellschaft* (FhG),
- the *Helmholtz-Gemeinschaft Deutscher Forschungszentren* (HGF)
- the *Wissenschaftsgemeinschaft Wilhelm-Gottfried-Leibniz* (WGL).<sup>32</sup>

On a much smaller scale there are several state based scientific academies, like the *Bayerische Akademie der Wissenschaften*, which receive both federal and state funding to coordinate research programs.

National policy coordination and advice comes through two organisations that span the federal-state divide. One is the *Bund-Länder-Kommission für Bildungsplanung und Forschungsfoerderung* (BLK), which is where federal and state ministers coordinate research policy and planning, though notably on formulae for jointly funded parts of the research system. The other is the *Wissenschaftsrat* (WR), an independent body appointed by the federal president to represent the needs of all relevant stakeholders. The WR has no funding role.

Universities are primarily financed by their state authorities, from which they receive basic funding for teaching and research. Especially in the recent past, many universities have increased their activities to get project funding. The most important source to finance universities on the basis of project funding is the non-profit organization, *Deutsche Forschungsgemeinschaft* (DFG), which is active in the whole range of disciplines – from engineering to arts and humanities. The DFG receives funding from both the federal and state governments based on a fixed formula.<sup>33</sup>

Federal research policy and funding falls under auspices of the Federal Minister for Education and Research (*Bundesministerium für Bildung und Forschung* – BMBF). The BMBF directs federal funds to higher education and public research institutes and is the main player in

---

<sup>32</sup> The MPG is a non-profit sector of 81 institutes that does basic research concentrating on cutting-edge interdisciplinary research. Its work is seen as being supplemental to university-based research. The FhG consists of 56 institutes that focus heavily on applied research in engineering and the natural sciences and works closely with industry. The HGF is an association of 15 large research institutes focusing on basic research in the physical, biomedical and technical sciences. Researchers in other sectors often use its facilities. The WGL is an association of 78 non-university institutes that are generally small, serve broad regional interests and which is funded directly by the government from what is known as the “Blue List”, a funding instrument agreed jointly between federal and state authorities.

<sup>33</sup> The newer formula, introduced in 2002, provides 58% from the federal government and the remainder from the states.

enacting relevant policy. State research policy and funding generally goes to higher education institutions, particularly through their annual block grant allocations.

Finally, a small (yet growing) number of private non-profit foundations provide a small percentage of German R&D funding. An example is the *Stifterverband für die Deutsche Wissenschaft*. They focus primarily on basic research but also support research cooperation between universities, business and public research institutes.

## 4.2 How is research output measured?

Where research performance measures have found their way into government produced publications, it is generally through the indicators collected at the instigation of the OECD.<sup>34</sup> The DFG, on the other hand, has begun to take more than a passing interest in research assessment. In the absence of a German equivalent to the UK research assessment exercise, the DFG produces university rankings on the basis of DFG grant allocations and a wide array of other criteria. The latest ranking was published in 2003 (the other two being in 1997 and 2000) and also provides international comparisons. The objective of the study is opaque but would seem to be an internal quality assurance mechanism that, given the interest in research rankings, has also been made public. Data collected comes from multiple sources including the BMBF, other research funding organizations in Germany (see section 4.1), the Federal Statistical Office, the DFG's own data and bibliometric studies done by research units in Switzerland and the Netherlands. Both the university and individual faculties are treated as the unit of analysis. The methodology behind the compilation of the statistics and figures is self-admittedly complex and the reader is referred to the source documentation for further information.<sup>35</sup> That said, the key performance indicators looked at in the study include the following (by university):

1. Number of DFG approved grants between 1999 and 2001
2. Number of professors
3. Number of scientists and academics in total
4. Total third party funding
5. Centrality in networks of DFG-funded coordinated programs
6. Number of DFG reviewers
7. Number of Alexander von Humboldt visiting researchers
8. Number of DAAD scientists and academics
9. Number of DAAD students/graduates
10. Participation in the European Union 5th Framework Program
11. Publications in international journals

The weekly magazine Stern annually publishes the Center for Higher Education Development's (CHE) ranking of universities, which boasts that some 250 institutions, 130,000 students, and more than 16,000 professors have taken part in their surveys.<sup>36</sup> Academic faculties are ranked according to criteria like facilities, student satisfaction, and research. Unlike the popular ranking systems currently in place in countries like the US, results are not grouped together to create a list of "best" and "worst" institutions but instead use a rough tier system that avoids what has been referred to as "over precision." The CHE rankings also gather and collect statistics on numbers of dissertations produced, bibliometric

---

<sup>34</sup> See for example, the publications, Germany's Technological Performance 2002 or Facts and Figures Research 2002 (both written for the Federal Ministry for Education and Research).

<sup>35</sup> Source: Deutsche Forschungsgemeinschaft Funding Ranking 2003 Institutions – Regions – Networks: DFG Approvals and Other Basic Data on Publicly Funded Research.

<sup>36</sup> The results listed in English can be found at the DAAD website. The results in German can be found at: <http://www.dashochschulranking.de/>



figures (which are weighted), numbers of promotions in the preceding four semesters and numbers of patents (particularly for engineering).<sup>37</sup> More specifically, under the heading research one can find indicators for faculties at a number of institutions including:

1. Citations per publication
2. Doctorates per professor
3. Patents per professor
4. Professors' judgment of the research situation
5. Publications per academic
6. Publications per professor
7. Qualifications to teach at professorial level per 10 professors
8. Third party funds per academic
9. Third party funds per professor

The detailed methodology behind how the data is converted into indicators can be found at CHE's Internet website, <http://www.dashochschulranking.de/>.

#### 4.3 Results from evaluation studies<sup>38</sup>

By and large, the use of explicit performance indicators has only begun to make headway in the German research system. Indeed, in the DFG's 2003 rankings publication the presidents of both the DFG and the Association of Universities and other Higher Education Institutions in Germany (*Hochschulrektorenkonferenz – HRK*) go so far as to say that, "the identification of institutions and centers of academic excellence has become a fixed component of science policy in Germany, a condition that will not likely change in the foreseeable future" (p. 5). Perhaps it is fitting then that the DFG has taken the lead in this area and commissioned a currently ongoing study into the development of such performance indicators.<sup>39</sup> As the project summary lists:

...the connections between governance and performance are analyzed in various projects and partial projects. The fundamental question arises, which indicators are appropriate for the measurement of performance and how to collect them. In this project patent and publication indicators will be generated on the one hand and made available to the other members of the research group. Furthermore, suggestions for other indicators will be elaborated and centrally gathered for the research group and evaluated, in order to find out which indicators or combinations of indicators are especially meaningful and suitable for application in practice. The secondary question is to what extent different indicators should be gathered in diverse disciplines.

The DFG is currently debating the merits of establishing a specific institute for scientific information (research information and quality assurance) which will be responsible for the assessment and quality assurance of the DFG's funding programs.<sup>40</sup> This will create a much broader monitoring system than the reliance on the tri-annual DFG research rankings. In all likelihood, those familiar with the discussion suggest that the establishment of such an

---

<sup>37</sup> The German Academic Exchange Service (*Deutscher Akademischer Austausch Dienst – DAAD*) provides the CHE's rankings in English.

<sup>38</sup> Much of the information contained in this section came from a telephone conversation with Dr. Ulrich Schmoch at the Fraunhofer Institute for Systems and Innovation Research.

<sup>39</sup> Source: [http://www.isi.fhg.de/ti/Projektbeschreibungen/us\\_performance\\_e.htm](http://www.isi.fhg.de/ti/Projektbeschreibungen/us_performance_e.htm). Research title: Performance Indicators for Research Institutions, in Particular Research Groups.

<sup>40</sup> Source: [http://www.britischebotschaft.de/en/embassy/r&t/notes/rt-fs003\\_DFG.html](http://www.britischebotschaft.de/en/embassy/r&t/notes/rt-fs003_DFG.html)

institute will probably not occur for another three to five years. The basic problem with the creation of such a centralized institution is how geographically large Germany is. If one were to take into account the evaluation of all research centers and research units, the total number would be on the order of 15,000. Sheer size will make bibliometric studies and comprehensive information collection extremely difficult, especially in light of staff members' mobility.

There is still much debate about what the most appropriate indicators for research are. Much of this discussion has stemmed from the work of the HRK, who organizes various workshops on evaluating research and what are appropriate indicators. At the macro-level the available evidence would suggest that the German science system performs very well. To many this means that there is no current crisis and no need to tie funding to performance indicators. Apart from this is a much broader concern, mainly from professors and scientists over how to accurately capture ethereal notions like creativity or originality. This has led many research performance indicator opponents to oppose nearly every type of performance measure that has been introduced. Professors and scientists have a clear preference for peer review. By and large, the opposition in Germany is much stronger than it is in countries like the Netherlands, United Kingdom, Switzerland and Austria.

As it turns out, university chancellors have been the main promoters of introducing and using research performance measurement systems. They do so mainly because they are the ones who must make the difficult financial decisions. Some suggest this clash will generate contentious debate between professors and chancellors in the coming years; performance measurement systems will indeed come but not be widely accepted. Those opposing indicators do so because it highlights deficiencies. Promoters, on the other hand, suggest that peer review works specifically to mask such deficiencies.

In November 2004, the German Science Council (*Wissenschaftsrat*) published an extensive report on standards for research rankings and outlined what could be considered "appropriate" indicators and whether it is even possible to establish such standards.<sup>41</sup>

#### **4.4 Are measurements used for funding?**

Higher education institutions receive a disproportionately large share of their financial support from the *Länder*. For a long time the criteria that the individual Land used were not very clear. However, because institutions today have greater autonomy to internally allocate funds than in the past, the development of research indicators has become increasingly important. It is likely that universities rely on some sets of internal indicators for their own internal allocation schemes.

While government units like the BMBF produce technology indicators it is not readily apparent from the available documents that such data figures into the allocation of any types of public funding, at either the federal or state levels. The same can be said for the DFG's rankings, though this is not surprising given that DFG funding is distributed competitively rather than through block grant allocations. Some suggest that approximately one-third of university income is from the DFG and thus it provides a relatively good performance indicator of institutional research. The CHE's rankings also do not help guide funding decisions; in practice CHE even states that the figures it produces cater more to prospective students seeking the best fit between themselves and an institution offering the program of their choice. In sum, widely published federal-level research performance measures are not directly tied to funding decisions.

---

<sup>41</sup> See the *Wissenschaftsrat* report, *Empfehlungen zu Rankings im Wissenschaftssystem Teil 1: Forschung*. Available [on-line] at: <http://www.wissenschaftsrat.de/texte/6285-04.pdf>

In some federal states, PI measures were intensified and do shape funding decisions. The two most notable examples are in the regions of North Rhine/Westphalia (*Nordrhein Westfalen*) and Lower Saxony (*Niedersachsen*). The former has relied on a simple, two-indicator scheme: external funding and numbers of PhDs. If universities performed poorly in both then funding was reduced. Lower Saxony has its own Academic Advisory Council (AAC), the *Zentrale Evaluationsagentur der niedersächsischen Hochschulen*, which was established in 1997-98 and began conducting evaluations of research in 1999.<sup>42</sup> The AAC was established not to redress any readily apparent shortcomings in the system but to encourage general improvement. The AAC has three main functions (research evaluation being the largest) and provides advice to state politicians on topics related to science and research.

Research assessments focus on both disciplines and institutions by linking systematic peer review with a number of indicators. The performance criteria are based on the quality and relevance of the research done as well as effectiveness and/or efficiency. The indicators used are typical for international research procedures: publications, funds for research projects, networking (how is the international network) and numbers of graduate students. Informed or evidenced based peer review is used to get richer information (through self reports) from universities before evaluation is conducted.

The research assessment does not directly tie outcomes to funding. However, there is an indirect link. Target agreements (*Zielvereinbarungen*) exist between universities and the State Ministry related to broad objectives of the university. The findings from assessment exercises are being incorporated into target agreements, which inform funding. This is a new idea of steering and reflects a big change in the relationship between government and the universities towards more of a partnership.

While Niedersachsen provides an illustrative example and in a way stands out because of the system in place, most of the other German states also have reform commissions and are thinking about the same problems. Discourse is comparable across German states.

---

<sup>42</sup> My thanks to Christof Schiene at the Wissenschaftliche Kommission Niedersachsen (WKN Hannover) for providing me with the useful information on the Niedersachsen system.



## 5 Measuring research performance in Norway

### 5.1 The Norwegian academic research system

The Norwegian research system can be divided into three levels: the political level, the strategic level and the performing level.<sup>43</sup> With regard to the political level, the Government (cabinet) and Parliament (*Storting*) are responsible for formulating the objectives, priorities and structure of the Norwegian research system. The various ministries are responsible for promoting and financing research within their respective areas, with the Ministry of Education and Research (MER) being the overall coordinator of research policies across sectors. MER provides close to half of the overall publicly-provided research funds. To ensure proper coordination at this level, special fora such as the Government Research Policy Board (*Regjeringens forskningsutvalg*), and the Interministerial Committee on Research and Development (*Departementenes forskningsutvalg*) have been established.

At the strategic level, the Research Council of Norway (RCN) is the most important institution as the principle adviser to the cabinet and parliament on research policy. Norway has only one research council. The council funds basic research, applied research and industrial research. Approximately one third of Norway's public sector research investment is channelled through the RCN. In addition to the RCN, there are a number of strategic bodies that focus on innovation and industrial development.

The performing level for research in Norway is usually considered to consist of three sectors:

- *The industrial sector*: Almost half of all Norwegian R&D expenditure is allocated to industrial research. Because Norwegian industry is mainly focused on raw materials exploitation, with traditionally low R&D investments, government aims at encouraging increased private research investment.
- *The institute sector*: This includes research institutes and other R&D-performing units outside the higher education system. There is an extensive network of more than 200 research institutes (70 purely devoted to research). Approximately one fourth of the total Norwegian R&D resources are spent in this sector. The RCN has the strategic responsibility for the institute sector through promoting a cohesive policy for the overall sector.
- *The higher education sector*: The higher education sector comprises 6 universities (including university hospitals), 5 specialised universities and 25 state university colleges.<sup>44</sup> More than one fourth of all R&D takes place in the higher education sector, mainly within the universities and the specialised university institutions. The higher education R&D funds mainly come through the university operating budgets, with supplementary funding from RCN. In recent years, contract research has become more important.

#### *Funding of research*

Norway aims to increase its R&D efforts to average OECD levels (measured as % GDP) by 2005. Research funding consists of a number of components.

The first component concerns the general research component for the general funding of the universities and the other higher education institutions. Until recently, this was primarily related to student numbers (making up about 70% of the university funds available for R&D).

---

<sup>43</sup> See: RCN (2003), *Report on Science & Technology Indicators for Norway 2003*, Oslo: Research Council of Norway, <http://www.forskningsradet.no>, and: RCN (2004), *Research policy*, Oslo: Research Council of Norway, <http://www.forskningsradet.no>.

<sup>44</sup> As from January 1 2005, one of the former specialised universities has changed its status to a regular university and one former state university college has also become a regular university.

In 2003, a new funding model was put in place. The model comprises three components: core funding (60%); teaching (25%); and research (15%). The research allocation, which will be redistributed among institutions, is based partly on performance and partly on strategic considerations. The performance-related research allocation will be distributed between universities based on degree completion specified by level, and distributed between colleges based on credit production and external co-operation. Additional measures of performance for both types of institutions include the number of posts (e.g. professorships) and competitive funding attracted from the EU and the RCN. The strategic element is still under development (fall 2004), but is intended to reward research of high quality and relevance and stimulate the institutions to develop research strategies that support the national objectives. The funds for the higher education sector are allocated directly to the institutions by the Ministry of Education and Research.

The second source of research funding is competitive grants for proposals that are judged to be of high quality in a process of peer assessment by the RCN. The RCN distributes public funding through more than 130 research programs and other activities with an increasing focus on large-scale programmes across traditional disciplinary boundaries. RCN allocates research funds for research programmes (strategic, targeted research efforts within a specified time-frame), for independent projects not affiliated with larger research programs, and for framework grants to research institutes to promote long-term professional development.

In addition, there are a few other ways of public research funding. First is the Centres of Excellence program, supported by the Research and Innovation Fund, by which RCN selects and funds centres of excellence as a focus for quality research. Second is a government endowed research fund to boost stable and long-term multidisciplinary research. Third, in order to attract increased private R&D funding, the government provides tax credits in stimulating private investments for certain types of R&D projects. Fourth, the ministries are required to increase their research investments. Finally, RCN provides funds for large-scale strategic programs across traditional disciplinary boundaries and aimed at stimulating stronger co-operation between research performers.

*Basic information on financing of R&D in higher education in Norway*

Table 1 provides some basic facts on the relative importance of different resources available for R&D in Norwegian higher education, including basic grants from government, research council funds and contract revenues from business and international sources.

Table 1: Key indicators for R&D and innovation in Norway (1995-2001, Mill. NOK and %)

	1995		1997		1999		2001	
	in NOK	%	in NOK	%	in NOK	%	in NOK	%
Government	3145.9	76	3592.6	74	4295.3	74	4418.3	70
RCN	579.5	14	680.8	14	753.7	13	997	16
Business	219.8	5	254.3	5	295.6	5	365.2	6
Abroad	64.5	2	128.8	3	168.1	3	169.7	3
Other	129,4	3	189,3	4	306,7	5	324,0	5
Total	4139.1		4845.8		5819.4		6274.2	

Source: R&D statistics NIFU STEP/SSB

Notes: Financing of R&D expenditures in higher education sector, not total financing. Statistics for 2003 not yet available. Government funding is basically basic grants, but a minor part can be contracts. Financing from business can be assumed to be contract funding. Financing from abroad is 60 percent from EU Commission (2001).

## 5.2 Measuring research output

In discussing the nature and role of measuring research performances in Norwegian higher education we distinguish between four major areas of measuring research outputs:

- Performance orientation in general funding of research
- Performance orientation in RCN-funded research programs and projects
- Research evaluations by RCN, including RCN-funded projects and institutes as well as disciplinary reviews
- General provision of R&D data

### *Performance orientation in general funding of research*

The performance-related research funds (15% of the total operating grants for universities) are allocated between universities based on degree completion specified by level, and redistributed between university colleges based on credit production and external co-operation. Additional measures of performance for both types of institutions include the number of posts (e.g. professorships) and competitive funding attracted from the EU and the RCN. But this mix of indicators is presently being changed and will now include a component that will be more directly focused on the output of research as measured in the number scientific and scholarly publications each year. The reason for this change is that the present financing model was not focused on results and quality in research while there was an incentive system for education. This was taking away the focus on research at the institutions.

Publications will be weighted according to publication form (articles, monographs) and publication channel (scientific journals, series, web-sites and book publishers). Both international publications and those in Norwegian are included in the bibliometric analyses. Some specified publications channels will be given more weight than others to create incentives towards quality in different areas of research. The relative weights are presented in table 2.

Table 2: Relative weights attached to types of publications important for research funding

Publication type	Level 1	Level 2
Scholarly books (ISBN)	5	8
Articles in series and periodicals (ISSN)	1	5
Articles in anthologies (ISBN)	0,7	1

A common documentation system for scientific and scholarly publishing has been introduced in the higher education sector in 2004.<sup>45</sup> In this system, publications are registered on the individual and department level and can be aggregated on any other level. The budgeting model is based on aggregated counts on the institutional level. When aggregated, co-publications between authors/departments/institutions are shared between them in the calculation.

There is no distinction being made between disciplines. All disciplines relate to the definition of scientific and scholarly publications and to the same model described above, but there are three different sets of rules for the nomination of publications for level 1 or level 2. These sets of rules have been made for three major groups of disciplines. As examples, neurology and physics are in group 1, economics and mathematics in group 2 and history and sociology in group 3.

<sup>45</sup> See: Universitets- og høgskolerådet (2004), Vekt på forskning. Nytt system for dokumentasjon av vitenskapelig publisering. <http://www.uhr.no/utvalg/forskning/vitenskapeligpublisering/DokumentasjonavvitpublSluttrapport121104.htm>

Next to the bibliometric criteria, external financing from other sources remains a part of the financing model for research (external co-operation) and will continue to be included. It now is also being considered to include patents and other indicators of innovation and commercialisation somewhere in the future.

Because the change to a performance measurement system has only been made recently, there is no hard information available on the outcomes in terms of changes in the relative funding levels between institutions. However, it is expected that the re-distributive character of the new funding model will be weak, at least in the beginning. At first it is important that institutions will get confidence in the new documentation system. The experience from research in the hospital sector, where a similar model was introduced two years ago, is that there was an overall increase in the funding along with redistribution of funds.

As for potential problems with the new system, it can be stated that there were some problems earlier on with regard to the thoroughness and objectivity of the publications data. However, these problems have now been solved by using bibliographic data sources (such as the journal indexes from Institute from Scientific Information – but not only those) and authority records that standardise names and codes for publications channels. The new system has been costly to develop, but will not be costly to maintain.

Finally, higher education institutions have been supportive to the idea of the new funding scheme. The system may seem relatively complex to administer, but now that the system is running, it is not too complex or cumbersome at the central and the institutional level. It is important to note that the process is part of a more widespread production of publication lists for several purposes (CV's, annual reports, applications, funding) into one common database with uniform rules. It takes one person per five thousand researchers to run the system, and the researchers themselves are being relieved from a workload that they had earlier on.

#### *Performance orientation in research council funds*

RCN allocates competitive grants for proposals that are judged to be of high quality by the relevant research community based on peer assessment. The major objectives of the RCN grants are to maintain a high quality research infrastructure, to stimulate innovation, strategic cooperation and to serve the needs of Norwegian society. Performance evaluations must show whether the RCN allocations contribute to goal achievement, quality and efficiency. Evaluations provide a status report and offer recommendations for potential changes to be implemented in the programs.

First of all, many different types of applications can be made for RCN research funds, ranging from proposals for single projects, single researchers, for entire research institutions, visiting grants, special events, networks, etc. The assessment criteria vary across application types. As such there is a long list of criteria, including the following:

- scientific merit
- research content
- degree of innovation
- project management
- the research group
- candidates for grants/fellowships
- risk
- feasibility
- international cooperation
- national collaboration
- communication of results
- strategic significance
- socio-economic utility value
- commercial relevance
- economic value
- additionality
- the environment
- ethics and equal opportunity
- relevance relative to the call for proposals
- research-related importance
- general project quality
- competence building
- national division of labour
- relevance to enabling technologies
- relevance to society



### *Research evaluations by RCN*

According to the statutes of the RCN, evaluation activities are one of the organization's ten main areas of responsibility.<sup>46</sup> The RCN is required to "implement and follow up the evaluation of research and research institutions". The evaluation activities conducted by the RCN aim at enhancing the quality, efficiency and relevance of the research sector. Evaluation activities also help to provide a good basis for determining how to allocate research funding, and for offering guidance on research-related issues internally within the RCN, to the various institutions, and to the authorities. In addition, evaluation activities may inform research policy, research strategy and research performance. Key issues to be examined include:

- Does the research lead to the desired and planned results?
- Is the research being carried out in an efficient manner?
- Is the research up to international standards?

The research evaluations concern both disciplinary and institutional reviews. These are used to determine a starting point for basic funding of the research institutes, and also to enhance institutes' ability to plan research. The disciplinary and institutional/departmental reviews have a number of objectives:<sup>47</sup>

- Offer a critical review of the strengths and weaknesses of the field, both nationally and at the level of individual research groups and academic departments. The scientific quality of the research will be reviewed in an international context.
- Identify departments which have achieved a high (potential) international level in their research,
- Identify areas of research that need to be strengthened
- To assess the situation regarding recruitment in the scientific field
- Provide institutions with the knowledge required to raise their own research standards
- Provide institutions with feedback on the scientific performance of individual departments, as well as suggestions for improvements and priorities
- Improve the knowledge base for strategic decision-making by RCN
- Represent a basis for determining future priorities, including funding priorities, within and between areas of research
- More generally, evaluations are to reinforce the role of RCN to advise Norwegian Government and relevant ministries.

Based on these objectives, the evaluation panels take into account a wide range of issues to be reviewed for the whole discipline, but also for the individual departments and research units. These are:

- Scientific quality and relevance, with regard to: International position, quality of the departments and appropriateness of their funding, strong and weak areas, relevance of the research.
- The institutional situation with regard to: the organisation, academic career structure, scientific leadership, gender and age; the graduate and postdoctoral contacts, training and mobility; national and international contacts and collaboration;
- Financial support: The general financial situation of the discipline and specific units; the balance between positions, projects and equipment
- Interchange of knowledge and technology between clinical practice and industry
- Specific discipline-related issues and questions
- Future developments and needs

---

<sup>46</sup> RCN (1997), *Evaluation strategy*, Oslo: Research Council of Norway, <http://www.forskningradet.no>.

<sup>47</sup> The objectives and major performance categories/criteria used in the research evaluations are deduced from studying a few research evaluation reports published by RCN.

To measure the relative performance of the scientific field / discipline and the individual research units and institutions, the following criteria are used:

- International publication in best journals
- International front position
- Originality of research
- Conceptualisation of own research within framework of public health research
- Total publication activity
- Success in academic training
- Relevance and influence of research – internationally, nationally
- Overall impression of research group/institute

Basically the RCN conducts the evaluation activities. The RCN implements and follows up evaluation activities within individual disciplines and research institutes in collaboration with the institutions involved. However, the responsibility for professional evaluation is assigned to the Research Boards. This arrangement is based on the specific expertise possessed by the various Research Boards in their relevant fields, either directly or through subordinate agencies. Regardless of the general framework of objectives and general criteria, several Research Boards have adopted their own strategies for evaluating the level and development of various fields, and for initiating specially organized, goal-oriented efforts (programs).

The quality reviews of research are complex because it is particularly difficult to come to final decisions and not to give biased opinions.<sup>48</sup> The costs of the RCN R&D evaluations amounted to about NOK 5 million (€600.000) per year in 1993-1994 for about 40 different evaluations. This corresponded to only 0.22% of RCN's total budget for 1994. This investment level did not measure up to international standards (about 1% - 2% of research expenses). No recent information is available.

#### *General provision of R&D data*

Finally, Norway collects a lot of information and indicators on Science and Technology. For example one can refer to the many NIFU / STEP publications, or the bi-annual Report on Science & Technology Indicators for Norway from RCN.<sup>49</sup> Many of such indicators are used for OECD purposes, but also at the national level. The reports, however, mainly focus on the national level rather than the institutional level. The indicators include all sorts of information, clustered in a number of major categories:

#### *Resources for R&D and Innovation*

R&D expenditure: total; as % of GDP; per capita; funded by government and industry; in higher education; innovation costs

#### *Human resources*

Percentage of population with higher education degree

R&D FTE workers per 1000 capita

R&D FTE per qualified researcher/scientist per 1000 capita

Percentage of PhD holders among qualified researchers/scientists

Percentage of women among qualified researchers/scientists

#### *Cooperation in R&D and innovation*

Extramural R&D expenditure compared to intramural R&D expenditure in industry

Percentage of all innovative companies involved in cooperation in innovation

Articles in international scientific journals co-authored by Norwegian and foreign researchers as % of all articles by Norwegian researchers

<sup>48</sup> See: Langfeldt, L. (2004): "Expert panels evaluating research: decision-making and sources of bias" *Research Evaluation*, Vol 13(1), pp. 51-62.

<sup>49</sup> RCN (2003), *Report on Science & Technology Indicators for Norway 2003*, Oslo: Research Council of Norway, <http://www.forskingsradet.no>.

*Results from R&D and innovation*

Percentage of innovative companies in industry

Number of articles in international scientific journals per 100000 capita

Number of patent applications from Norwegian assignees (domestic) per 100000 capita.

The information on such indicators is drawn from several sources, such as general national statistical databases as produced by NIFU/STEP, as well as from international databases, like OECD and Eurostat.



## 6 Measuring research performance in Australia

### 6.1 The Australian academic research system

In the university sector of the post-secondary education system in Australia there are 44 institutions, most of them called ‘university’ (four are specialist institutions). Forty of these receive government (i.e. federal, that is Commonwealth) funding under the Higher Education Funding Act (HEFA 1988) – either on a triennial (i.e. three-year), or contract basis. The higher education sector is dominated by the public universities, that cater to almost 97 percent of the total student load (554,000 students) in higher education.

In Australia, education is a responsibility shared by the Federal Government and the State and Territory Governments. The ministers of education meet in a Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA). The statutes under which universities operate are prescribed by the individual States. In higher education, the influence of the federal government is substantial because it is the major source of finance.

Research activity is widely distributed across the university sector, but 5 of the 36 universities receive nearly half the research income. New government policies<sup>50</sup> that cover research and research training in Australian universities were introduced in 1999 following consultation on a discussion paper by the Department (i.e. ministry) of Education, Science and Technology (DEST).<sup>51</sup> In December 1999, the government introduced new funding arrangements for higher education research. The arrangements were designed to encourage institutions to be more flexible and responsive in developing a strategic portfolio of research activities and research training programmes.

#### *Funding of research – basic data*

The system involves a performance-based system for block funding of university research and research training activities administered by DEST and peer-review competitive grants administered by various research granting agencies, of which the largest are the Australian Research Council (ARC), funded by the Department of Education, Science and Training, and the National Health and Medical Research Council (NHMRC) within the Department of Health. Funding for performance-based programmes is appropriated through HEFA.

Commonwealth funding for university research operates through a “dual system”, in which funds for research activities flow through two streams:

- performance-based block grants and university operating grants which are provided for teaching, research, research training and general operating purposes, such as capital works and community service - these can be expended at each university’s discretion; and
- targeted research grants provided specifically for research, which are allocated competitively on the basis of qualitative, internationally peer review-judged criteria, such as those used by the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC) and for funding Cooperative Research Centers (CRCs).

---

<sup>50</sup> DETYA (1999), *Knowledge and Innovation: A policy statement on research and research training*, Department of Education, Training and Youth Affairs, Canberra.

<sup>51</sup> DETYA (1999), *New Knowledge, New Opportunities: A Discussion Paper on Higher Education Research and Research Training*, Department of Education, Training and Youth Affairs, Canberra.

In 2003, the Commonwealth Government is allocating to universities A\$4,952 million (including HECS) for general operating purposes. The general operating grant covers teaching, research and community service. The allocation of resources to each major activity is the responsibility of institutions, which do not record expenditure by activity. Furthermore, there is overlap in some activities since most academic staff are responsible for conducting research as well as teaching and do not divide their time precisely between their activities. So it is impossible to say accurately how much of the general operating grant is allocated to research at any one university or for the sector overall. Kleeman<sup>52</sup> estimates that 16% of institutions' general operating grants is allocated to research.

As indicated in the table below, the balance between competitive and block funding in the dual funding system is weighted in favour of block funding. About one quarter of higher education research funding is allocated on the basis of competitive, internationally peer review-judged criteria, such as those used by the ARC and the NHMRC. A further 15 per cent is allocated competitively through a range of other programs such as CRCs, but not necessarily on the basis of international peer review.

**Table 3**  
**Commonwealth Funding of University Research 2001-02**

\$2.6 billion					
Block Research Funding			Competitive Research Funding		
	\$ m	%		\$ m	%
<b>Formula based block funding</b>	1,005	39	<b>ARC Competitive funding</b>	270	10
<ul style="list-style-type: none"> <li>• Institutional Grants Scheme (\$271m)</li> <li>• Research Infrastructure Block Grants (\$114m)</li> <li>• Research Training Scheme (\$516m)</li> <li>• APAs/IPRs (\$102m)</li> <li>• Regional Protection Fund (\$2m)</li> </ul>			<b>NHMRC competitive funding</b>	371	14
<b>University Operating Grants</b>	570	22	<b>Other competitive programs</b> eg CRCs and Systemic Research Infrastructure Grants	379	15
<b>Total</b>	<b>1,575</b>	<b>61</b>	<b>Total</b>	<b>1,020</b>	<b>39</b>

Sources: Compiled from Federal Government Portfolio Budgets and Priorities 2001-02, DISR 2001 (<http://www.dest.gov.au/science/archive.htm>) and Higher Education, Report for the 2002-2004 Triennium, DEST, 2002. A description of the block-funded schemes administered by DEST is provided in Attachment B.

Source: DEST (2002) *Higher education report for the 2003-2005 triennium*, chapter 4.

#### *Block research funding*

The significant majority of funds – 61%, or \$1.575 billion – is allocated in the form of block funding, and almost one third of this (the component of the operating grant spent on research as identified by the universities) is not subject directly to any qualitative or quantitative

<sup>52</sup> Kleeman, J. (2002), *Steerage of research in universities by national policy instruments*. Paper presented at IMHE General Conference, 16-18 September 2002.

research-related performance criteria. It is driven largely by the enrolment patterns of undergraduate students.

The Commonwealth has 2 schemes for allocating formula-based block funding for research: the *Institutional Grants Scheme* (IGS) and the *Research Infrastructure Block Grants Scheme* (RIBG). The IGS supports institutions' research and research training activities. The Scheme commenced in 2002 and replaced the Research Quantum and the Small Research Grants Scheme. IGS funding is distributed across universities by a formula comprising research income (60 per cent) and publications (10 per cent, using the two most recent years' data), and higher degree research student places (30 per cent, using full-time equivalents and the previous year's data).

The RIBG Scheme aims to support high quality research by:

- meeting project-related infrastructure costs associated with Australian competitive grants;
- ensuring that areas of recognised research potential have access to the support necessary for their development;
- enhancing support for areas of existing research strength; and
- remedying deficiencies in research infrastructure.

The scheme allocates grants to publicly funded universities on the basis of an index which measures institutional success in obtaining competitively awarded research funding. RIBG allocations, which can be paid to universities only, are derived from data collected for the two most recent calendar years.

The data source for 2004 RIBG allocations is income over the 2001 and 2002 calendar years as reported in the *Higher Education Research Data Collection* (see next section).

The *Research Training Scheme* (RTS) allocates higher degree research student places to institutions. The Scheme aims to improve the quality of the research training environment, reduce attrition rates and improve the completion times of students. As research students complete or discontinue their studies, their vacated RTS places become available for reallocation through a performance-based formula. This formula distributes funding across universities based on successful research student completions (50 per cent), research income (40 per cent) and the number of research publications (10 per cent).

From 2003-2004 institutional gains in the IGS from the previous year will be capped at 5 per cent. Funds exceeding the cap will be distributed to institutions incurring the greatest proportional losses under the new arrangements. Funding will also be provided to regional institutions from the Regional Protection Fund, which was announced in 1999 to reduce the impact of any decrease in funding under the new system.

The main objectives of the *Australian Postgraduate Awards* (APA) programme are to:

- support postgraduate research training in the higher education sector; and
- provide financial support to postgraduate students of exceptional research promise who undertake their higher degree by research at an eligible Australian higher education institution.

Under the APA Scheme the allocation of awards to participating institutions is based on a formula that is reflective of their overall research performance and consistent with the funding obtained under the Research Training Scheme (RTS) for new research students.

#### *Competitive research funding*

There are numerous research project grant schemes in various Commonwealth and State portfolios. The ARC and the NHMRC are the most important agencies for the funding of

projects and fellowships. The Commonwealth allocates some \$145 annually for cooperative research centres in natural science and engineering. There are currently 62 CRCs in manufacturing technology (11 centres), information and communications technology (7), mining and energy (8), agriculture and rural based manufacturing (12), environment (15) and medical science and technology (9).

## 6.2 Measuring Research Output

Reporting of research output is part of the accountability framework in Australian higher education. As was mentioned above, research performance measurement also feeds back in the funding of research. We will discuss two important channels through which research output is reported to the Commonwealth Department of Education, Science and Training (DEST). These are:

- Research and Research Training Management Reports
- Higher Education Research Data Collection

### *Research and Research Training Management Reports*

Universities have been required to provide Research and Research Training Management Reports (RRTMRs) as part of the Knowledge and Innovation reforms announced in December 1999. These reports form a major part of the accountability requirements for universities. Amendments to HEFA make approval of grants under the block research funding schemes contingent on universities having an approved RRTMR in place.

RRTMRs were introduced in 2000. They are prepared annually and serve a number of important objectives, including:

- increasing transparency in the setting and reporting of institutional goals for research and research training;
- serving as an important public accountability mechanism to assist in assuring the quality of research and research training performed in Australia's higher education sector;
- facilitating institutional plan-setting and goal-setting by institutions, and providing a snapshot of the way each institution directs its research efforts, its areas of strength and how it performs in those areas, current and future directions for research and research training, approaches to managing their research and research training activities, notable recent achievements and past performance;
- providing an overview of each institution's distinctive contribution to the national innovation system within a uniform reporting framework; and
- informing prospective students, collaborative research partners and industry, in a commensurate way as to how each institution has chosen to direct its research and research training activities.

Institutions reported on their performance for all research in terms of three broad 'research clusters', namely:

- science and technology;
- health and medical research; and
- arts, humanities and social sciences.

RRTMR reports comprise two sections:

- Part A - in which institutions describe their objectives for research and research training, their future directions, practices and policies for managing research and research training, processes used to ensure quality research training experience, collaboration and partnerships, and arrangements to manage intellectual property



- issues, the commercialisation of research outcomes and contractual arrangements;  
and
- Part B - in which institutions report on their research and research training performance in a standardised format, which will enable performance trends over time and between institutions to be detected. Part B is intended to demonstrate the extent to which institutions have implemented the spirit and substance of the reforms announced in the 1999 *Knowledge and Innovation (K&I)* policy, focussing in particular on institutions' identified research strengths.

During 2003, 42 eligible institutions submitted a report. The Minister approved all 42. The 2003 RRTMRs are published by DEST on its website.<sup>53</sup>

The Minister, Brendan Nelson, has advised the DEST that he has decided not to require a Research and Research Training Management Report (RRTMR) in 2004 as a prerequisite for funding in 2005. However, it is likely that some form of reporting will be required in 2005 and at regular intervals thereafter pending development of and consultation with the sector over the implementation of the Research Quality Framework that was announced in *Backing Australia's Ability - Building our Future through Science and Innovation*.<sup>54</sup>

The RRTMR reports for each university have about 20 to 30 pages and provide information and data tables and some broad observations relating to:

- what universities have reported as their research strengths;
- a range of research performance measures relating to research income, publications, HDR (higher degree by research) supervision and representation of HDR students in areas of research strength;
- indicators of commitment to assuring the quality of research training; and
- the efforts universities are making to commercialise their research and how they are dealing with intellectual property.<sup>55</sup>
- the efforts universities are making with regard to collaboration and partnerships (between universities, government, industry and overseas partners).

When it comes to this list, universities adopt a variety of approaches in identifying their research strengths. Some use a round of competitive submissions and review; others undertake consultations; while others make use of a range of performance indicators, such as level of research income, peer review recognition, publications, patents, research collaborations, establishment of strategic partnerships, HDR student commencement and completions load.

Summarising, the RRTMR reports provide the following measures of research performance:<sup>56</sup>

- research income

<sup>53</sup> See: <http://www.dest.gov.au/highered/respubs/rrtm/default03.htm>

<sup>54</sup> In December 2003, the *Higher Education Support Bill 2003* was finally passed through the Australian Parliament. The legislation contains many of the major structural reforms that were proposed in the Government's "*Our Universities: Backing Australia's Future*" reform package, which contained the proposals announced by the minister for Education, Science and Training in response to a review of the Australian higher education sector carried out in 2002. The Commonwealth's consultation process concluded in late October 2002. The Commonwealth's decisions were announced in the Ministerial reform package, *Our Universities Backing Australia's Future*, released with the Federal Budget on 13 May, 2003. See:

[http://www.backingaustraliasfuture.gov.au/policy\\_paper/policy\\_paper.pdf](http://www.backingaustraliasfuture.gov.au/policy_paper/policy_paper.pdf).

<sup>55</sup> The reports show data on the stock of patents held by Australian universities and/or their controlled entities (Australian patents and overseas patents) and data on the revenue from licences issued.

<sup>56</sup> See: <http://www.dest.gov.au/highered/research/rrtmr.htm>.

- research active staff
- number of staff who generated research income
- number of staff who generated publications
- number of staff eligible to supervise higher degree by research (HDR) students
- number of staff who supervised HDR students
- number of areas of research strength<sup>57</sup>
- HDR students and HDR commencing students in areas of research strength
- HDR supervisor characteristics
- Summary of commercialisation activity in Australian universities and controlled entities

### *Higher Education Research Data Collection*

The *Higher Education Research Data Collection* comprises research income and research publications data submitted by universities each year. Data collected is used, along with data from the Higher Education Student Collection, for determining allocations to universities under performance based funding schemes (IGS, RIGB, RTS, APA). The Department of Education, Science and Training issues detailed specifications on the collection and submission of the data on research outputs. The data is collected in a database, the so-called *Higher Education Research Data Collection* (HERDC). The purpose of the HERDC is to collect performance data on Research Income and Research Publications in accordance with the requirements of the Higher Education Funding Act 1988.

The research income data measures each university's ability to attract income from a wide variety of sources. The research publications data reflect research output. As it is not practical to conduct a full count of all research output, the publications component is intended as a broad indicator only and is not a comprehensive count of all quality research output. The data is not intended to be used for internal funding allocations. The idea is that universities develop their own allocative mechanisms to take account of excellent research, particularly in areas not covered by the Commonwealth's publications measure.

For the purposes of the HERDC, research needs to be defined. It is defined as an activity that leads to publicly verifiable outcomes that are open to peer appraisal. Sticking with internationally accepted (OECD) definitions, research is defined as follows:

*Research and experimental development comprises:*

- *creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of humanity, culture and society, and the use of this stock of knowledge to devise new applications.*
- *any activity classified as research and experimental development is characterised by originality; it should have investigation as a primary objective and should have the potential to produce results that are sufficiently general for humanity's stock of knowledge (theoretical and/or practical) to be recognizably increased. Most higher education research work would qualify as research and experimental development.*

Research includes pure basic research, strategic basic research, applied research and experimental development.

In the HERDC, universities provide DEST with two tables of information:

- Table 1: Source of Total Research Income
- Table 2: Type of Research Publication

---

<sup>57</sup> The number of research strengths reported by each university varied widely; between 2 and 54.

The first table collects data on the source of *research income* received by a university and all of its controlled entities for the reference year. The second table collects publications data for the reference year. The research income and publications data are stored in the form of an electronic spreadsheet.

Research income is defined as any income received by a university and its controlled entities<sup>58</sup> and is provided specifically for research purposes. The data is reported in each of the following categories:

- Australian Competitive Grants
- Other Public Sector Research Funding
- Industry and Other Funding for Research.

The HERDC specifications make it clear what income is eligible and what income is not eligible to be counted as research income. A register of Commonwealth competitive grant schemes<sup>59</sup> lists the many corporations, programs, schemes, foundations, trusts, etc that qualify as funding bodies that award *competitive* grants. For the other two categories there are less long lists, but there are requirements that state that the income reported was provided specifically for research purposes.

For the purpose of research performance measurement (and research funding), *research publications* are to be reported to DEST. For this, the term research publication is defined and general requirements on reporting are specified. The requirements deal with issues such as multiple authors, classification of publications, year of publication, institutional affiliation, publishers<sup>60</sup>, peer review, electronic publication, etc. Because research is defined as an activity that is open to *peer appraisal*, the publication will have to be assessed by independent experts (*before* publication and by appropriately qualified experts, independent of the author). The concept of a commercial publisher is used as a surrogate for any formal peer review requirement. For journal articles, this means that (with a few exceptions) the journal is listed in one of the Institute for Scientific Information (ISI) indexes or in another acceptable directory of journals.<sup>61</sup> Acceptable means: satisfying the peer review requirements.

There are four categories of publications, each having a weight:

- Books, authored research (weighting: 5)
- Book chapters (weighting: 1)
- Articles in scholarly refereed journals (weighting: 1)
- Full written conference paper – refereed proceedings (weighting: 1)

For each of these, specific eligibility requirements apply. Items that are unlikely to meet the criteria for a publication in a specific category are mentioned as well (e.g. translations, edited books, textbooks, forewords, articles in newspapers, book reviews, editorials, papers in workshops).

For verification purposes, institutions will have to keep records of the publications reported. Research publications data may be subject to audit. Institutions will be given the opportunity to verify data provided to DEST.

---

<sup>58</sup> A *controlled entity* is an independent operation that acts as a corporate entity and in legal terms is separate from the university.

<sup>59</sup> The Australian Competitive Grants Register (ACGR).

<sup>60</sup> There is a register of commercial publishers.

<sup>61</sup> Ulrich's International Periodicals Directory ([www.ulrichsweb.com](http://www.ulrichsweb.com)) or DEST's Register of Refereed Journals are mentioned.

As an example, we provide the 2002 figures provided to DEST by the *Australian National University* (for its Faculty of Arts, and broken down by School and Centre):

Table 1. 2002 Publications summary (Australian National University - Faculty of Arts)

Faculty/ Centre	Department	DEST Categories - unweighted					Weighted TOTAL
		Books	Chapters	Journal articles	Conference papers	Total	
Arts	Arch & Anthropology	1.50	6.10	10.77	1.00	19.37	25.37
	Language Studies	2.00	20.83	10.00	7.83	40.66	48.66
	Humanities	3.00	14.45	7.50	1.00	25.95	37.95
	Social Sciences	4.00	17.14	17.00	8.00	46.14	62.14
	CAIS	0.00	5.00	3.00	0.00	8.00	8.00
	ANDC	1.00	0.00	0.00	0.00	1.00	5.00
	Total	11.50	63.52	48.27	17.83	141.12	187.12

Source: <http://www.anu.edu.au/ro/data/2002statistics.php>

The Australian Vice Chancellors' Committee (the council of university presidents) publishes the HERDC time series data (per university!) on its website:

<http://www.avcc.edu.au/documents/publications/stats/HERDCTimeSeriesData1992-2003.xls>

### 6.3 Results, experiences, problems

The results of the 1999 Knowledge and Innovation (K&I) reforms were evaluated in 2003/2004 by an External Reference Group chaired by Chris Fell.<sup>62</sup> The primary aim of the evaluation was to identify areas in which existing arrangements need to be improved, rather than to formally assess whether K&I (or its constituent elements) has achieved its objectives.

#### *Performance indicators and performance-based funding*

The evaluation provides an assessment of the arrangements by which the government distributes block research funding to Australia's universities. The evaluation was informed by a consultation process including workshops and submission of written statements by stakeholders (universities, research organizations, postgraduate student associations, industry bodies and individual researchers). The K&I principles of excellence, autonomy, linkage and collaboration, contestability and accountability were largely supported by the stakeholders. The approach adopted under K&I of making all research block funding provided to universities subject to performance formulae has had a positive impact. Universities and university bodies broadly oppose any move of current research funds away from performance-based block funding for the universities towards the funding councils.

The evaluation provided substantial comment on the need to assess universities on the *quality* of their research outputs and the desirability of using such assessments as a tool for resource allocation. There was discussion about whether Australia should adopt some variant of the United Kingdom's Research Assessment Exercise (RAE). The majority of the stakeholders oppose any quality mechanism that does not avoid the problems of the RAE. While the majority of respondents see significant problems with the RAE, the External Reference Group considers that there would be value in exploring whether it is possible to design an approach to quality assessment that avoids the RAE's drawbacks.

<sup>62</sup> See: [http://www.dest.gov.au/highered/ki\\_reforms/default.htm](http://www.dest.gov.au/highered/ki_reforms/default.htm).

There was broad support for the Institutional Grants Scheme (IGS) overall, including the equal weighting of all research income in all formulae which was seen to contribute to the linkages to the national innovation system.

There was general support for retaining a count for publications in the formulae for RTS and IGS. However, a number of respondents put forward the following criticism:

- the tendency of the publications count to promote quantity at the expense of quality.
- publications counts adds limited value as a device for allocating funds, since institutions' publication numbers are highly correlated with other elements in the performance-based block funding formulae (especially research income)
- publication counts are subject to criticism in terms of the publications it does or does not include.

Some are citing especially the importance of this measure for the social sciences, arts and humanities, which find it more difficult than other disciplines to attract research income. While a majority support the introduction of quality measures in principle, many are uncertain how this might be achieved, citing particularly concerns over equity between disciplines and unreasonable or unmanageable compliance costs. Although many have concerns with procedural aspects of the publications measure, it was generally considered crucial that the funding formulae retain some measure of research output.

Some advocate increasing the weighting of the publications count beyond the current 10 per cent. However, the representatives of Australian universities (the AVCC) argue that raising the publications weighting would tie substantial funding to each single publication, thus increasing the risk of over-publication or inaccurate claims.

However, other stakeholders point to a range of problems. In essence, these arise from the difficulties inherent in using a simple numerical measure as a proxy for the highly diverse and complex outcomes that are desired of the research system. Many stakeholders cite the findings of Linda Butler<sup>63</sup> that a rise in the number of publications has been accompanied by a significant decline in citation impact. Stakeholders appreciate that reliance on proxies can induce aberrant behaviour, both on the part of university administrators as they seek to optimise their institution's positions, and on the part of individual researchers. According to the Academy of the Social Sciences in Australia, it could discourage ground-breaking and longer-term research: If productivity is defined and counted in terms of units of output rather than quality, it is inevitable that the system will reinforce safe well-travelled research paradigms, immediate results and quick publication at the expense of bold new directions over long time horizons (Academy of the Social Sciences in Australia).

Problems of a more procedural nature identified in submissions included:

- the difficulties raised by differences in publication practices between fields
- the diversity of publication types, raising the question of how each publication type should be weighted in the overall count – some argue that the weightings for books and book chapters are too low vis-à-vis standard refereed articles
- the exclusion of certain classes of publications, including publications in the creative arts and in systematics, and patents
- the costs of collecting and auditing publications data.

Linkage and collaboration is a key goal in the K&I policy. However, there is relatively little in the performance-based block funding framework which serves as a direct driver to achieve this goal, either within formulae or other mechanisms. The evaluation did not provide much further detail on the kinds of drivers or indicators that might be incorporated into funding

---

<sup>63</sup> Butler, L. (2003), Explaining Australia's increased share of ISI publications—the effects of a funding formula based on publication counts. *Research Policy* Vol. 32, pp. 143–155.

structures or to encourage the growth of business expenditure on R&D and industry-university efforts. Evidence from stakeholders' consultations, RRTMRs, levels of commercialisation activity and trends in research income received from non-Commonwealth sources all point to substantial levels of connectivity within the university research system and between it and the remainder of the national innovation system. However, there are no simple measures of the amount of connectedness, and there is no agreed way of assessing how much connectedness is optimal. Growth in research income from sources other than National Competitive Grants provides some grounds for inferring that the connections have grown stronger since K&I was announced. Since 1999, research income earned from non-Commonwealth public sector and industry sources has grown both in absolute terms and in relative terms. Trend data on universities' levels of commercialisation activity since the introduction of the K&I reforms are not available. However, benchmark information relating to the year 2000 has been gathered through a National Survey of Research Commercialisation conducted by the ARC in conjunction with the CSIRO and the NHMRC.<sup>64</sup>

#### *The HERDC and the RRTMRs*

The evaluation also addressed the usefulness of the *Higher Education Research Data Collection*, including categories under which its data are provided. Are there more viable data alternatives for the system as a whole? Just over 10 per cent of respondents provided comment on HERDC. All but one agree that the collection is useful and should be retained. Broadly there appear to be no significant issues with HERDC.

Respondents in the evaluation study were evenly split on the value of RRTMRs, with a slight majority wishing to discontinue them. The main arguments for discontinuation were that:

- RRTMRs have added a costly administrative burden, for example:
- RRTMRs suffer from lack of clarity about objectives. There are more effective mechanisms of reporting and accountability, which receive more scrutiny
- There is little conviction about the value-adding of the RRTMRs either in terms of the university itself or in terms of prospective students, as was intended
- The RRTMRs are not necessarily a good indicator of strategy.

A high proportion of respondents urged further streamlining of RRTMR reporting requirements. Some key suggestions included:

- reducing the frequency of reporting that would ease the burden on institutions, perhaps either through a biennial or triennial reporting regime
- tightening data definitions in guidelines that would help with consistency and comparability of data
- developing new electronic means of reporting that would aid comparisons.

There was still confusion about the precise objectives of RRTMRs. Some suggested there is a difference between a more strategic approach identified with a plan, as distinct from the straightforward information gathering required of a report. While there need not be such a dichotomy, some submissions identified it as a source of possible confusion. On the positive side, many institutions (large and small) acknowledged the benefits of RRTMRs, in terms of an improved accountability process and as an aid in strategic planning.

#### *Impact on particular disciplines*

A quarter of respondents provided comment on how particular disciplines were faring under K&I. Over 10 per cent of respondents expressed the view that the humanities and social sciences are disadvantaged under the K&I framework.

---

<sup>64</sup> See: Australian Research Council, Commonwealth Scientific and Industrial Research Organisation and National Health and Medical Research Council, National Survey of Research Commercialisation (Canberra: Australian Research Council, 2002).

Other propositions that the K&I reforms influenced a changed institutional behaviour include:

- a bias toward high-cost courses (humanities is a low cost)
- disproportionate pressure on course completions for humanities students
- an incentive to align research strengths with high-cost (that is non-humanities) courses
- discouragements to enrol Masters by Research (which affects humanities disproportionately)
- an incapacity to adequately capture the value of humanities research.

Most of the other comments related to how vulnerable the humanities and social sciences would be, should publications be removed from funding formulae for the IGS and RTS.





## 7 Measuring research performance in New Zealand

### 7.1 The NZ research system and its funding

Higher education in New Zealand is regarded as part of the greater whole of ‘tertiary education’, or ‘post compulsory education and training’. The public tertiary education organisations (TEOs) include universities, colleges of education, polytechnics and *wananga*.<sup>65</sup> Apart from these so-called Crown (i.e. state) entities, there are a great number of private training establishments (PTEs) that provide post-secondary education of some sort, government training establishments, industry training organisations, and continuing education organizations. Currently there are 38 public TEOs, enrolling almost a quarter of a million students: 8 universities,<sup>66</sup> 23 polytechnics, 4 colleges of education and 3 *wananga*.

Universities are the single most important source of basic research in New Zealand. In the 1990s, research within universities was funded on the basis of the number of students. This was known as the EFTS (equivalent full-time student) ‘top-up’ funding for research. Funding incentives therefore were not explicitly aimed the goal of creating a focus on excellence, but rather they made important research areas vulnerable to volatile student demand. In addition, it was widely recognised that research in New Zealand’s tertiary education system was too often disconnected from the rest of the national innovation system and reflected too little concentration and focus by individual institutions on their areas of strength.

Universities receive public funding through a combination of core funding from the ministry of Education and funding out of a special fund that allocates funding for Centres of Excellence. In addition to this, there are sources such as the Marsden Fund and other ‘public good’ funding allocated by the *Foundation for Research, Science and Technology*, the *Health Research Council* and *Technology NZ*. Furthermore there are a small number of *Centres of Research Excellence* (CoREs) that receive funding from the CoRE fund, established in 2001 to promote excellent research carried out in collaborations between tertiary education organisations (TEOs) or TEOs in linkages with other public research organisations and industry.

To help focus the national research efforts and resources around areas of excellence and encourage high performance, the government set up a working group in 2001 to develop a performance-based research fund (PBRF). The Working Group consulted widely with experts from within and outside the tertiary education sector. A major review was undertaken of the UK, Australian, and Hong Kong research funding models, and the claim is that the resulting PBRF avoids the pitfalls and capitalises on the better aspects of these models.<sup>67</sup> The report *Investing in Excellence* published by the Working Group recommended a PBRF model that subsequently was adopted and accepted by the Cabinet. A *Tertiary Education Commission* (TEC) was established by the government in July 2002 to advise the government on the implementation of the new funding model. The model came to be known as the *Performance-Based Research Fund* (PBRF). The PBRF allocates funds on the basis the quality of the research produced in TEOs. The TEC is overseeing the implementation of the PBRF model.

---

<sup>65</sup> Wanangas are institutions that provide tertiary education and training, while also teaching Maori traditions, customs and language.

<sup>66</sup> Lincoln U, Massey U, U of Auckland, U of Waikoto, U of Canterbury, U of Otago, and Victoria U of Wellington.

<sup>67</sup> For example, a major difference is that the PBRF involved *all staff*, as a mandate by the NZ Education Amendment Act (1990) is that a criterion of teaching in a degree course is that research and teaching are closely interdependent and that most of the teaching must be conducted by people who are active in advancing knowledge.

Between 2004 and 2007, the PBRF is progressively replacing the 'top-up' funding for research. Funding allocations through the PBRF will not be fully implemented until 2007. In the meantime, the bulk of the research funding will continue to be allocated through degree "top up" funding arrangements (i.e. on the basis of student enrolments). These will be phased out gradually and replaced by funding based on the PBRF funding formula. The funding rates for the "top up" component of undergraduate degree and research postgraduate degrees will reduce to 90% of the 2003 rates in 2004, to 80% in 2005, and to 50% in 2006; and the "top ups" will be completely phased out in 2007.

## 7.2 Measures of research performance

The PBRF fund is intended to provide incentives to increase the average quality of research. The report of the Performance-Based Research Fund Working Group in December 2002 also states another purposes of the PBRF – that is to '*improve the quality of information on research output*'. The publication of the PBRF results and access to the information collected in the course of the PBRF Quality Evaluation (see below) and the subsequent analysis of it is producing standardised and transparent information on research outputs that is made available to users. Data is collected for the PBRF through the Ministry of Education's (MoE) existing data collection mechanisms, as well as by the TEC in the PBRF exercise. The data submitted by the institutions is known as the *Evidence Portfolio* (EP), discussed below.

The main aim of the PBRF is to ensure that research continues to support degree and postgraduate teaching, increase the average quality of research, improve the quality of public information on research output, and underpin the existing research strength in the tertiary education sector. The emphasis is on "excellence" in terms of *quality* of research and *not* on quantity. The fundamental question is whether researchers are undertaking world-class quality research, gaining esteem from this excellence, and similarly contributing to the research environment to further enhance research excellence.

There are three parts to the PBRF model:

- A periodic *Quality Evaluation*: this measure allocates 60% of performance-based funding;
- *Research Degree Completions*: 25% of performance based funding; and
- *External Research Income*: 15% of performance based funding.

### *Quality Evaluation*

The component that has attracted most comment, and certainly involvement is the first: the *Quality Evaluation* of academic research. This component rewards and encourages the quality of researchers. All 'eligible staff' is assessed *individually*, by one of 12 peer review panels, on the basis of an *Evidence Portfolio* containing information about their research. This EP has three parts: *Research Outputs* (70%), *Peer Esteem* (15%), and *Contribution to Research Environment* (15%). The three components are elaborated further below.

In the Research Outputs section, academics nominate up to 4 "nominated research outputs" (NROs) that best illustrate the quality of their research and up to 50 other research outputs, completed in the last 6 years. The academic has to indicate whether the NRO was quality assured. 'Quality assured' means that the "output must have been subject to formal independent scrutiny by those with the necessary expertise and/or skills to assess its quality (including, where relevant, its rigour, logic, clarity, originality, intellectual significance, impact, applications, artistic merit, etc.)". Such quality assessment processes could include: blind peer review or refereeing processes and, unlike the peer esteem component (see below), this review occurred *before* its dissemination or publication.

The Peer Esteem section is concerned with the recognition of the staff member's research by her or his peers. Indicators of peer esteem include: research-related fellowships, prizes, awards, invitations to share research knowledge at academic and end-user conferences and events, the ability to attract graduate students or to sponsor students into higher level research qualifications, positions or opportunities because of research reputation, and research-related citations and favourable reviews.

The Contribution to Research Environment is concerned with the contribution of a researcher to the development of research students, to new and emerging researchers and to a vital high-quality research environment. Indicators relate to: research and disciplinary leadership – including membership of research teams, and contributions to disciplinary development and debate and public understanding of the discipline; contribution through students and emerging researchers (such as supporting students to achieve post-graduate qualifications and to develop as researchers); and/or contribution to organisational vitality – supporting the development of research both within and across institutions (e.g. hosting visiting researchers).

#### *Research degrees*

The *Postgraduate Research-based Degree Completions* (e.g. Doctoral, Masters) measure is a measure of the number of research-based, postgraduate degrees completed within a TEO. The Completions measure is weighted in the funding formula for the following factors:

- The cost of the subject area
- Maori and Pacific student completions (an equity weighting), and
- The volume of research in the programme (i.e. the programmes' research component).

Research Degree Completions information is collected annually using the Ministry of Education's Single Data Return (SDR).

#### *External research income*

*External Research Income* (ERI) is the total of a TEO's research income that is received by the TEO and/or any 100% owned subsidiary of the TEO. Only research funding from outside the tertiary sector (and contestable funding from within the tertiary sector) can be included as ERI. ERI includes income for research purposes from trusts, like *the Community Trust*, *Wellcome Trust* or *Lion Foundation*. All eligible forms of ERI are treated equally in the funding formula.

### **7.3 The peer review process**

The peer review of the evidence portfolios (EPs) observes guidelines for handling potential conflicts of interest, moderation across panels, and includes audits of the EPs, sharing "benchmark" EPs across panels, detailed statistical analyses within and between panels, and stipulations that at least 25% of NROs should be read by the panel members. The panel members use the following quality ratings (on a scale of 0 to 7) for judging the three parts of the EP:

- 0-1 = Not meeting the requirements for active researcher (R),
- 2-3 = Regular application of existing research methodologies with acknowledgement by peers of sound research basis (C);
- 4-5 = Original or innovative research that is recognised within New Zealand or elsewhere and is esteemed by the academic community beyond the researcher's own institution (B);
- 6-7 = Highly original or innovative research that ranks with the best of its kind in the world and is esteemed by the international academic community (A).

The scores for the three parts of the EP are summed (using the weightings cited above) and an overall Quality Indicator of A, B, C, or R is assigned. This Quality rating is then multiplied by the "content" weightings (e.g., Arts, Business, Teacher training is weighted 1, Science,

Nursing, Music is weighted 2, and Engineering, Agriculture, Medicine, Audiology is weighted 2.5). The nature of appointment (fractional to full-time) is included in the final formulae, and the scores are then aggregated by Department/School/Discipline, and then by TEO. The final weighted sum is then added to the other two components (degree completions, and external research income) and this is used to determine the Research funding allocation to the TEO.

#### *Outcomes of the 2003 peer reviews*

The results of the first round of PBRF were announced in early 2004. Of the many TEOs in NZ, 45 are eligible for PBRF funding, although not all took part in the 2003 assessment. Eight universities, two polytechnics, four colleges of education, one wananga and seven private training establishments participated. This is half of the eligible institutions. In 2006 the next Quality Evaluation is planned to take place.

For the 2003 evaluation, of the 8,013 PBRF-eligible staff in the participating TEOs, 5,771 had their Evidence Portfolios assessed by a peer review panel. There were 12 such panels covering 41 designated subject areas. A Moderation Panel comprising the 12 panel chairs and an independent chair oversaw the work of these expert panels. Altogether, there were 165 panel chairs and members, 33 from overseas. Research degree completions were notified by 13 TEOs. Roughly two-thirds of the completions were for masters courses, with the remainder being doctorates. The external research income generated by the 15 TEOs that provided data totalled about NZ\$ 195 million for the 2002 year. All but about NZ\$ 1 million was generated by the eight universities.

Because of the complexity of the assessment system, simple characterisations of “A”, “B”, “C” and “R” are difficult to make. In very broad terms:

- “A” signifies research of a world-class standard
- “B” signifies very good quality research
- “C” signifies good quality research
- “R” signifies that the Evidence Portfolio did not meet the requirements for a “C”.

It should be noted that not all staff that produced research outputs that were deemed to be of a world-class standard secured an “A”. In many cases, for instance, high-calibre researchers were assigned a lower Quality Category because they failed to demonstrate either the necessary level of peer esteem or contribution to the research environment. It is important to recognise that “R” does not necessarily signify “research inactive” or indicate poor-quality research. The “R” category includes many new and emerging researchers of high potential. Being in the early stages of their research career, most had not yet been able to acquire a substantial measure of peer esteem or make a major contribution to the research environment.

The results of the 2003 Quality Evaluation, and especially the quality score data, reflect the nature of the assessment methodology that has been employed and the particular weightings applied to the four Quality Categories – i.e. “A” (10), “B” (6), “C” (2), and “R” (0). The FTE-weighted quality score for the 22 participating TEOs is 2.6 (out of a potential score of 10).

Under the approach adopted, the maximum quality score that can be achieved by a TEO (subject area or nominated academic unit) is 10. In order to obtain such a score, however, all the PBRF-eligible staff in the relevant TEO would have to receive an “A” Quality Category. With the exception of very small academic units, such an outcome is extremely unlikely. No sizable academic unit, let alone a large TEO, could reasonably be expected to secure a quality score even close to 10. Much the same applies to quality scores at the subject-area level. Likewise, there is no suggestion that a quality score of less than 5 constitutes a “fail”. These considerations are important to bear in mind when assessing the results of the Quality Evaluation.

Furthermore, the quality scores provide only one way of depicting the results and do not furnish a complete picture. For instance, the subject area of Education (i.e. didactics) achieved a relatively low quality score (1.02 FTE-weighted), yet it contains no less than 24.4 A-rated staff and 70.3 B-rated staff (FTE-weighted). The low quality score reflects the very large number of staff whose Evidence Portfolios were assigned an “R”.

#### 7.4 Further information on the PBRF process

For the PBRF process, research is defined in pretty much the same way as in the case of Australia and the UK (see separate chapters in this report). In contrast to the UK though, *all* academics are submitted to the Quality Evaluation – not just the researchers that are meeting specific (high) standards. All academic staff members that are expected to contribute to the learning environment and to make a significant contribution to research activity and/or degree teaching are expected to participate in the Quality Evaluation. The TEO should nominate to Category R any staff who are eligible to participate but who are research inactive or are research active but do not meet the requirements for Categories A to C. The TEO will need to submit to the TEC only those EPs of staff in Categories A to C. EPs do not need to be submitted for staff in Category R.

When it comes to collecting the research performance data it is good to mention that the PBRF is primarily concerned with *quality*, not with quantity. The EP should provide an overview of a staff member’s outputs and contributions during the assessment period. Where a staff member has more material than can be included in the EP, he/she should select their best research outputs and their most significant examples of peer esteem and contribution to research environment from the assessment period. Staff members are only allowed to select their best research outputs produced during the assessment period for inclusion as their up to four nominated research outputs. The *research outputs* to be submitted include:

- Published academic work (such as books, journal articles, conference proceedings, and Masters or PhD theses)
- Work presented in non-print media (such as films, videos and recordings), and
- Other types of outputs (such as intellectual property, materials, products, performances and exhibitions).

Researchers are limited to providing 30 examples of *peer esteem* during the assessment period on their EP. Examples are prizes and membership of learned societies. Likewise, staff members are limited to providing 30 examples of *contribution to the research environment* during the assessment period. As with other items of research output, the items have to be classified in specific categories listed in the guidelines to the PBRF Quality Evaluation. All in all the guidelines and procedures for the Quality Evaluation cover more than 300 pages.<sup>68</sup>

In sum, the following information was publicly reported for each participating TEO in the 2003 Quality Evaluation round:

- The average quality score for all eligible staff (weighted on a FTE basis). The average quality scores are calculated on the basis of the (A, B, C) quality scores of individual staff members for the three components of EPs.
- The proportion of all eligible staff (weighted on a FTE basis) that received the scores of A, B, C or R.
- The total number of eligible staff (weighted on a FTE basis) at the census date.
- The proportion of all staff at the census date that are involved in research and/or degree-level teaching.

---

<sup>68</sup> See: PBRF Integrated Guidelines (PBRF A Guide for 2003), at [http://www.tec.govt.nz/downloads/a2z\\_publications/pbrffinal-july03.pdf](http://www.tec.govt.nz/downloads/a2z_publications/pbrffinal-july03.pdf).

- The proportion of all eligible staff (weighted on a FTE basis) submitted to the TEC for a quality rating.
- The total number of student places (EFTS) at the undergraduate, taught postgraduate, and wholly research postgraduate levels for 2002.
- The total number of research postgraduate completions in 2002 (including equity weightings).
- The external research income (i.e. that which is eligible for the purposes of the PBRF) received in 2002.
- Basic demographic data at an aggregated level.

This data is also published for each TEO, panel, subject area, and academic unit level.

## **7.5 Evaluation of the PBRF and the Quality Evaluation**

An evaluation of the first round of the PBRF and the Quality Evaluation that is part of it took place over 2003/2004.<sup>69</sup> We will present the main conclusions.

No TEO or other stakeholders advocated any fundamental re-design of the PBRF or the Quality Evaluation process. On the contrary, there was a broad expectation that the 2006 Quality Evaluation would remain more or less the same and that the on-going operation of the PBRF would proceed with continuous improvements. The implementation of the PBRF and the conduct of the 2003 Quality Evaluation were possible because of a high level of trust and cooperation between Tertiary Education Organisations (TEOs), their staff and the Tertiary Education Commission (TEC).

Overall, the evaluation suggests that attention should focus on modifying the existing processes in time for the less hurried conduct of the 2006 Quality Evaluation. The evaluation does not suggest that attention be given to a fundamental redesign of the PBRF or the Quality Evaluation for three reasons:

1. there was no evidence of major design failure reported or observed in the evaluation;
2. while participants identified problems with design elements and aspects of the implementation, these were not seen as immediate and fatal threats but rather as remedial; and
3. it is too soon to observe or evaluate the impact of the PBRF upon TEOs and individuals.

The evaluation showed that sector confidence in the PBRF needs to be fostered. Giving priority to demonstrating fairness and transparency, rather than to the information needs of the public or policy makers, is more likely to sustain the PBRF during its transition years (2003-2007).

It was proposed that the main features of the 2003 Quality Evaluation should be retained in the next Quality Evaluation (be that 2006, or later). Thus a mixed model of peer review and quantitative performance indicators, and based on the individual staff member as the unit of assessment in the Quality Evaluation.

On the issue of the submission of researchers in the peer review process we mention the fact that TEOs are allowed to nominate the academic units into which Quality Categories were to be grouped for reporting purposes. This had two effects:

---

<sup>69</sup> See WEB Research (2004), Evaluation of the implementation of the PBRF and the conduct of the 2003 Quality Evaluation. Available on-line at:  
[http://www.tec.govt.nz/downloads/a2z\\_publications/pbrfwebreport\\_final29jul.pdf](http://www.tec.govt.nz/downloads/a2z_publications/pbrfwebreport_final29jul.pdf).

- some TEOs chose to define relatively small academic units (despite a prior warning that this would make it easier to infer individual Quality Categories); and
- the range of decisions made by TEOs made it impossible to reliably compare the relative performance of nominated academic units within similar disciplinary areas. The use of subject-area quality scores in the reporting of results helped to overcome this problem; but this caused other problems such as apparent distortions arising from the groupings chosen.

Participants in the evaluation understood that it is not possible to devise a completely objective, or error-free, qualitative evaluation of research. They also understood the considerable difficulty in putting such a scheme into operation. Their expectation was that the broadest consensus will be developed around what the standards are to be; how they are understood within the TEC and TEOs and how they are to be interpreted by the peer review panels. The focus, between Quality Evaluations, should be on continuous improvement to the fairness and consistency of the practice of the Quality Evaluations.

#### *Suggestions for improvement*

There were some detailed *recommendations* made with respect to the design of the PBRF. These issues would need to be resolved to guarantee the success of the PBRF as a funding mechanism in the future.

One of the recommendations was that the criteria for *determining staff participation* in the PBRF should be reviewed and revised. For TEOs in the 2003 Quality Evaluation, identifying PBRF-eligible staff was both a compliance task and a strategic task, as they were required to identify PBRF-eligible staff and obtain Evidence Portfolios (EPs) from those staff. Determining the total number of PBRF-eligible staff is critical for a TEO because all PBRF-eligible staff are included in the denominator for calculating the quality score. However, under the current policy framework, a TEO faces conflicting pressures in applying the eligibility guidelines. On the one hand, it has an incentive to minimise the number of eligible staff (especially staff likely to secure an ‘R’) in the interests of reducing the size of its denominator and thus increasing its likely quality score. On the other hand, if it excludes staff members who have a chance of securing at least a ‘C’, it runs the risk of reducing the funding to which it is entitled. However, within the limits of the evaluation, no examples of ‘creative’ interpretations of the eligibility criteria were found.

The PBRF was intended to integrate a funding allocation mechanism with a measure of research quality so as to *stimulate excellence in research* and lively research environments across the tertiary sector. Although the PBRF provided an opportunity for all TEOs to have the excellence of their research assessed, not all TEOs participated in the 2003 Quality Evaluation. If the overall intent of the PBRF in its 2003 Quality Evaluation was ‘to reward and encourage excellence [in research]’ in all TEOs, then the responses of most polytechnics, some wananga, and some private training enterprises were not congruent with that objective as they elected not to participate. The main participants in the Quality Evaluation are likely to be the universities. As universities currently contain most of the PBRF-eligible staff within the tertiary education sector, this will mean that the policy intent will largely be met.

Another recommendation concerned the *administrative and compliance costs* of the PBRF and the Quality Evaluation. Remember that the costs of a TEO participating in the PBRF are to be met by the institution itself. The intention was that for the Ministry of Education (MOE), Tertiary Education Council (TEC), and the higher education institutions (TEOs) these costs be no more than 2% of the total PBRF funding allocated for the period 2007 to 2012. However, estimates of costs indicate that the percentage of administrative and compliance costs as a proportion of the total funds allocated is likely to fall from a range between 12% and 17% for the period 2004-2006 to a range between 1.45% and 1.93% for the period 2007-2012. A limit of 2% was seen to be a realistic goal.

The structure and composition of the *review panels* was problematic. The fundamental decision to adopt a small number of multidisciplinary panels rather than a larger number of disciplinary panels (like in the UK) was considered to be inevitable for a small country. This has produced some apparent inconsistencies in assessments of EPs in subject areas where there was no person from that subject area on the panel, or where one disciplinary area was assessed by more than one panel.

Yet another issue was the *definition of research*. The PBRF Working Group adopted a ‘new definition of research, which is specific, yet wide ranging and enables excellence to be recognized wherever it occurs’. This definition arguably formalises what exists as tacit practice in universities. It is largely based on the definitions used in the UK RAE, the New Zealand Qualifications Authority (NZQA), and the OECD. However, it is not identical to the NZQA definition; and for non-university TEOs this was a critical issue. In addition, for some subjects there remains a question as to whether the definition itself, or its qualitative interpretation by peer review panels, did in fact adequately recognise excellence wherever it occurred. The issues relating to the interpretation of the definition of research are complex, and not susceptible to simple solutions.<sup>70</sup> They include:

- the interpretation of its meaning by the peer review panels, and especially its application to a continuum of activities where many fall ambiguously on the boundary of what is included and what is not;
- the problem of applying the definition in ‘non-traditional’ research areas, such as the creative and performing arts; and
- the question of whether modifiers (such as the following) might be better included in the definition itself:

The quality evaluation process will give full recognition to work of direct relevance to the needs of industry and commerce, and all research, whether applied or basic/strategic, will be given equal weight (Ministry of Education and Transition Tertiary Education Commission, 2002, p.18).

Although the *Guidelines* for panel members specifically stated that ‘research outputs (ROs) that deal with topics or themes of primarily *local, regional or national focus* can be of world standard’, in practice peer review panel members often found it difficult to judge whether this was so for any particular RO unless it was one that they personally sighted and had personal subject expertise about. Proxy evidence used by panels did not always allow the panel to make an accurate judgement. The most commonly used proxy indicator of quality was the international standing of journals, although publication in a lesser journal, or in a New Zealand journal, could not necessarily be taken to indicate lack of quality.

There is a broad consensus that new and *emerging researchers* were unfairly treated in the 2003 Quality Evaluation. This was largely due to an unanticipated effect of the scoring system, whereby it was difficult for young researchers to produce the evidence of Contribution to Research Environment (CRE) and Peer Esteem (PE).

About the preparation of Evidence Portfolios it was recommended that an improved *IT platform* was developed to facilitate the creation, collection and processing of EPs, to minimise errors in data presentation, and reduce the costs of providing and maintaining PBRF data. Now that TEOs have seen what EPs are used for, what they should contain, and how they are to be submitted, they are able to offer suggested improvements in the design and

---

<sup>70</sup> The definition of research and its interpretation/application for assessing the quality of research outputs in the next Quality Evaluation is the subject of a recent consultation round. See: TEC (2004), Performance-based research fund. 2006 Quality Evaluation. Definition of Research Consultation Paper. (available at: [http://www.tec.govt.nz/downloads/a2z\\_publications/pbrf2006-definitionresearch-feedback.doc](http://www.tec.govt.nz/downloads/a2z_publications/pbrf2006-definitionresearch-feedback.doc) ).



management of the EP preparation, assessment and submission process. TEOs expect to make substantial savings in their indirect costs from improving their own EP process and from integrating that process with their research recording, reporting and management systems.

All in all, the evaluation of the New Zealand PBRF model produces evidence of a relatively successful research performance measurement exercise. Clearly, the exercise needs some refinement, but the overall rationale and design seems to be largely accepted by the academic community. Currently, a number of consultations are underway to see where the design of the Quality Evaluation and other parts of the PBRF need adjustment.<sup>71</sup> This means that the process is a clear example of 'learning by doing'.

---

<sup>71</sup> For instance, see: TEC (2004), Performance-based research fund. 2006 Quality Evaluation. Assessment Framework Consultation Paper. (available at: [http://www.tec.govt.nz/downloads/a2z\\_publications/pbrf2006-assessment-feedback\\_doc.html](http://www.tec.govt.nz/downloads/a2z_publications/pbrf2006-assessment-feedback_doc.html)). Other aspects of the PBRF are also addressed. See: [http://www.tec.govt.nz/funding/research/pbrf/consultation\\_documents.htm](http://www.tec.govt.nz/funding/research/pbrf/consultation_documents.htm).



## 8 Synthesis and Analysis

As the different country studies show, research performance measurement (RPM) is a diverse and, to this point, largely subjective undertaking. This section pulls together the information presented in the case studies. Our objective here is twofold: to synthesize what is occurring in these different systems and get a better idea about how it may be practically applied in other systems, like that in the Netherlands.

### 8.1 General observations about RPM

In the past several years, many governments have been exploring how to best encourage high quality research and improve the quality of education received by students in higher education. Implicit in the desire to improve research and educational outcomes is a need to assess the current performance levels of different types of institutions and compare these outcomes over time. The assessment systems to be used for this include mechanisms that provide useful information regarding the research performance of universities.

Measuring research output and research performance is inherently problematic. Quantitative indicators are all laced with qualitative aspects; so much so that higher education research that uses research indicators or the data behind them tend to regard measurement concerns as the quintessential methodological shortcoming (cites, dates). What the country studies attest to more than anything else is that research performance measurement does not collapse neatly into a single composite number or value. Indeed, specifically for this reason places like the United Kingdom, Germany and Australia are increasingly trying to evaluate research performance by considering a broad mix of input, output and process indicators.

However, despite the growing use of multiple research performance measures, there is as yet no consensus over how to weight the individual indicators. Such weightings can be found, the British Funding Council's (HEFCE) application of the Research Assessment Exercise (RAE) results is probably the best example; what cannot is an empirical or theoretical basis in their support. Should input measures be given greater importance or should output measures? Should faculty inputs be given greater weight than doctoral students or research income? Should basic science publications outweigh applied ones and should either outweigh the numbers of doctoral degrees granted? Questions such as these have yet to be answered and in all likelihood will remain that way for some time to come. In the absence of more detailed knowledge about how precisely universities transform inputs into outputs (i.e. the long sought after higher education production function), it is simply not possible to know where priority must be given.

And what is research? All serious efforts at developing research performance indicators attempt to first define "research," though what one finds is that the expected precision from defining the object is seriously curtailed by characterizing it with highly generalized descriptors. The obvious implication is that such vagueness makes its way into RPMs. Economists generally describe productivity (or performance) in ratio form: outputs per inputs. Though such a formula is conceptually simple, practical implementation is remarkably difficult (cite, date). Analyses of multi-product organizations must account for things like whether inputs are used in the production of several outputs or the priorities given to producing one output over another. Given the rich body of economic literature relative to productivity measurement, it is surprising if not disappointing to find research performance measures being developed based on ad hoc, piecemeal inclusion of intuitively appealing indicators rather than serious analyses of the institution's internal production dynamics. While several attempts are made to develop research indicators "per scientist or academic" a

good number are still strictly output- or input-based. In this regard, practically all of the performance indicators identified in this study suffer from two basic (yet serious) shortcomings: 1) they only capture half of the productivity equation, and 2) they seriously neglect the dynamics behind using multiple inputs in the production of multiple outputs.

Finally, producing research performance indicators can be a time-consuming task. Coordinating the collection of relevant information from multiple sources across what may include hundreds of institutions and ensuring that data is uniform in its definitions requires considerable time and effort. It also depends on the type of information being collected. Quality indicators, for example, require considerably more effort to obtain the necessary information. This is particularly evident in the countries we looked at. In more cases than not performance measures are produced annually (e.g., the SET in the UK, NIFU-STEP in Norway, Flanders and Australia). The RCN's evaluation of research and research institutions in Norway are done on a bi-annual basis, while in Germany and New Zealand the exercise takes place every three years. Finally the RAE in the UK runs on a six year cycle.

## 8.2 Rationale for RPM

This section briefly explores three broad potential rationales for research performance measurement (RPM). Continuing the discussion we started in the first chapter of this report, we already mentioned some of the potential users of RPM. These are: policy-makers, funding bodies, institutional managers, academics, (research) students, business and industry and the general public. In other words, the outcomes of RPM can play a role in:

1. enhancing accountability frameworks;
2. informing policy-making (including decision-making on research funding allocations);
3. informing students, academics and potential clients about the institution's capabilities, thus providing a tool for institutional marketing.

Naturally, these potential rationales for RPM are interrelated and each of them operates on the level of the higher education institution and the national (or even international) level. We will now briefly discuss each of the potential rationales for conducting RPM.

### *Enhancing accountability*

In the six countries covered in this report – and indeed across the OECD in general – there has been a general trend within government towards greater emphasis on evidence-based policy development and accountability for programme outcomes. Accountability is maintained through the use of processes whereby publicly funded organisations, and the individuals within them, are responsible for their decisions and actions and submit themselves to external assessment. The higher education sector in most OECD countries has also felt the effects of this movement towards greater 'accountability' for its use of public funds. Indeed, in the Netherlands, one of the stated aims in the Science Budget is to improve the quality of public information about research outputs, making the research performance of universities more transparent.<sup>72</sup>

### *Informing policy-making*

RPM may provide an important tool for underpinning strategy and policy-making at national and institutional levels. Governments and institutional managers have to make decisions about research activities to fund, the particular disciplines, and the recipient groups. RPM may help them prioritise the use of resources. For example, RPM may enable university leaders to ascertain their competitive position relative to other universities, and in that way establish direction and strategies for the institution. By examining an institution's research

---

<sup>72</sup> *Wetenschapsbudget 2004. Focus op Excellentie en Meer Waarde*, Ministerie van OCW. p. 27.

performance in different fields of research relative to other institutions, policy-makers and budget holders on the national or the university level may be better equipped to decide which areas are most effectively and efficiently utilising their resource allocations. It would seem that one of the most often expressed rationales for developing and relying on performance indicators for research is the need to know where one's research stands in an international perspective

#### *Institutional marketing*

Given the increasing global mobility of both researchers, research students and research funding, the relative research standing of a university may have a greater bearing on its ability to attract and retain staff and students and funding than was the case when international movement of people and resources was more limited and research budgets were less competitive. As this trend continues it becomes increasingly important for universities to market their capabilities to the students and staff that might participate in or benefit from the research capabilities of the university. RPM may serve to improve the quality of public information about research output. If performance comparisons show strong performance they may enhance the reputation of an institution or department and assist it in attracting the best quality researchers and research students — both domestic and international — to participate in its programs. Clearly this brings us to the topic of ranking, and institutional rating tables.

#### *National policies' influence on RPM*

The conclusion from our discussion on the rationale for RPM clearly is that it is critical not to forget that even informative research performance measures have some purpose.

Understanding national science policy and, more specifically its objective is key to applying one country's methods to another country's research system. For example, one of the key characteristics of the UK's Research Assessment Exercise is that the quantity produced really does not matter, what does matter is the quality. Indeed, one of the major underlying factors behind this decision was the result of making the polytechnics into universities in the early 1990s. Because these post-1992 universities historically did not have strong research programs, assessment based on the quantity of research produced inherently put them at a disadvantage vis-à-vis the existing universities. Putting the emphasis on quality and limiting the amount of publications that could be submitted created a much flatter playing field in the UK. This can also be seen in Norway. One of the major concerns there has been to raise the level of investment in R&D (as a percentage of GDP) to the OECD average. It is evident from the Norway chapter that both the performance indicators used and the policies recently enacted are designed to help achieve that particular objective.

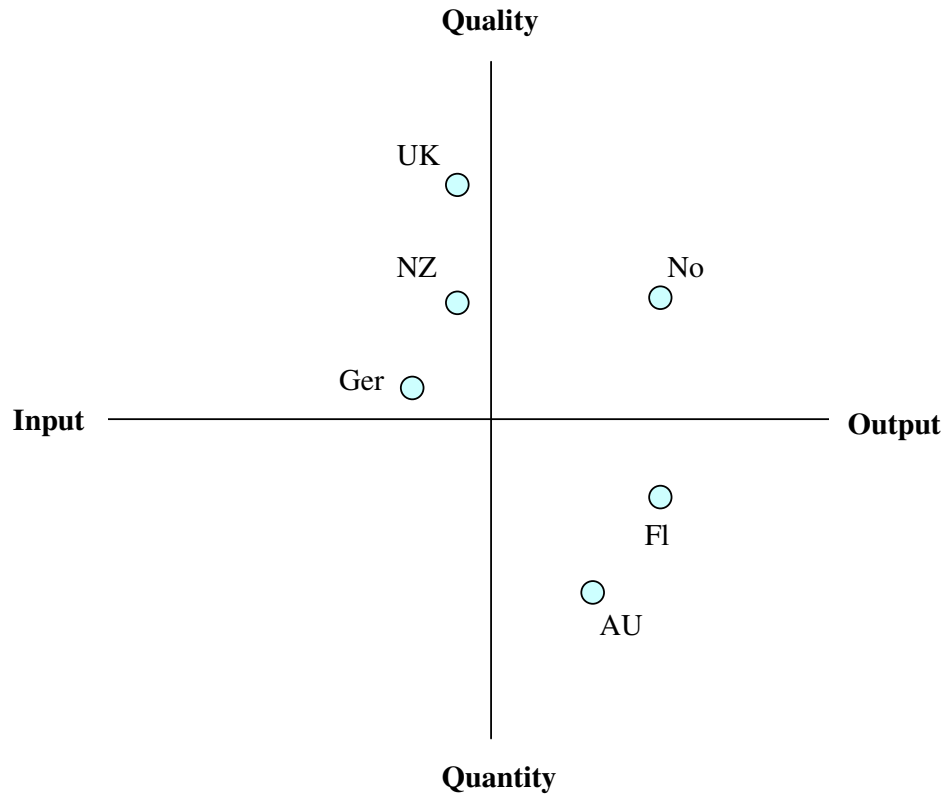
### **8.3 Two dimensions**

A useful way to characterize research performance measurement across different countries and by different types of organizations is to use a map like that in Figure 1 below. The horizontal axis is a continuum representing the extent to which RPM systems are based on inputs or outputs. The vertical axis is a continuum that indicates whether the RPM systems are quantitative or qualitative in nature. Keeping in mind what was stated in the previous sections on the rationale for RPM, one could add a 3<sup>rd</sup> axis to characterizes the purpose of RPM. At one extreme RPM tends to be used only for informative purposes; at the other it bears heavily on funding (resource allocation) decisions.

In short, what one sees from looking at Figure 1 is that types of research performance measurement are a function of who provides the resources. Where competitive (or selective) funding is distributed (like in the UK RAE-based allocations, or the New Zealand PBRF) research performance measurements tend to be both input and quality oriented. On the other hand, where funding is distributed in the form of basic subsidies (often intertwined with

education appropriations, like in Australia and Flanders) the performance measures used tend to emphasize outputs and quantitative measures of them.

**Figure 1.**



It would seem that the primary purpose of generating research performance indicators is to simply inform policymakers about the state-of-the-art. These indicators are also being used to allocate research funding but the extent to which they play any role varies considerably across the countries. At one extreme there is the RAE in the UK, which the Higher Education Funding Council for England (HEFCE) maintains that 98% of their annual research funding is based upon. At the other extreme are places like Flanders, in which it constitutes only a fractional part of annual distributions, Germany where the various *Länder* partially allocate research funding to publics, and Norway, where it is just being introduced.

#### **8.4 Data used for RPM**

A key limiting constraint in connection to RPM is the range of data that are available in relation to the research performance of universities across different disciplines. An important consideration when thinking about producing performance indicators is the extent to which the necessary data is already being collected. The ability and extent to which meaningful indicators can be readily produced is clearly a function of the quantity and quality of the data on hand. The table in the appendix shows that in all of the countries we surveyed at least some of the data is automatically being collected and in at least two (Germany and Norway), much of it is.

In practice, however, performance data collected and publicly released tends to differ considerably across (and even within) the countries. Available data can be loosely grouped into seven categories, namely:

1. bibliometric data;
2. awards to individual researchers;
3. research student data;
4. research faculty data;
5. research income from external sources;
6. research commercialisation performance data; and
7. outcomes from peer review processes.

Information available in each of these categories is briefly considered below. The next section then looks at the indicators we have come across for the six countries. Before doing so it is important to note at the outset that some measures of research quality may not be universally applicable across research fields. For instance, even the most commonly used existing bibliometric measures of quality such as citations have varied levels of acceptance across different fields of research. While broadly accepted as a valid quality indicator in the physical and natural sciences they are not as widely accepted as valid indicators of quality in the humanities and social sciences community.

#### *Bibliometric data*

Bibliometric data is widely available in relation to articles published in academic journals. It is possible to access information for a range of fields of research regarding the number of publications produced by a given researcher, department, institution or country. It is also possible to access information regarding the frequency with which such publications are cited in other articles published in academic journals.

Thomson-ISI (*Institute for Scientific Information*) is the dominant supplier of such bibliometric services. It produces three journal citation indexes – the *Science Citation Index Expanded*, the *Social Science Citation Index*, and the *Arts and Humanities Citation Index*. Each index provides a record of the number of times a publication has been cited within articles published in any of the 7000 academic journals that they monitor. Based on analysis of citation data, Thomson-ISI also compiles lists of the most highly cited research across twenty-one research fields. Recognition as a highly cited researcher means that an individual is among the 250 most cited researchers for their published articles within a specific time-period. Thomson-ISI can also provide information on the raw publication counts of researchers, institutions and countries.

#### *Awards to individual researchers*

Information regarding the results of prestigious academic awards is publicly available and it is generally possible to identify the institutional connections of award winners. In addition to academic awards, information is publicly available regarding winners of prestigious scholarships/fellowships, such as *Fulbright scholarships*. Lists of memberships of learned societies, for example the Royal Society, or the Academy of Arts and Sciences (KNAW, in the Netherlands), and the institutional connections of members are also generally publicly available.

#### *Research student data*

In a number of countries, such as Australia, New Zealand and the United Kingdom, extensive information is published by government agencies and funding bodies on research student numbers within different higher education institutions. Information available includes such measures as:

- research student enrolments by institution and field of study; and
- research student completions (diplomas awarded, e.g. PhDs) by institution and field of study.

#### *Research faculty data*

As is the case for student data, in a number of countries detailed information is collected and publicly released in relation to the number of full-time equivalent academics by institution and field of study. In addition to such general information of numbers of faculty, information is also sometimes available in relation to matters such as the proportion of faculty holding PhDs.

#### *Research income from external sources data*

As is the case with student and faculty data, a range of indicators is generally available regarding the levels and source of research income of different universities. In addition to a basic breakdown of income quantity and sources, information is also available for some institutions in relation to success rates in winning competitive research grants. Often this information tends to be aggregated at the institutional level rather than being provided on a department by department basis.

#### *Research commercialisation performance data*

Detailed information regarding performance in the commercialisation of university research and the generation of income from such commercialisation is only available on a comparable basis in Anglo-Saxon countries (Australia, Canada and the United States). These countries run similar periodic surveys on commercialisation performance. Data is collected and reported in categories such as:

- technology licences held;
- licensing revenue generated;
- patents held; and
- spin-off companies formed.

As is the case with data relating to income from external sources, commercialisation performance data is often aggregated to the institutional level and is not provided on a department by department basis.

#### *Data from peer review processes*

A final category of data that is available in relation to research performance is that contained in the results from peer review based research assessment exercises that have been conducted on a system wide basis, for instance the UK Research Assessment Exercise (RAE), the New Zealand PBRF exercise and the evaluations carried out by the Research Council of Norway (RCN). These peer review based assessment exercises provide information regarding how either individuals, departments or institutions are rated against a defined set of performance categories. These exercises provide a considerable body of publicly available information regarding performance. However, they are methodologically complex and often regarded as costly enterprises. Moreover, in no two countries have such peer review based assessment exercises followed an identical methodology.

### **8.5 RPM indicators used in the six countries**

#### *Indicators by type of research*

In most of the cases we studied, research performance measurement still focuses heavily on traditional forms of research inputs and outputs: numbers of research staff or doctoral students, research income and publications produced. To date, only marginal attention has been given to assessing the many ethereal aspects of research production/performance like collaboration and contribution to innovation. For many reasons this is not surprising.

The European countries in our study tend to try and distinguish between basic and applied research as well as focus on issues like innovation and commercialisation. A good example is



the indicator employed by the German DFG, “centrality in networks of DFG-funded coordinated programs,” which seeks to capture the extent to which collaborative research takes place and the extent to which different units take on larger/smaller roles in such projects. In Australia, reference is made to such issues in the *Research and Research Training Management Reports* (RRTMRs) that universities were required to annually complete. In New Zealand the ‘contribution to research environment’ indicator seeks to capture this idea. In Flanders, yet another set of indicators is employed. The general observation to be made about performance indicators in this area is the overall lack of consensus about what should or should not be included. Patents are almost invariably accounted for but tend to favour commercialisation activities in the biological, medical and physical sciences. Applied research income also shows up frequently, which probably gives the social sciences more leverage than patent activities, but again one runs into the problem of relying on a partial and input-based measure. While most agree that research output is not strictly limited to publications, the impact of commercialisation, innovation and strategic development remains ill defined and highly subjective across countries. If such activities are part of RPM systems at all, they are often assessed by means of peer review exercises instead of by quantitative assessment.

#### *Typical components of RPM systems*

Despite the diversity of national or system-level objectives, contexts, and priorities, it is possible to identify a number of often-used performance measures that span the case studies presented here. Per capita publication counts seem to be the most popular, with some variation. Drawing on the ISI’s Citation Indices, measures such as “share of total citations,” “publications per scientist,” “publications in international journals” and “citation impact scores” are all in use. Another commonly used set of measures involves PhD students. Indicators like “number of research assistants” or “number of PhD graduates” tend to be used, though this is not necessarily surprising especially in countries where higher education providers receive substantial government funding for research through annual block grants. In nearly all of the cases above, the measures tend to be disaggregated by academic fields.

The other very commonly employed measure was “research income,” which in some cases went further towards “contract-based or applied research income.” Where commercialisation is concerned the most commonly used indicators are “numbers of patents” and “income from patents.” Finally, we can identify at least two instances where indicators based on participation levels in the EU Framework Programs were also used (Germany and Flanders).

Where quality indicators come into play, overlapping approaches are the norm. Both the UK and New Zealand systems rely, for example, on measures such as “peer esteem” or how well scientist/researchers are regarded by their fellow colleagues. In practice this generally includes such categories as academic awards, editorships, visiting scholars, faculty members’ distinctions and the giving of policy advice. In relative terms the RAE still relies quite heavily on peer review, ironic since the prior system of allocating competitive funding through subject-based committees was considered to be not very transparent. On the other hand, New Zealand’s PBRF takes a very systematic approach to defining quality; it simply requires that any submitted publications (or other output) must have first gone through a formalized quality assurance framework (e.g., peer review). What is more, the PBRF also includes more ethereal considerations such as “contribution to research environment.” While the criteria here overlap notably with that which the RAE requires in its peer esteem category, the main difference is that the PBRF makes the criteria explicit. The RAE is more open-ended in the sense that the assessed researchers have more flexibility (and less guidance) to determine what must be included. Nevertheless, some of the categories in the PBRF’s research environment section are still notably vague and subject to much interpretation (e.g., “contributions to the research discipline”). Though the overlap between the two is clearly evident it is worthwhile to note that New Zealand’s system unashamedly co-opted what it regards as the best components of the RAE. Moreover, the NZ system is still under construction in the sense that the national

authorities have embarked on a wide range of consultations with various representatives in the sector to improve and streamline the RPM and the research assessment system. In other words, New Zealand regards the implementation of its PBRF and Quality Evaluation system as a process of ‘learning by doing’ and in this process very much regards trust and cooperation between national authorities and universities as essential ingredients.

## **8.6 Trends, and some words of caution**

The available evidence suggests that over time RPM systems have increasingly become more quantitative and output-oriented than in the past. No clear-cut explanation for why is readily apparent, though the country examples do provide some indication that two factors have jointly played a role. These are the growing scarcity of public funding and the drive for greater transparency.

The gradual adoption of tuition fees, ambitious European-wide proposals to increase resources available for research and development, and fears about the impending brain drain to the United States all support the idea that changes in the way Europe pursues and conducts research are imminently necessary. This seems to be putting those actually doing the research in a complicated position. However, this report will not treat this undoubtedly important issue.

Before discussing any (wider) implementation of RPM in the Netherlands and drawing from the international experiences in this area, it is useful to highlight three background issues, irrespective of which approach is adopted. These issues are particularly important when RPM is used for making international performance comparisons. The three issues are:

- the context of broader innovation systems differs between countries – the role of universities within a country’s broader innovation system will affect expected performance on key quality measures such as research publications and citations. For instance, higher publication counts can be expected in countries where universities largely focus on curiosity driven basic research, than in countries where they are expected to engage heavily in applied research activities in collaboration with industry;
- the focus and structure of institutions may differ between and within countries – both within and across countries, different higher education institutions may have a different degree of focus on the conduct of research. This will affect research output performance. For instance, institutions with a heavy focus on teaching rather than research would not be expected to generate the same level of research outputs as institutions that have research-only centers or where academics have minimal time commitments for teaching;
- performance across different subject areas will be driven to an extent by broader economy wide factors – if a country or region has economic strengths in areas related to a particular field of research, it may be expected that its universities will tend to record stronger performance in such fields of research than in fields where there is no connection to the wider economy. This may be explained by the fact that areas of research with connections to the wider economy will tend to generate greater external funding support and researchers will benefit from collaborations with related researchers within industry. For instance, it would be expected that Eindhoven university would be relatively stronger in electrical engineering than the University of Twente.

In other words, RPM systems will have to be flexible enough to allow institutions to show their particular mission and areas of strength. RPM systems, and the indicators that feed into them, will need to be fair. The practices we encountered in the six countries all show that no RPM system is perfect at the outset.

## Appendix to chapter 8. – Summary of 6-country findings

	UK	Germany	Norway	New Zealand	Australia	Flanders
<b>Purpose</b>						
Inform policy decisions	x	x	x	x	x	x
Allocating funding	x		x (from 2005)	x	x	x
<b>Indicators</b>						
Input oriented	x	x				
Output oriented	x	x	x	x	x	x
Process oriented	x			x	x (RRTMR)	
Quality oriented	x			x		x
<b>Indicators distinguish between hard/soft sciences or different academic disciplines?</b>	yes	yes	?	yes	no	yes
<b>Who collects data</b>						
Institutions				x	x	x
Government	x	x		x	x	x
3rd Parties		x	x			
<b>Research-performing institutions already required to collect indicator data?</b>	Yes, some (research degree completions)	yes	yes	yes, some (research degree completions)	yes, some (research degree completions)	yes
<b>Indicators based on existing internationally-collected statistics (e.g. OECD)?</b>	yes for some	yes for some	yes for some	no	no	yes for most
<b>How often are indicators produced?</b>	SET - annually RAE - every 6 years	DFG - every 3 years BMBF – periodically CHE – annually	RCN - every 2 years NIFU/STEP every year	research income and degree completions: annually; quality data: every 3 years	annually (but RRTMR reports are temporarily (?) stopped)	annually
<b>Indicators distinguish between basic, applied and strategic research?</b>	yes	yes	?	no, only in a very indirect way	no, but in RRTMR it is addressed	yes
<b>Indicators focus on innovation and commercialization</b>	yes	yes	yes	no	no, but addressed qualitatively in RRTMR	yes, but more on innovation

## 9 Research Prestatiemeting in Nederland

### 9.1 Inleiding

Onderdeel van het project “prestatiemeting universitair onderzoek” was het voorbereiden van een expertbijeenkomst over het centrale onderwerp, de meting van onderzoeksprestaties. In de bijeenkomst werd van de aanwezige deskundigen een oordeel gevraagd over de mogelijkheden en haalbaarheid van prestatiemeting in Nederland en van gedachten gewisseld over prestatiebekostiging. De discussie vond plaats aan de hand van een aantal stellingen – over de doelen gesteld aan RPM, de buitenlandse ervaringen, en de onderliggende gegevens. De stellingen zijn in de appendix bij dit hoofdstuk opgenomen. Alvorens verslag wordt gedaan van de expertbijeenkomst (in paragraaf 9.3) maken we kort enkele opmerkingen over de stand van zaken met betrekking tot research prestatiemeting in Nederland.

### 9.2 RPM in Nederland

We bespreken in deze paragraaf in vogelvlucht de bronnen/instaties waar informatie over onderzoeksprestaties van Nederlandse universiteiten is te vinden. Achtereenvolgens gaan we in op de onderzoeksvisitaties, de VSNU, het NOWT, het CBS en de ministeries.

#### *Onderzoeksvisitaties*

De meest in het oog springende wijze waarop researchprestaties in Nederland in beeld worden gebracht is via het systeem van onderzoeksvisitaties. Deze externe onderzoeksbeoordeling, die sinds 1993 wordt georganiseerd, is een integraal onderdeel van de kwaliteitszorg van de universiteiten. Elke vijf tot zes jaar wordt al het universitair onderzoek in een bepaald vakgebied op grond van een landelijk overeengekomen (VSNU-) protocol beoordeeld door een onafhankelijke commissie van gezaghebbende experts in het veld van onderzoek. Per jaar zijn zo rond de zes commissies in hun respectievelijke velden actief. Het (openbare) visitatierapport geeft een overzicht van de kwaliteit, productiviteit, de relevantie en de duurzaamheid en vitaliteit van het beoordeelde onderzoek aan de Nederlandse universiteiten. Een en ander komt tot uitdrukking in een cijfer (schaal 1-5). De onderzoeksbeoordelingen vonden tot 2004 plaats onder auspiciën van de afdeling Kwaliteitszorg van de VSNU. In 2004 is deze afdeling verzelfstandigd in de Stichting *Quality Assurance Netherlands Universities* (QANU). Het (recent vernieuwde) *Standard Evaluation Protocol 2003-2009 for Public Research Organisations* schrijft voor dat universiteiten niet alleen de inhoud van het onderzoek, maar ook management, strategie en missie van een instituut laten beoordelen. Het protocol laat ruimte voor een beoordeling van één of meer instituten binnen één universiteit of voor een vergelijking met collega-instituten in binnen- en buitenland.<sup>73</sup> In vergelijking met het voorgaande protocol betekent dit dat niet meer gesproken kan worden van een dekkend systeem met comparatieve informatie over onderzoek in een discipline.

In de recente onderzoeksvisitatie van de discipline Farmacie werden niet alleen de gebruikelijke uitkomsten van onderzoek zoals publicaties beoordeeld (zie paragraaf 1.4 van dit rapport), maar ook de maatschappelijke impact – *the academic and societal impact and networking* – van het onderzoek. Dit vond plaats met de Sci-Quest methode.<sup>74</sup> Deze impact is met een aantal indicatoren – voorlopig nog op experimentele wijze – in beeld gebracht. Het

<sup>73</sup> Zie: <http://www.qanu.nl/comasy/uploadedfiles/Samenvatting%20SE%20Protocol.doc>.

<sup>74</sup> Zie het QANU-rapport voor het Farmacie-instituut van de Rijksuniversiteit Groningen: <http://www.qanu.nl/comasy/uploadedfiles/PharmacyRUGUCPinternet.pdf>.

gaat hier om zaken als de mobiliteit van researchers, hun interactie met de omgeving, en de selectie van onderwerpen van onderzoek.

#### VSNU

Een tweede bron van gegevens over academische onderzoek is de VSNU. Kengetallen over het universitaire onderzoek worden verzameld op basis van afspraken tussen de VSNU en OCW. De kengetallen betreffen specifiek gegevens over:

- de inzet van het wetenschappelijk personeel, naar geldstroom, naar universiteit en HOOP-gebied;
- de output van dat onderzoek, onderverdeeld naar dissertaties, wetenschappelijke publicaties en vakpublicaties;
- de instroom en het rendement van promovendi (aio's en oio's).

In tegenstelling tot de onderzoeksvisitaties gaat het hier om gegevens over de kwantiteit van het onderzoek. De gegevensverzameling staat bekend als: *Kengetallen over het universitair onderzoek* (KUOZ). De cijfers werden tot voor kort beschikbaar gesteld via de website van de VSNU – de zogeheten *Digitaal Ontsloten Cijfers* faciliteit. De meest recente cijfers zijn opgenomen in een nieuwsbrief van het Ministerie van OCW.<sup>75</sup>

#### Het NOWT

Het *Nederlands Observatorium van Wetenschap en Technologie* (NOWT) is een samenwerkingsverband tussen het *Centrum voor Wetenschaps- en Technologie-Studies* (CWTS) van de Universiteit Leiden en het *Maastricht Economic Research Institute on Innovation and Technology* (MERIT) van de Universiteit Maastricht. Het NOWT heeft tot taak om cijfermateriaal te leveren en kwantitatieve analyses uit te voeren, om de stand van zaken en trends in het Nederlandse kennissystemen in beeld te brengen. Het gaat daarbij vooral om de positie en prestaties van Nederland in een internationaal vergelijkend perspectief.<sup>76</sup> Het NOWT heeft vijf zogeheten Wetenschaps- en Technologie-Indicatoren Rapporten uitgebracht (in 1994, 1996, 1998, 2000 en 2003). Deze rapporten bevatten veel gegevens over het Nederlandse wetenschapssysteem op het macro-niveau (d.i. Nederland als geheel): R&D personeel, R&D-uitgaven en –financiering, bibliometrische gegevens (aantallen onderzoekspublicaties en de citatie-impact van die publicaties in de internationale wetenschappelijke tijdschriften), octrooien en octrooi-aanvragen, start-up en spin-off bedrijven voortgekomen uit de universiteiten en andere kennisinstellingen in de publieke sector. Verder worden bibliometrische indicatoren gepresenteerd met betrekking tot de maatschappelijke impact van wetenschap in termen van het gebruik van wetenschappelijke kennis door het bedrijfsleven.

Financiële gegevens per universiteit ontbreken echter. Wel zijn er gegevens omtrent de input in de vorm van wetenschappelijk onderzoekspersoneel (WP). Het wetenschappelijk personeel (WP) kan, conform de verschillende financieringsstromen, worden onderverdeeld in WP1, WP2 en WP3. WP1 omvat de onderzoekers die worden gefinancierd uit de 1ste geldstroom, de reguliere financiering vanuit het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW) in de vorm van universitaire fondsen. WP2 omvat de onderzoekers die worden gefinancierd vanuit de 2de geldstroom via NWO en KNAW. WP3 ten slotte omvat de onderzoekers die gefinancierd worden door onderzoeksopdrachten voor derden. Tot de belangrijkste opdrachtgevers behoren de verschillende ministeries, met name OCW en EZ, de Europese Commissie, het bedrijfsleven, en charitatieve fondsen.

<sup>75</sup> Zie: <http://www.minocw.nl/feitenencijfers/nieuwsbriefowb/25.pdf>

<sup>76</sup> De Nederlandse score wordt afgezet tegenover die voor acht focuslanden:

Australië, België, Canada, Duitsland, Finland, Verenigd Koninkrijk, Zweden en Zwitserland.

Ook zijn er gegevens (gebaseerd op de ISI statistieken) omtrent de wetenschappelijke prestaties in de vorm van de publicatie-output en de citatie-impact scores per universiteit (voor het geheel en per discipline).

De commerciële benutting van resultaten van universitair onderzoek kan – zij het op gebrekkige wijze – via de octrooiwervingsactiviteiten van universiteiten in beeld worden gebracht. De NOWT Indicator rapporten laten hiertoe het totaal aantal octrooien van alle universiteiten te zamen zien. De exploitatie van intellectueel eigendom bij publieke kennisinstellingen vindt ook plaats door het creëren van nieuwe bedrijven die zijn gericht op het vercommercialiseren van die wetenschappelijke kennis en technische expertise via technologieën en/of daaraan gerelateerde diensten. Het NOWT vermeldt hierbij gegevens afkomstig uit een studie in opdracht van het Ministerie van Economische Zaken.<sup>77</sup>

#### *CBS*

Jaarlijks publiceert het Centraal Bureau voor de Statistiek (CBS) het rapport “Kennis en Economie”. Inmiddels zijn er negen edities van deze publicatie verschenen.<sup>78</sup> Terwijl het hierboven genoemde NOWT-rapport zich vooral richt op het wetenschapssysteem, richt het CBS-rapport zich meer op het innovatiesysteem. In *Kennis en Economie* beschrijft het CBS kennisprocessen bij bedrijven en andere instellingen, uitgaande van het raamwerk van het nationaal innovatiesysteem. De belangrijkste onderwerpen die aan bod komen, zijn het menselijk kennispotentieel, onderzoeksinspanningen bij bedrijven en publieke kennisinstellingen, kennisverspreiding en de resultaten daarvan.

De gegevens die in *Kennis en Economie* zijn vermeld over het hoger onderwijs hebben betrekking op het personeel, de inkomsten, de uitgaven (o.a. voor R&D) en de studenten van de universiteiten als totaal. Gegevens over afzonderlijke universiteiten worden niet getoond.

Vermeldenswaard is nog de *Community Innovation Survey* (CIS), die periodiek wordt afgenomen in de EU-lidstaten door EUROSTAT in samenwerking met de nationale statistische bureaus. De Nederlandse CIS-enquête wordt uitgevoerd door het CBS. In deze innovatie-enquête wordt tevens melding gemaakt van het belang van universiteiten en niet-universitaire (publiek en/of private) onderzoeksinstituten als informatiebron voor innoverende Nederlandse bedrijven en de samenwerking van bedrijven met universiteiten. De cijfers uit deze enquête worden wederom gepresenteerd op het niveau van de sector universiteiten als geheel.

#### *Ministeries*

Ook de ministeries van Onderwijs, Cultuur en Wetenschap en van Economische Zaken publiceren op gezette tijden informatie over de onderzoeksprestaties. We noemen hier de OCW-publicatie *Kennis in Kaart 2004 – Gegevensbasis HOOP*, een van de resultaten van het vorig jaar gepubliceerde Hoger Onderwijs en Onderzoek Plan (HOOP) 2004. De gegevens die *Kennis in Kaart* bevat over onderzoeksprestaties zijn evenwel schaars (onder andere een tabel met per universiteit het aantal promoties), maar zullen in de toekomst met andere indicatoren verder worden aangevuld.

#### *Ten slotte*

Uit het bovenstaande mag worden opgemaakt dat vergelijkende informatie over de researchprestaties van Nederlandse universiteiten relatief schaars is. Dit geldt niet voor bibliometrische gegevens (die zijn in redelijk grote mate en op detailniveau beschikbaar), maar wel voor andere informatie over het academisch onderzoek. Veel prestatiegegevens zijn alleen op geaggregeerd niveau beschikbaar en worden niet jaarlijks gepubliceerd.

---

<sup>77</sup> *Researchers op ondernemerspad: internationale benchmark naar spin-offs van kennisinstellingen*, Ministerie van Economische Zaken, 2003.

<sup>78</sup> De nieuwste is getiteld: *Kennis en Economie 2004* (CBS, 2005)

### 9.3 Verslag expertbijeenkomst RPM

Deze paragraaf bevat een verslag van de expertbijeenkomst die in het kader van het onderhavige project is georganiseerd<sup>79</sup>. We geven een overzicht van gedachten, meningen en vragen rond de kernthema's die bediscussieerd zijn:

- de doelstellingen van research prestatie meting;
- het leren van de buitenlandse ervaringen;
- de te hanteren methodologie en onderliggende gegevens (met name met betrekking tot het meten van de valorisatie van kennis).

#### *Rationale van research prestatie meting*

Het uitgangspunt voor de discussie over prestatie meting is dat het Ministerie van OCW op dit moment niet inzichtelijk kan maken aan de volksvertegenwoordiging en andere ministeries wat de onderzoeksprestaties zijn die met publieke middelen door universiteiten tot stand worden gebracht. Daarmee ontbreekt een deel van de legitimiteit van de bekostiging van de instellingen. Door onderzoeksprestaties te meten, worden verschillen en overeenkomsten tussen instellingen duidelijk. De sterktes en zwaktes op instellingsniveau worden zichtbaar. Als die duidelijkheid er is, kan vervolgens de vraag beantwoord worden of dit consequenties zou moeten/kunnen hebben in termen van beleidsvoering (inclusief bekostiging) en, zo ja, welke. Een aantal reacties:

- Er ontbreekt een doelstelling: prestatie meting is nodig om kwaliteitsbeleid te kunnen voeren, of – in iets andere woorden – om sterkte/zwakte analyses van het onderzoekslandschap te kunnen uitvoeren, of – in nog iets andere bewoordingen – om te bepalen hoe middelentoewijzing nog beter kan (op nationaal niveau en op het niveau van de instelling).
- De aanleiding voor prestatie meting mag dan wel juist zijn, maar is dit nu een probleem tussen overheid en instellingen of meer een probleem tussen ministerie en Tweede Kamer? Met andere woorden: het moet in de discussie en beleidsvoering duidelijk zijn of het gaat om “beter uitleggen (transparantie en communicatie versterken)” of “prestaties verbeteren (kwaliteit)”;
- Hieraan gekoppeld lijkt er binnen de groep enige discrepantie te zijn in de analyse van de problematiek: (a) we weten nog te weinig over prestaties; en/of (b) we weten genoeg, maar kennen vooral uitvoeringsproblemen (operationalisatie, meting).
- Er wordt gewezen op de eventuele discrepantie tussen bekostigingsprikkel op nationaal niveau, instellingsniveau en op individueel niveau (onder het motto: onderzoek is mensenwerk). Hoewel er natuurlijk gewaakt moet worden voor elkaar tegenwerkende prikkels, wordt ook aangegeven dat er geen één-op-één relatie hoeft te zijn tussen nationale en instellings- en individuele prikkels. Dat wil zeggen: een instelling kan kiezen voor een (gedeeltelijk) andere set van prestatieprikkel dan de overheid.
- Gezien de ervaringen van de Nederlandse instellingen – er wordt benadrukt dat Nederland absoluut niet in de kinderschoenen staat waar het het meten van onderzoeksprestaties en prestatiebekostiging betreft – kan er in de beleidsvoering

---

<sup>79</sup> Deelnemers aan de bijeenkomst, die plaatsvond op 8 maart 2005 en onder voorzitterschap stond van dr. C. van Bochove (directeur Directie OWB, Ministerie van OCW), waren: Van Steen, Van Yperen, Dijkstra, Van der Meer (OCW), Jongbloed, Huisman (CHEPS), Canton (CPB), Spaapen, Pen (KNAW), De Boer (TUD), Meurink (CBS), Zijdeveld (NWO), Baggen (AWT), Van Leeuwen, Van Raan (CWTS), Van Steijn, Otten (VSNU), Van der Meulen (UT), Klomp, Bakker (EZ), Van der Duyn Schouten (UvT), Bennink (QANU).

in plaats van voor een top-down benadering (nationale prestatiemeting stuurt instellingen) ook gekozen kan worden voor bottom-up (wat kan de nationale overheid leren van de ervaring met prestatiemeting en -bekostiging op instellingsniveau?).

- Doelstellingen rond prestatiemeting moeten voldoende rekening houden met de verscheidenheid van de processen en producten binnen universiteiten. Dit geldt zowel voor de verschillende functies van de instellingen (fundamenteel onderzoek, onderwijsgebonden onderzoek, maatschappelijke dienstverlening, etc.) maar geldt ook voor de verschillende karakters van de disciplines (bijvoorbeeld ontwerpende faculteiten versus onderzoekende faculteiten).
- Er wordt geopperd dat een systeem van prestatiemeting aan verschillende doelen tegelijk (accountability, onderbouwing beleid, instellingsvergelijking – zie Appendix, stelling 1) tegemoet kan komen, waarbij verwezen wordt naar de verschillende doelen van de *Research Assessment Exercise* (RAE) in het Verenigd Koninkrijk (zie Hoofdstuk 3).
- De doelstelling van prestatiemeting hoeft op dit moment niet zozeer het bevorderen van de kwaliteit van onderzoek te zijn (die kwaliteit is namelijk in orde in Nederland), als wel de benutting (valorisatie) van academische kennis.

### ***Buitenlandse ervaringen***

Op basis van het onderzoeksmateriaal poneren de onderzoekers dat een incrementele *learning by doing* aanpak vermoedelijk tot de beste resultaten zal leiden. Deze stelling leidt tot de volgende opmerkingen:

- Er wordt gewaarschuwd dat *learning by doing* beperkingen heeft, vooral als het uiteindelijk om status, reputatie en verdeling van middelen gaat;
- Kosten-baten analyses van de buitenlandse stelsels/instrumenten zijn er niet of nauwelijks. Evaluatiestudies naar ervaringen van betrokkenen geven wel aan dat er sprake is van gewinning, acceptatie en toename van de transparantie.
- Er wordt aangegeven dat die ervaringen niet zo belangrijk zijn, het zou uiteindelijk om gedragseffecten moeten gaan. Accountability is daarom al een goed uitgangspunt (*no regret* optie), omdat het gedragseffecten genereert (overigens ook ongewenste).
- Voorbeelden van ongewenste effecten zijn duidelijk in het VK: er worden vakgroepen/afdelingen opgeheven bij lage RAE-scores (als gevolg van bezuinigingen), ondanks de nuttige functie die ze vervullen (o.a. via benutting van kennis en bijdrage aan diversiteit in hoger onderwijsstelsel). Denk aan de transactiekosten wanneer zo'n afdeling een herstart moet maken.
- Als positief effect van de RAE wordt genoemd: de bijdrage aan de concentratie van onderzoeksgeld.
- Ook kan men vraagtekens zetten bij het aggregatieniveau van de RAE in het VK. Hoewel het om een beoordeling van onderzoeksgroepen gaat, hebben instellingen/ departementen vrijheid in het bepalen van wie er meedoet in de RAE en of een departement inzet op hoge kwaliteit bij enkele leden van het departement of inzet op kwaliteit in de breedte. RAE weerspiegelt dus niet altijd onderzoekskwaliteit van de gehele discipline aan een instelling. Het is een strategisch spel met hoge transactiekosten.
- Uit het onderhavige rapport blijkt dat we niet zozeer naar “achterliggers” Duitsland en Vlaanderen moeten kijken, maar goed moeten kijken naar voor- en nadelen (en effecten) van Angelsaksische initiatieven.
- In dit verband wordt gevraagd of het wel zin heeft om op korte termijn in te grijpen, omdat er nog niets bekend is over de effecten van de huidige onderzoeksevaluatie in Nederland volgens het (nieuwe) Van Bommel-protocol.



### **Methodologie en onderliggende gegevens**

Discussie over methodologie gaat vooral over de derde stelling die aangeeft dat bij de meting van de valorisatie van kennis nauwelijks gebruik gemaakt kan worden van kwantitatieve indicatoren. De discussie spitst zich dan ook vooral toe op het meten van de valorisatie van (onderzoeks)kennis.

- De stelling impliceert dat vooral *peer review* inzicht in de valorisatie zou moeten geven. Kan dat wel door middel van peer review? Hoewel het (nog) niet volmaakt is, lijkt het Nieuw-Zeelandse voorbeeld te werken.
- In het algemeen is er een hoge correlatie tussen verschillende prestatie-indicatoren in het onderzoek (bijv. veel prestatieindicatoren hangen samen met bibliometrische indicatoren; krachtige wetenschappelijke universiteiten geven ook een krachtige impuls aan de regio). Toch moet voorzichtig worden omgesprongen met het meten van valorisatie/commercialisering. De grenzen van de meetbaarheid zijn snel in zicht.
- Sommigen uiten zorgen over de meetbaarheid van de valorisatie van kennis (buiten de wat meer tastbare resultaten zoals inkomsten uit contractactiviteiten, spin-offs, patenten, licenties e.d.), anderen zijn wat optimistischer en verwijzen naar bepaalde technieken (stakeholder assessments, Sci-Quest methode – zie paragraaf 9.2) die hierbij gebruikt zouden kunnen worden.
- Een goed uitgangspunt zou zijn om instellingen af te rekenen op hun missie. Als valorisatie hoog in het vaandel staat, dan moet een instelling daar ook rekenschap over afleggen.
- Er wordt met betrekking tot het ontwerpen van de methodologie en de gegevensverzameling op gewezen dat een stapsgewijze aanpak waarbij gestreefd wordt naar het creëren van draagvlak tot de beste resultaten zal leiden. Ook wordt aandacht gevraagd voor de bestuurlijke zorgvuldigheid: onzorgvuldig overheidshandelen kan tot juridische stappen leiden (zie de ervaring met deerschikking van de B-delen/onderzoeksbudgetten over ‘oude’ en ‘jonge/kleine’ universiteiten).

Overige opmerkingen rond methodologie en gegevensverzameling:

- De meningen zijn verdeeld waar het gaat om de (publieke) beloning van derde geldstroomactiviteiten. Waar enerzijds wordt aangegeven dat derde geldstroomactiviteiten een indicatie zijn van valorisatie en dienstbaarheid aan de samenleving en dat zulke activiteiten dus beloond zouden kunnen worden, wordt anderzijds geargumenterd dat een deel van de maatschappelijke activiteiten toch wel (kostendekkend, marktconform) uitgevoerd zal worden, zonder dat er een prestatieprikkel nodig is.
- Een prestatiemetingmodel moet aandacht schenken aan de variëteit van inputs, processen en outputs (tussen en binnen instellingen). Het *basket* model spreekt aan, maar roept vragen op over de vergelijkbaarheid van prestaties van verschillende instellingen op verschillende terreinen (humaniora versus economie, resp. techniek.).
- Er moet zowel aandacht zijn voor input als output en de relatie daartussen, anders krijgen we nog geen verbetering van de transparantie. Ook wordt aandacht gevraagd voor de proceskant.
- Niet alle indicatoren hoeven van te voren gedefinieerd te worden en “opgelegd” aan instellingen en hun afdelingen. In het verleden hebben verschillende groepen zichzelf gebogen over vraagstukken als de identificatie van prestaties, waaronder de valorisatie van kennis in bepaalde disciplines en de wezenlijke taak van die discipline (vernieuwen en vergroten van kennis versus toepassen van kennis). Het aggregeren van prestaties naar instellingsniveau wordt dan echter moeilijk.

- Rond prestatie­meting(en) is er vaak het probleem van scheve verdelingen: er zijn soms grote verschillen tussen de prestaties van individuen in een grotere onderzoeksgroep. Als (te) grote groepen worden voorgedragen voor evaluatie, treden verdunningseffecten op.

De voorzitter sluit de bijeenkomst af met de opmerking dat de discussie constructief was en veel inzichten heeft opgeleverd over wat kan en wat nuttig is bij research prestatie­meting. Er is echter duidelijk gebleken dat de materie complex is. Onderzoeksprestaties – zowel de wetenschappelijke prestaties als de benuttingsdimensies daarvan – zullen gedifferentieerd moeten worden benaderd en in het licht van de missie van de betreffende universiteit moeten worden gezien.

## Appendix bij hoofdstuk 9: Stellingen ten behoeve van de expertbijeenkomst

De expertbijeenkomst over *research prestatie meting* (RPM) bestreek vier onderwerpen:

1. Rationale van RPM
2. Buitenlandse ervaringen met RPM-systemen
3. Methodologie
4. Onderliggende gegevens

Voor elk van deze onderwerpen zijn een aantal stellingen geformuleerd, voorafgegaan door een korte opsomming van de bijbehorende issues.

### 1. Doelen gesteld aan RPM

RPM systemen dienen de volgende doelen:

- Verantwoording (*accountability*) van ingezette (publieke) middelen
- Onderbouwen van beleid en bekostiging op nationaal en instellingsniveau
- Instellingsvergelijking, *benchmarking* (nationaal, internationaal) en marketing

**Stelling 1:** Systemen van RPM verschillen al naar gelang het doel dat ermee wordt nagestreefd. Er bestaat geen systeem dat aan alle doelen tegemoet kan komen.

### 2. Buitenlandse ervaringen

Wil Nederland kunnen leren van buitenlandse ervaringen met RPM-systemen dan zal met de volgende zaken rekening moeten worden gehouden:

- Het belang van de specifieke nationale context
- De toenemende aandacht – internationaal gezien – voor prestaties (‘performance’)
- De overeenkomsten en verschillen tussen de zes RPM systemen beschreven in de CHEPS studie

**Stelling 2:** Hoewel de rol van universiteiten in de nationale R&D systemen verschilt tussen de diverse landen staat dat geen algemene conclusie over de invulling van RPM systemen in de weg:

Een gemeenschappelijke set van RPM-technieken en -indicatoren is een kwestie van tijd en ‘learning by doing’. Buitenlandse ervaringen leren dat een RPM systeem niet perfect hoeft te zijn om het te kunnen implementeren ter onderbouwing van het nationale onderzoeksbeleid.

### 3. Methodologie van RPM

Uit onze internationale verkenning komen de volgende thema’s naar voren:

- De definitie van onderzoek (welke activiteiten/prestaties/uitingen/vormen van dienstverlening vallen eronder/erbuiten?)
- Indicator-gedreven (op *kwantiteit* gerichte) technieken
- *Peer review* gedreven (op *kwaliteit* gerichte) technieken

**Stelling 3:** Als men naast excellentie tevens is geïnteresseerd in de maatschappelijke relevantie van onderzoek (-sprestaties) zal meer gebruik moeten worden gemaakt van systemen van (modified) peer review. (Kwantitatieve indicatoren voor valorisatie en commercialisatie van academisch onderzoek zijn onvoldoende voorhanden.)

**Stelling 4:** Een ‘basket’ van indicatoren die geschikt is om de omvang van de onderzoeksprestaties van een onderzoeksgroep uit te drukken zal (minimaal) moeten bestaan uit de volgende indicatoren (gemeten over een aantal jaren):

- aantallen publicaties (gewogen naar soort en impact),
- omvang van in competitie verworven onderzoeksinkomsten bij research councils, omvang van contractonderzoek uitgevoerd voor publieke en private partijen
- aantal verstrekte PhD en Master’s diploma’s

**Stelling 5:** Wil een systeem van RPM breed worden gedragen (en voldoende worden gevoed) door de academische gemeenschap dan zal het in hoge mate moeten zijn gebaseerd op peer review. Dit ongeacht het doel van de RPM. Het is de vraag of het in Nederland recent ingevoerde systeem van kwaliteitszorg in zijn huidige vorm geschikt is om te worden gebruikt voor RPM.

#### **4. Onderliggende gegevens**

Kwesties die spelen bij de verzameling en verwerking van de informatie die ten grondslag ligt aan RPM zijn:

- de reeds beschikbare gegevens versus de niet-beschikbare (en de redenen daarvoor)
- definitiekwesties (afbakening, wegging, discipline-specifieke kwesties)
- wie verzamelt, bewerkt, en controleert de benodigde gegevens en wat mag dat kosten?

**Stelling 6:** Hoewel ten aanzien van de meeste RPM-informatie overeenstemming over definities zal zijn te krijgen en gegevens voorhanden zijn, zullen zich niettemin problemen voordoen bij de verzameling van gegevens over de onderzoeksinkomsten en de ‘zachte’ variabelen over relevantie en valorisatie van onderzoek.

**Stelling 7:** Om rekening te kunnen houden met disciplinespecifieke zaken en invalshoeken zal de gegevensverzameling ten behoeve van RPM betrekking moeten hebben op een voldoende groot aantal groepen van disciplines.

**Stelling 8:** Mede door de in stelling 6 en 7 genoemde zaken zal een aanzienlijk deel van de gegevensverzameling niet jaarlijks plaats kunnen vinden maar slechts eens in de 4 tot 6 jaar.

## **Lijst van contactpersonen**

De volgende personen hebben ons van informatie voorzien bij het schrijven van dit rapport:

Pierre Verdoodt, Ministerie van de Vlaamse Gemeenschap

Luanna Meyer, Massey University, New Zealand

Ulrich Schmoch, Fraunhofer Institute, Germany

Christof Schiene, Wissenschaftliche Kommission Niedersachsen

Stig Slipersaeter, NIFU-STEP, Norway