

E · V A · L U · E · R E N

Het beoordelen van wetenschappelijk onderzoek in de praktijk

Rapport in opdracht van
Meta Evaluatie Commissie Kwaliteitszorg Wetenschappelijk Onderzoek

FEBRUARI 2009

p/a Koninklijke Nederlandse Akademie van Wetenschappen (KNAW)
Kloveniersburgwal 29
Postbus 19121
1000 GC Amsterdam
T 020 5510700
Dit rapport bestellen? Zie: www.knaw.nl/cfdata/publicaties/publicaties.cfm

Inhoud

Inleiding

1. Het stelsel van evalueren in Nederland 3
 2. Het Standaard Evaluatie Protocol en de Peer Review Committee 13
 3. De navolging van de aanbevelingen van de externe evaluatie 21
 4. Conclusies 25
- Bijlagen 30

Inleiding

De Meta Evaluatie Commissie heeft zich sinds 2005 - het jaar waarin zij met haar werkzaamheden is begonnen - gebogen over vele externe evaluatierapporten en bijbehorende zelfevaluatierapporten uit allerlei disciplines van de instellingen voor wetenschappelijk onderzoek in Nederland. Zij heeft zich geïnformeerd over de discussies over evalueren die in Nederland de ronde doen, voerde in 2005 en 2006 gesprekken met alle besturen van instellingen en bracht in 2007 het rapport 'Trust but Verify'¹ uit.

Najaar 2007 is er overleg geweest tussen de Commissie en haar drie opdrachtgevers, de VSNU, de KNAW en NWO. Daar is afgesproken – naast de bestudering van een nieuwe tranche evaluatie rapporten - in 2008 een verdiepingsslag te maken.

Werkwijze

Voor deze verdiepingsslag zijn 26 evaluatierapporten en bijbehorende zelfevaluatierapporten uit 9 disciplines (periode 2003-2007) bestudeerd: Archeologie, Biologie, Taal en Cultuur, Psychologie, Diergeneeskunde, Filosofie, Geschiedenis, Theologie en Techniek. De twaalf uitgangspunten van 'Trust but Verify' waren daarbij de leidraad. Verschillende onderwerpen die van belang worden geacht voor de evaluatiesystematiek zijn intensiever bestudeerd. Zo zijn: het aggregatieniveau van de evaluatie, het in het SEP gehanteerde systeem van de scores, het nut van het formuleren van een benchmark en de maatschappelijke relevantie van het wetenschappelijk onderzoek aan een nadere beschouwing onderworpen. Daarnaast zijn deze rapporten gebruikt om de bestuurlijke hantering van de uitkomsten nader te beschouwen.

Vervolgens is gekozen om voor 4 disciplines - Taal en Cultuur, Biologie, Techniek en Psychologie – al deze onderwerpen nader uit te diepen door het houden van interviews. Er is getracht een zo breed mogelijke spreiding over de verschillende onderzoeksinstellingen (KNAW, NWO en Universiteiten) en soorten evaluaties (locale en landelijke beoordelingen) te verkrijgen.

In 2008 zijn gesprekken gevoerd met 13 Nederlandse voorzitters van verschillende evaluatiecommissies. Op decentraal niveau binnen de betrokken instellingen (decanen, wetenschappelijk directeuren, instituutdirecteuren) zijn 15 gesprekken gevoerd. Binnen NWO en KNAW zijn elk 2 instituten bezocht, binnen de universiteiten 11 onderzoeksgroepen. (bijlage 2).

Alle interviews zijn gehouden aan de hand van een lijst met onderwerpen die vooraf was toegestuurd (bijlage 3 en 4). Ook werden additionele interviews gevoerd met een drietal secretarissen van de betrokken evaluatiecommissies.

De resultaten van deze activiteiten hebben hun beslag gekregen in dit rapport. In hoofdstuk 1 wordt het stelsel van evalueren besproken, in hoofdstuk 2 de Peer Review Committee en het gebruik van het Standaard Evaluatie Protocol. Hoofdstuk 3 behandelt de bestuurlijke hantering van de aanbevelingen van de evaluatie commissies. Het rapport wordt afgesloten met een samenvatting en een aantal algemene conclusies. Op grond van dit rapport is het briefadvies van de Meta Evaluatie Commissie aan haar drie opdrachtgevers (bijlage 5) vervaardigd.

1. Het stelsel van evalueren in Nederland

Het stelsel van evalueren is in 2001 vastgelegd in afspraken tussen de drie instellingen van wetenschappelijk onderzoek: de VSNU, KNAW en NWO. De basis voor deze afspraken was het rapport 'Kwaliteit verplicht'², waarin het gemeenschappelijk doel: het verbeteren van kwaliteit van wetenschappelijk onderzoek werd verwoord. De uitgangspunten waren:

- het verbeteren van de kwaliteit van wetenschappelijk onderzoek door een beoordeling naar internationale standaard van kwaliteit en relevantie
- het verbeteren van het management en leiderschap binnen de instellingen en
- het breed afleggen van verantwoording over de resultaten van het onderzoek.

Op basis van deze afspraken is het Standaard Evaluatie Protocol³ ontwikkeld, dat in 2003 is ingevoerd.

De vertrekpunten van het protocol zijn relevant voor deze rapportage:

- Er is gekozen voor een model waar een Peer Review Committee (PRC) de beoordeling uitvoert.
- De evaluaties zijn zowel retrospectief als prospectief. Dat betekent dat de PRC zowel kijkt naar de resultaten uit het verleden maar ook naar de plannen die de onderzoeksgroep heeft voor de toekomst.
- De beoordeling van deze PRC vindt plaats op basis van vier criteria: Productiviteit, Kwaliteit, Relevantie en Levensvatbaarheid.

Deze vier criteria zijn dusdanig omschreven in het protocol dat het mogelijk is zowel op basis van kwantitatieve als kwalitatieve gegevens te werken.

De bevindingen

Uit de bestudeerde externe evaluatierapporten en bijbehorende zelfevaluaties van de onderzoeksgroepen blijkt dat het stelsel in procedurele zin werkt, maar dat er ruimte voor verbeteringen is. Over doelstelling en uitgangspunten van het stelsel zijn door de geïnterviewden weinig opmerkingen gemaakt en de beoordeling vindt volgens de afspraken plaats. Er is dus geen reden het stelsel te veranderen. Er blijven echter een aantal discussiepunten:

Een nadere beschouwing van vier onderwerpen, het stelsel betreffende zijn van belang:

- Het uitgangspunt dat het verbeteren van management en leiderschap binnen de instelling verbeterd kan worden

- Het criterium maatschappelijke relevantie
- en twee gevolgen van afspraken over het stelsel:
 - De evaluatielast
 - De vrijheid de eigen onderzoekseenheid te kiezen: het aggregatieniveau

Het verbeteren van management en leiderschap

Nu evalueren en verantwoorden binnen universiteiten en instituten gemeengoed is geworden, lijkt het vragen van een extern oordeel over dit onderwerp enigszins gedateerd. Binnen de universiteiten, de KNAW en NWO wordt op een eigentijdse wijze binnen de organisaties geëvalueerd. Uit dit onderzoek is gebleken dat de universiteiten, de KNAW en de NWO op verantwoorde en degelijke wijze omgaan met de evaluatie van de organisatie van hun onderzoeksgroepen en instituten. Management en leiderschap zijn daarbij zonder twijfel onderwerp van gesprek. Bovendien zijn bij de universiteiten de krimpende eerste geldstroom en de groeiende aandacht voor het verwerven van gelden uit tweede en derde geldstroom – met de daarbij behorende voorwaarden – al jaren reden om efficiënt om te gaan met de middelen voor wetenschappelijk onderzoek.

PRC's blijken weinig aandacht te schenken aan de zeer uitgebreide gegevens die hen bereiken: management en managementstijl, besluitvormingsprocessen, de wijze van communiceren, de wijze van supervisie, het werkklimaat en werkcultuur, om er maar enkele te noemen. Deze informatie is voor een PRC niet relevant en wordt dan ook niet gebruikt.

Overweging

De verbetering van management en leiderschap kan beschouwd worden als een taak van de instellingen zelf, waardoor het protocol op dit onderwerp sterk vereenvoudigd kan worden. Alleen gegevens op hoofdlijnen betreffende de personele opbouw, de bronnen van financiering en de verwachte ontwikkelingen zijn belangrijk voor een PRC. Bij haar aanbevelingen over de toekomstige ontwikkeling van een onderzoeksgroep heeft zij deze informatie nodig.

Maatschappelijke relevantie

Het rapport van de MEC, 'Trust but Verify', geeft onder het kopje maatschappelijke en wetenschappelijke relevantie een beschouwing over en een aantal voorbeelden van het maatschappelijke belang van wetenschappelijk onderzoek. Het stuk eindigt met de opmerking dat 'een goede omschrijving van de wetenschap-

pelijke relevantie van het onderzoek ook inzicht biedt in de maatschappelijke waarde ervan’.

Uit de analyse van de externe evaluatierapporten blijkt echter dat uit het oordeel voor ‘Relevance’ niet duidelijk is af te leiden of naast de wetenschappelijke relevantie ook rekening is gehouden met de maatschappelijke relevantie, bijvoorbeeld met onderzoeksresultaten die een maatschappelijke waarde hebben. Dat is ook begrijpelijk. Het SEP vraagt wel aandacht voor maatschappelijke en economische relevantie, maar het criterium ‘Relevance’ staat dan voor zowel wetenschappelijke als maatschappelijke relevantie.

Uit dit onderzoek blijkt dat de helft van de geïnterviewde voorzitters van PRC’s geen probleem heeft met de beoordeling van maatschappelijke relevantie. Men acht zich prima in staat een oordeel te geven. Bij de beoordeling van ‘Relevance’ staat bij hen meestal niet de maatschappelijke maar de wetenschappelijke waarde voorop. Het criterium ligt dan voor de hand: de bibliometrische analyse gebaseerd op publicaties in erkende wetenschappelijke tijdschriften. De publicaties voor een breed publiek – zoals wel plaats vindt in alfa- en gammawetenschappen, dan wel een belangrijk patent of een technische toepassing in de bèta wetenschappen – worden impliciet meegewogen. Een enkele voorzitter is van mening dat maatschappelijke relevantie een non-item is. Het gaat immers om goed of slecht onderzoek dus is er geen verschil tussen maatschappelijk en wetenschappelijke relevantie.

Ook geeft een aantal voorzitters aan moeite te hebben met het beoordelen van maatschappelijke relevantie: waar staat het begrip precies voor? Wie bepaalt wat maatschappelijk relevant is? Uit de zelfevaluatie-rapporten zou duidelijk moeten blijken wat de onderzoekers verstaan onder maatschappelijke relevantie. Pas dan is de PRC in staat te oordelen.

Bij de onderzoeksdirecteuren lopen de meningen over maatschappelijke relevantie ook duidelijk uiteen. Sommigen vinden het geen onderwerp van discussie: ‘Wetenschap is per definitie niet maatschappelijk relevant. Dat is ook nooit zo bedoeld’ en ‘Ook fundamenteel onderzoek kan maatschappelijk relevant zijn omdat het uiteindelijk op de lange termijn kan bijdragen aan de oplossing van een maatschappelijk probleem’ en ‘aan praktijkonderzoek dat probleemoplossend gericht is, valt in wetenschappelijke zin weinig eer te behalen. Dus doen wij dat niet en dus zijn we niet maatschappelijk relevant bezig’.

Anderen geven, zogevoegd, voorbeelden wat in hun ogen maatschappelijke relevantie zou kunnen inhouden: het geven van onderwijs, het verspreiden van onderzoek via media en optredens, het verkrijgen van derde geldstroom onderzoek, het verrichten van regionaal onderzoek, het ontwerpen van een bouwwerk, het schrijven van een mooi essay dat bijdraagt aan inzicht in het cultureel erfgoed, of het oprichten van een aan de universiteit gelieerde holding met spin-off bedrijven.

Overweging

Er blijkt dus bij zowel de beoordelaars (de PRC's) als de onderzoeksdirecteuren grote onduidelijkheid en een verschil van inzicht te bestaan over de maatschappelijke relevantie van onderzoek. Dat is problematisch omdat het aangeven van 'het maatschappelijk nut' van wetenschappelijk onderzoek breed in de belangstelling staat.

Bij de rijksoverheid en in de politiek is de maatschappelijke relevantie van wetenschappelijk onderzoek een belangrijk onderwerp. Valorisatie en vraaggestuurde programmafinanciering – met de impliciete veronderstelling dat wetenschappelijk onderzoek maatschappelijk rendement moet hebben – zijn al jaren speerpunten van beleid. Dat blijkt ondermeer uit de agenda van het innovatieplatform⁴ en uit de adviezen van de AWT⁵.

Ook bij onderzoeksfinanciering en extra financiële beloningen voor onderzoek is maatschappelijke relevantie een belangrijk criterium. Door de toenemende omvang van de tweede geldstroom wordt het aantonen van maatschappelijke relevantie belangrijker: Het is een criterium van NWO.

De VSNU bemerkt enthousiasme voor valorisatie op alle universiteiten⁶ en de Colleges van Bestuur bevorderen de zichtbaarheid en de maatschappelijke impact van 'het' onderzoek van 'hun' universiteit^{7,8}.

Het opstellen van een theoretisch kader⁹ en evaluatiemethodieken¹⁰, heeft de afgelopen jaren zeker geleid tot meer inzicht in de ingewikkelde problematiek van de maatschappelijke relevantie van wetenschappelijk onderzoek, maar niet tot een beter begrip in de wetenschappelijke wereld.

Concluderend kan worden gesteld dat er behoefte bestaat aan meer inzicht omtrent de maatschappelijke relevantie van wetenschappelijk onderzoek, maar dat tegelijkertijd onderzoeksgroepen moeite hebben met het beschrijven daarvan. Tijdens de interviews met de directeuren van onderzoeksgroepen bleek dat men,

door de extra aandacht die tijdens het gesprek op dit punt werd gevestigd wel op belangrijke gedachten kwam om maatschappelijke relevantie bij een volgende gelegenheid beter te omschrijven. Een procesmatige aanpak valt in deze derhalve te overwegen.

Nieuwe theoretische of methodische inspanningen zullen niet méér bruikbare inzichten opleveren om onderzoekers en PRC's te faciliteren bij hun evaluatietaken.

Als onderzoeksgroepen zelf op heldere wijze de maatschappelijke impact van hun specifieke onderzoeksterrein beschrijven kunnen PRC's een oordeel vallen over maatschappelijke relevantie naast wetenschappelijke relevantie. De PRC's zouden dan wel in staat moeten zijn een oordeel in deze te vellen. Mogelijk kunnen zij reviewers vragen hen daarin bij te staan.

De evaluatielast

In 'Trust but Verify' wees de Meta Evaluatie Commissie al op de – in haar ogen – ongewenst stapeling van allerlei soorten evaluaties die onderzoekers, administratief personeel, hoogleraren en bestuurders bezig houden. Uit vele kanten van de wetenschappelijke wereld werd haar destijds gemeld dat de veelheid aan te schrijven evaluaties als een last werd beschouwd.

Een overgrote meerderheid van de ondervraagde voorzitters van PRC's en onderzoeksdirecteuren zijn echter van mening dat regelmatig evalueren een groot goed is en dat er geen reden is tot klagen als men de zaak administratief goed op orde heeft. Evalueren, zo zegt men, is goed, 'het houdt je scherp'. De zesjaarlijkse evaluatie volgens het SEP wordt dan ook niet als 'last' ervaren.

De meeste onderzoeksdirecteuren geven aan veel waarde te hechten aan een mid-term evaluatie, halverwege de cyclus van zes jaar. Mid-term evaluaties worden vooral intern uitgevoerd, ter voorbereiding op de zelfevaluatie en de externe beoordeling. Een enkele keer werd een mid-term evaluatie aangetroffen met een zware beoordelingsronde door een externe commissie of een verzoek aan de 'oude' PRC om een hernieuwd oordeel.

Een jaarlijks gesprek van een onderzoeksgroep of instituut en het bestuur over de voortgang acht men eveneens zinvol, als dat maar niet leidt tot weer meer in te vullen papieren. Mondeling overleg is prima. Vaak is dat ook al usance.

Minder te spreken is men daarentegen over de veelsoortigheid van protocollen die verwerkt moeten worden ten behoeve van aanvragen bij NWO, andere financieringsinstanties en de accreditatie van onderzoeksscholen en de accreditatie van het onderwijs (NVAO).

De eigen doeleinden van deze instanties genereren andere criteria en vereisen dus andere aan te leveren teksten. Zo zijn bijvoorbeeld de zwaartepunten van NWO regelmatig aan verandering onderhevig en daarmee ook de criteria voor aanvragen. Daar de instellingen – als uitvloeisel van het beleid van OCW – de gelden van NWO harder nodig hebben dan voorheen om hun onderzoek op peil te houden, wordt NWO vaak genoemd in deze.

Ook de accreditatie van onderzoeksscholen (KNAW) heeft een eigen protocol, gedeeltelijk volgens de lijnen van het SEP, maar toch weer even anders. Ook de rijksoverheid kent haar eigen vereisten voor het verkrijgen van onderzoeksgelden.

Over het integreren van onderwijsbeoordelingen in de bestaande onderzoeksevaluaties oordelen de respondenten vooral negatief: Men acht deze twee zaken te verschillend van karakter. De opname van de evaluatie van de promotieopleidingen in het SEP zou verder geëxploreerd kunnen worden.

Geconcludeerd kan worden dat de evaluaties volgens het SEP geen probleem vormen, maar de veelheid aan andersoortige protocollen voor andere doeleinden wel.

Sommige instellingen hebben zelfs aparte medewerkers en stafafdelingen aan het werk om aan deze vereisten te voldoen. De veelheid aan gevraagde gegevens bij evaluaties en financiering genereert daarmee dus –naast papier en ergernis – een ongewenste administratief potentieel.

Overigens is bij het vaststellen van het evaluatiestelsel in 2001 (Kwaliteit Verplicht) dit gevaar nadrukkelijk is onderkend, maar blijkbaar heeft dit nog niet geleid tot afstemming en van de protocollen.

Het aggregatieniveau

Bij de invoering van het SEP (2003) is bewust gekozen de instellingen de vrije hand te geven in de wijze waarop zij hun evaluatie willen inrichten. Daarmee werden de voorheen verplichte beoordeling per discipline (de z.g. VSNU-beoordeling) optioneel.

Deze vrijheid om een eigen onderzoekseenheid van beoordeling te kiezen heeft tot gevolg gehad er een enorme diversiteit aan evaluatierapporten bestaat: Evaluaties van instituten, evaluaties van onderzoeksgroepen van een universiteit, gehele of gedeeltelijke facultaire evaluaties en gehele of gedeeltelijke landelijke discipline evaluaties.

De discussie over de meerwaarde van een gezamenlijke evaluatie per discipline versus de evaluatie van afzonderlijke onderzoeksinstituten is te typeren als eeuwigdurend. Men komt er niet uit omdat aan beide opties gedegen voor- en nadelen aan verbonden zijn.

De resultaten van het onderzoek onder de voorzitters van PRC's en directeuren van onderzoeksgroepen zijn een getrouwe kopie van deze discussie over de voor- en nadelen maar geven ook enkele nieuwe gezichtspunten. Er is een onderscheid te constateren tussen de groepen geïnterviewden. Een grote meerderheid van de voorzitters van PRC's is voorstander van een disciplinair opgezette evaluatie. Men heeft een beter overzicht over het vakgebied en kan programma's met elkaar vergelijken.

Bij de onderzoeksdirecteuren ligt het genuanceerder. De directeuren onderzoek van de universiteiten zijn in meerderheid vóór een landelijke disciplinaire beoordeling. De directeuren van onderzoeksinstituten van KNAW en NWO geven – veelal om begrijpelijke redenen – de voorkeur aan een aparte beoordeling van het eigen instituut.

Overweging

Eenzijds heeft een landelijke evaluatie het voordeel dat – als de opdracht door meerdere instellingen wordt gegeven – er maar één PRC nodig is. De internationale experts op een vakgebied hoeven maar eenmaal te komen en de kosten zijn geringer.

Het is echter wel zo dat de toenemende vertakking van onderzoeksvragen in vele disciplines tot een inhoudelijke maar ook organisatorische verknoping leidt, waardoor een landelijk disciplinaire evaluatie schier onmogelijk wordt. Bovendien is een aparte evaluatie noodzakelijk als er geen sprake is van een gelijksoortig onderzoeksinstituut in Nederland (WUR bijvoorbeeld). De toename van organisatorische koepels van zeer verschillende disciplines (bijvoorbeeld in de Geesteswetenschappen) maakt dat er zoveel vakgebieden onder vallen dat lokaal evalueren voor de hand ligt.

Daarenboven speelt een ander fenomeen op een ander bestuurlijk niveau: universiteiten zijn in toenemende mate bezig met lokale profilering. In dat streven past een landelijke evaluatie minder goed. Deze tendens blijkt ook al uit de verminderde belangstelling voor landelijke onderzoeksscholen ten gunste van lokale arrangementen.

De vrije keuze van de instellingen blijft een groot goed. Ook in 'Trust but Verify' werd daar reeds op gewezen.

Met de veelsoortigheid aan te evalueren onderzoekseenheden kan echter geen inzicht worden gegenereerd in de stand van de kennis van een vakgebied in Nederland. Ook het traceren van potentiële samenwerkingsverbanden en het ontwikkelen van landelijke speerpunten behoort niet tot de mogelijkheden. Vergelijken kan ook niet meer, hoewel het maar de vraag is of dat überhaupt mogelijk is en wat de meerwaarde zou zijn.

Aan deze behoefte aan disciplinebreed overzicht aan inzicht in afstemming en samenwerking kan voorzien worden door een aparte opdracht aan de PRC. Enkele voorzitters van PRC's uit het MEC-onderzoek geven aan wel geïnteresseerd te zijn in het leveren van een bijdrage aan de landelijke plaatsbepaling van een vakgebied. Er is ook al een voorbeeld van een extern evaluatierapport dat een thematisch overzicht geeft van de discipline. Een andere optie zou kunnen zijn de beroepsgroepen per discipline te vragen dit onderwerp nader ter hand te nemen.

2. Het Standaard Evaluatie Protocol en de Peer Review Committee

De Meta Evaluatie Commissie heeft onder andere tot taak: ‘het evalueren van de werking van het Standaard Evaluatie Protocol en het doen van aanbevelingen ter verbetering van de doelmatigheid en effectiviteit van de daarin omschreven processen van kwaliteitszorg’

Daarnaast onderzoekt de MEC of de PRC over de vereiste deskundigheid en onafhankelijkheid beschikt en of haar rapporten helder zijn gepresenteerd en of haar argumentatie en analyse kwaliteit bevat”

De afgelopen jaren is gebleken dat er weinig kritiek is op het in 2003 ingevoerde protocol dat gezamenlijk door de drie instellingen VSNU, KNAW en NWO is vervaardigd.

De twee thema's: het Standaard Evaluatie protocol (SEP) en de Peer Review Committee's (PRC's) zijn onderwerp van gesprek geweest in de interviews bij voorzitters van PRC's en directeuren van onderzoeksgroepen.

HET STANDAARD EVALUATIE PROTOCOL (SEP)

Tevredenheid

Zowel voorzitters van de PRC's als de onderzoeksdirecteuren blijken in het algemeen tevreden te zijn over het SEP. Men beschouwt dit protocol allerwegen als een prima raamwerk om 'effectief te evalueren'. 'Er moet nu eenmaal een protocol zijn, en dit voldoet als leidraad prima'. Uiteraard zijn er concrete op- en aanmerkingen en suggesties voor verbeteringen:

Bij alle evaluaties wordt dit Standaard Evaluatie protocol gehanteerd, veelal aangevuld met een eigen protocol van de instelling en soms aangevuld met extra vragen van het Bestuur van de betreffende instelling. De aanvullende protocollen wijken nauwelijks af van het SEP, ze verdiepen en spitsen toe.

De omvang en de criteria

Zo de voorzitters en onderzoeksdirecteuren opmerkingen over het protocol zelf maken, hebben deze betrekking op vereenvoudiging. Het protocol is op zichzelf al dusdanig omvangrijk dat de secretarissen vooraf een excerpt van het SEP maken en dat de PRC's vooraf toezenden. Dat is ook de mening van de geïnterviewden. Een aantal geïnterviewden wil ook af van 'ellenlang proza'. Uitgebreide verhande-

lingen over interne administratieve aangelegenheden, personeel, financiën, leiderschap en management, zijn niet relevant voor een PRC. Ook uitgebreide lijsten met allerlei publicaties zijn overbodig en worden niet gelezen. Naar de mening van de geïnterviewden kan volstaan worden met een korte typering van de onderzoeksgroep in termen van omvang, financiering met daarnaast drie tot vijf belangrijkste wetenschappelijke en overige publicaties. Dit spoort met het KNAW rapport over kwaliteitszorg¹¹. In tegenstelling tot de opvatting van de Commissie Kwaliteit van de KNAW meent men echter dat het criterium ‘levensvatbaarheid’ niet geschrapt mag worden maar in afgeslankte vorm in het stelsel gehandhaafd moet blijven.

De scores

Wetenschappers zouden hun vak verloochenen als er geen discussie zou zijn over het al dan niet gebruiken van cijfers bij beoordelingen. Uiteraard verschillen dus de meningen van de geïnterviewden over het gebruik van de vijfpunts-schaal in het SEP.

In feite gaat het hier om de achterliggende keuze voor een evaluatiesystematiek. Valt de keuze op een kwantitatieve, meetbare beoordeling van het onderzoek, dan moeten er exacte criteria ontwikkeld worden op grond waarvan er een cijfermatig, objectief oordeel plaats kan vinden. In dat geval – zo zei een voorzitter – ‘kan de PRC wel thuis blijven en wordt de beoordeling een administratieve aangelegenheid per mail’.

Valt de keuze op een oordeel van gewaardeerde experts in een vakgebied, dan betekent dat per definitie een menselijk oordeel. Men geeft zich dan over aan het oordeel van een deskundige vakgenoot en geeft deze commissie van wijzen vertrouwen. Een ‘kwalitatief’ oordeel, het woord zegt het al.

In feite is er met de keuze voor het SEP in 2003 gekozen voor het laatste, maar met gebruik van – in de zelfevaluaties aangedragen – cijfermateriaal en met een eindoordeel in cijfers. Zowel de deelprogramma’s als de totale geëvalueerde onderzoekseenheid in totaliteit krijgt een waardering in cijfers.

Hoewel er altijd discussie zal blijven over de gebruikte methode, zijn zowel de voorzitters van PRC’s als de onderzoeksleiders het grondig eens dat ‘deze verschillende methoden elkaar aanvullen’. ‘Er moet zowel geoordeeld worden op een basis van kwantitatieve gegevens (bibliometrie) als op basis van kwalitatief materiaal (gesprekken, ervaring en kennis). Beide benaderingen zijn waardevol’.

Uit de externe evaluatierapporten blijkt dat de PRC's inderdaad de objectieve gegevens die vermeld worden in het zelfevaluatierapport (bibliometrie, stakeholders-analyse, aantal promovendi, verworven subsidies) gebruiken en meewegen in hun (uiteindelijk cijfermatig) oordeel. Een cijfer is herkenbaar en is inhoudelijk onderbouwd, en, zo blijkt, men kan daar – disciplinebreed – prima mee uit de voeten.

Uit de interviews komen geen duidelijke suggesties die het verlaten van de vijfpuntschaal rechtvaardigen.

De erosie van de scores

In 'Trust but Verify' werd er al op gepreludeerd. Er is sprake van hoge waarden, zowel het overall cijfer als de cijfers op programmaniveau zijn vaak 'in score vier of vijf' te vinden. (zie bijlage 6). Als er al een lage score wordt uitgedeeld is er duidelijk wat aan de hand: een groep die al langer slecht presteert, een hoogleraar die met emeritaat gaat, te weinig financiële expansie of een magere lijst excellente publicaties. Een hoge score (vier maal vijf) is ook goed herkenbaar: een excellente onderzoeksgroep.

De erosie van de scores werd herkend door de geïnterviewden. Soms geeft dat aanleiding tot beschouwingen, maar meestal wordt de conclusie getrokken dat het onderzoek in Nederland gewoon van goede kwaliteit is.

Het zou aanbeveling verdienen dat zowel de Peer Review Committees als de instellingen zich realiseren dat een vijfpuntschaal alleen goed werkt als het realistisch gemiddelde: een drie, als uitgangspunt wordt genomen.

DE PEER REVIEW COMMITTEE

Bij de bestudering van de eerste tranche rapporten is in het rapport van de MEC 'Trust but Verify' geconstateerd dat er geen reden was om opmerkingen te maken over de samenstelling en rekrutering van de PRC's. Naast de bestudering van de tweede tranche rapporten is nu ook de direct betrokkenen, de voorzitters van PRC's en de directeuren van onderzoeksgroepen, gevraagd naar de PRC en haar werkwijze.

Op een PRC wordt een zware wissel getrokken. Een onafhankelijke integere beoordeling is van groot belang omdat de consequenties van de beoordeling groot kunnen zijn. De externe beoordeling werpt zijn schaduw vooruit.

Zo staan plannen om zwak onderzoek te laten fuseren of voornemens tot een reorganisatie meestal al vermeld in de zelfevaluatierapporten. Het Bestuur stelt vooraf vaak extra vragen aan de PRC, soms direct, soms via eigen protocollen. Vragen die bestuurlijk leven. Dat kan variëren van een advies over de naamgeving van een onderzoeksdepartement tot een expliciete vraag naar de levensvatbaarheid van een onderzoeksgroep. Een PRC opereert in dit krachtenveld.

Kwaliteit, onafhankelijkheid en rekrutering

De PRC leden zijn ervaren hoogleraren in het te evalueren vakgebied, veelal uit het buitenland. Ze zijn nieuwsgierig naar de ontwikkelingen van hun vak in Nederland, en kwijten zich serieus van hun taak. Dat vinden zowel de – veelal Nederlandse – voorzitters, maar ook de beoordeelden, de onderzoeksleiders en decanen. PRC's hebben dus kwaliteit.

De PRC's zijn breed samengesteld. Voor diverse stromingen en richtingen binnen het vakgebied is er meestal een PRC-vertegenwoordiger. Essentieel is een bekwame voorzitter (bij voorkeur een Nederlander): een goede voorzitter bewaakt het proces, zorgt voor evenwicht tussen de commissieleden en hun inbreng en is bekend met de Nederlandse setting.

De PRC-leden worden – altijd in overleg met de betrokken onderzoeksgroep –, soms in overleg met de eerder aangezochte voorzitter – gevraagd door het Bestuur. In drie gevallen werd een opmerking gemaakt over de buitenlandse leden van een Peer Review Committee. Eén maal om aan te geven dat een specifiek Nederlands onderzoeksterrein niet gebaat is met een buitenlands commissielid. Tweemaal werd opgemerkt dat de internationale leden weinig kennis hebben van het Nederlandse onderzoeksbestel. Dat uitte zich bij de beoordeling van output en relevantie. Een kort schriftelijk resumé vooraf over de structuur van de Nederlandse onderzoekswereld kan nuttig zijn voor buitenlandse commissieleden.

In een aantal gevallen kreeg de PRC van het bestuur de boodschap mee dat er bestuurlijke consequenties zouden worden verbonden aan een lage score. Meerdere voorzitters hebben deze boodschap als oneigenlijke druk ervaren.

Werkwijze

De PRC-leden hebben vooraf omvangrijke zelfevaluatierapporten doorgenomen. Ze gaan altijd op site-visit waar ze inzicht in het onderhavige onderzoek krijgen en zich breed oriënteren en gesprekken voeren met de onderzoekers. Uiteindelijk blijkt dat men zich – in de paar dagen dat er in Nederland gewerkt wordt – een goed beeld kan vormen.

Vrijwel alle commissies streven ernaar om aan het einde van hun bezoek een conceptevaluatierapport gereed te hebben, dat alleen nog in detail later uitgewerkt moet worden. Er wordt meestal gewerkt met een eerste en tweede reviewer per deelprogramma.

De overeenstemming in de commissie over het oordeel en de aanbevelingen verloopt snel en soepel. De uiteindelijke conclusies komen dus relatief gemakkelijk tot stand. Zowel de commissieleden zelf als de beoordeelde zijn tevreden over de kwaliteit van de uiteindelijke rapportage. Een enkele kanttekening wordt gemaakt, bijvoorbeeld over de eigen voorkeuren van commissieleden en de daaraan gekoppelde deelbelangen.

Breed is men van mening dat in het SEP opgenomen zou moeten worden dat een PRC altijd op site-visit gaat en ook gesprekken voert met aio's en postdocs. In ons onderzoek maakten wij niet anders mee, maar het kan geen kwaad dat verplicht op te nemen.

Hoor en wederhoor

De beoordeelde onderzoeksgroep krijgt het rapport vooraf te zien en kan in alle gevallen reageren op feitelijke onjuistheden. Soms leidt dat tot lichte wijzigingen, soms ook niet. Een aantal malen is de reactie van de directeur onderzoek danwel de decaan opgenomen in het uiteindelijke rapport.

Procedureel is het gebruikelijk dat de betreffende decaan/onderzoeksleider een reactie vervaardigt op het rapport en dat het College van Bestuur een standpunt inneemt. Dat kan soms veel tijd vergen. Het komt een enkele keer voor dat de periode tussen het afrondende eindrapport van de PRC en het formele eindoordeel van het College van Bestuur meer dan een jaar is. Meestal is het sneller.

Na het afronden van de rapportage horen de PRC-leden zelden of nooit wat er met hun conclusies is gebeurd. Een uitzondering troffen we aan bij een mid-term evaluatie, waarbij de 'oude' PRC werd ingeschakeld om een advies te geven.

Het oordeel van de PRC

De analyse van de 26 evaluatierapporten geeft geen aanleiding tot zwaarwichtige opmerkingen over de helderheid van het oordeel van de PRC. De argumentatie is op programmaniveau goed herleidbaar en inzichtelijk. Een enkele PRC voorzitter vindt dat de PRC beter zou kunnen motiveren hoe het cijfer tot stand is gekomen, maar de meesten zijn tevreden over hun prestaties.

De onderzoeksdirecteuren kunnen met de wijze waarop het oordeel wordt gemotiveerd goed uit de voeten.

Waarop baseert de PRC haar oordeel?

In ‘Trust but Verfy’ wordt de noodzaak van een helder omschreven doelstelling in de zelfevaluaties gekoppeld aan een ‘prospectieve beschrijving waar men op uit is’ zwaar aangezet. Met deze aanscherping van de gebruikelijke handelwijze in de zelfevaluaties zou men dan duidelijke (toekomstige) doelstellingen moeten formuleren waarop men beoordeeld wil worden. Ook zou de onderzoeksgroep zelf aan moeten geven welk ijkpunt zij daarbij voor ogen heeft. Deze aanbevelingen uit ‘Trust but Verfy’ kunnen nog niet van invloed zijn geweest op het huidige onderzoek van de MEC; ze dateren immers van voor 2007. Wel zijn de zelfevaluatierapporten bezien op dit onderwerp en is in de interviews ingegaan op deze problematiek.

De bestudeerde zelfevaluatierapporten geven wel doelstellingen aan – met een veelal vage of zeer algemene formulering – maar zelden of nooit een heldere omschrijving van een streven of missie waar concreet meetbare onderwerpen uit af te leiden zijn. Mochten de onderzoeksgroepen al een duidelijk streven hebben, dan is dat vrijwel nooit geformuleerd in hun zelfevaluatierapport.

In de interviews is zowel aan de voorzitters van PRC’s als aan de directeuren onderzoek gevraagd wat zij als ijkpunt hebben gebruikt of zouden kunnen gebruiken in hun respectievelijke rapporten.

Het blijkt een lastig onderwerp. In eerste instantie wordt vaak gerefereerd aan de bibliometrische analyse van publicaties en de CWTS scores, een relatief eenvoudig kwantitatief ijkpunt. Al snel komt dan – zowel bij voorzitters van een PRC als bij onderzoeksdirecteuren – de bemerking dat een CWTS-analyse weliswaar nuttige informatie verschaft, maar nooit als een verkapt ijkpunt gehanteerd mag worden. Ook noemt men het aantal promovendi, de hoeveelheid verworven subsidies uit 2e en 3e geldstroom of een instituut in het buitenland met hetzelfde karakter als voorbeelden van een ijkpunt.

Beide geïnterviewde groepen noemen wel regelmatig de wenselijkheid dat de onderzoeksgroep zelf een eigen ijkpunt en een duidelijk doel waar men naar streeft de komende jaren formuleert in de zelfevaluatie.

De overgrote meerderheid van de PRC-voorzitters vindt dat een goede commissie in staat is te oordelen op basis van eigen expertise: hun kennis van het vakgebied en wat daar actueel speelt. Een ijkpunt ontwikkelt de PRC werkendeweg.

Het collectieve besef in de commissie over goed, matig en slecht onderzoek in het vakgebied is maatgevend. Eigenlijk is de commissie zelf het ijkpunt. De directeuren onderzoek kwamen uit zichzelf niet met dit idee.

Overweging

Het formuleren van duidelijke doelstellingen van het onderzoek en het aangeven van een ijkpunt waar het onderzoek geplaatst kan worden is geen algemeen geaccepteerd onderwerp in de kringen van wetenschappelijke onderzoeksgroepen. PRC's blijken er ook niet zo'n behoefte aan te hebben: zij redden zich wel.

Plannen en verwachtingen voor de nabije toekomst van de te evalueren onderzoeksgroep vormen echter zeer belangrijke informatie voor het werk van een PRC. Een korte, krachtige goed gemotiveerde formulering van de voorgenomen activiteiten – zowel inhoudelijk als organisatorisch – zowel in kwalitatieve als kwantitatieve termen – is zinvol om oordeel en aanbevelingen te kunnen formuleren. Als dit in de zelfevaluaties duidelijk omschreven is, kan een PRC oordelen of die doelstelling behaald is en of de weg erheen begaanbaar is. Het onderzoek wordt dan beoordeeld op basis van de eigen aangegeven maatstaven. De PRC kan daar met haar wijsheid en expertise een oordeel over vellen en een aanbeveling over formuleren.

Het uiteindelijke oordeel en de aanbevelingen van de PRC kunnen dan scherper, meer 'to the point' en meer overwogen geformuleerd worden.

Overige aanvullingen voor het SEP

Uit de interviews zijn verder de volgende waardevolle aanbevelingen gekomen, die de opstellers van het nieuwe SEP-protocol kunnen overwegen.

- Enkele geïnterviewden geven aan dat het zinvol zou zijn de evaluatie van de onderzoeksmaster en de Phd trajecten mee te nemen bij de evaluatie van het onderzoek. Er wordt overigens geen pleidooi gehouden voor het gezamenlijk evalueren van onderzoek en onderwijs.
- De onderzoeksgroep zou in de zelfevaluatie de eigen niche in haar onderzoek moeten definiëren. In het SEP ware daartoe de vraag te stellen: "Wat heeft de onderzoeksgroep bijgedragen aan de ontwikkeling van de wetenschap in deze discipline?"
- In het SEP zou de specifieke bestuurlijke kwaliteiten die noodzakelijk zijn voor het in stand en op peil houden van een onderzoeksgroep tot uitdrukking moeten komen. Het leiden van een onderzoeksgroep, het handhaven van in-

ternationaal aanzien, het op peil houden van de kwaliteit en het onderhouden van relevante netwerken, dat alles vereist bekwaamheid en inspanning.

3. De navolging van de aanbevelingen van de externe evaluatie

De Meta Evaluatie Commissie heeft de taak 'de invloed van de rapporten van de externe evaluatiecommissies op het beleid en bestuur van de universiteiten, KNAW en NWO' aan een nader onderzoek te onderwerpen.

Het betreft hier het sluitstuk van het stelsel. Nadat de onderzoeksgroepen een zelfevaluatie hebben vervaardigd, het basismateriaal voor de werkzaamheden van de externe evaluatiecommissie, volgt het werk van die commissie. Zij gaan op site-visit en hebben gesprekken met betrokken onderzoekers en besturen, beoordelen het werk van de onderzoeksgroep, en schrijven hun rapport. Over de conclusies en aanbevelingen vindt hoor- en wederhoor met de onderzoekers. Dan is het rapport definitief en zijn de onderzoeksgroep en de betreffende instelling aan de beurt: de universiteit, KNAW en NWO.

Wat gebeurt vervolgens met de conclusies en aanbevelingen van een Peer Review Committee?

De beschreven aanbevelingen in de rapporten van de PRC's hebben de leidraad gevormd voor de 15 gesprekken met de directeuren van de onderzoeksgroepen van de instellingen. Daarnaast zijn de bestuursstandpunten op basis van de rapportages van de PRC's opgevraagd en besproken met dezelfde directeuren van de onderzoeksgroepen.

De inhoud van de aanbevelingen

De aanbevelingen zijn zeer verschillend van karakter. Soms zijn het suggesties en ideeën, maar in de meeste gevallen serieuze conclusies en overwegingen. Qua inhoud is de variatie nog groter: er is sprake van meer van praktische zaken zoals uitbreiding van apparatuur of bibliotheek en meer inhoudelijke aanbevelingen zoals het ontwikkelen van een nieuwe onderzoeklijn. PRC's geven ook aanbevelingen van meer organisatorische aard: de herstructurering van een onderzoeksgroep of faculteit of het verwerven van meer tweede geldstroomonderzoek of het aanstellen van meer promovendi. Het betreft vaak aanbevelingen die een extra inspanning en/of een extra financiële impuls vereisen.

De onderzoeksgroep en de aanbevelingen

Uit de gesprekken met de betrokken directeuren van een onderzoeksgroep blijkt dat oordeel en aanbevelingen van de PRC zelden echte verrassingen bevatten. De

zwakke punten waren al eerder bekend, soms al vermeld in de zelfevaluatie rapporten en meestal waren veranderingsprocessen al ingezet. Het feit dat een externe beoordeling aanstaande is betekent voor onderzoeksgroepen dat – meestal vlak voor de externe evaluatie – het zelfevaluatie rapport wordt afgerond. Daarmee wordt men – ook al door het protocol – gedwongen kritisch naar het eigen onderzoek en de onderzoeksorganisatie te kijken.

Het schrijven van een zelfevaluatie rapport gekoppeld aan het uiteindelijk oordeel van de Peer Review Committee ziet men als een belangrijk handvat om veranderingen binnen een onderzoeksgroep door te voeren. Vaak wordt de term ‘stok achter de deur’ gebruikt

De aanbevelingen van de PRC – zo blijkt uit de interviews – worden altijd serieus bezien door de directeuren onderzoek. Meestal zijn of worden ze geëffectueerd of zijn onderwerp van discussie met een hoger echelon, een faculteit of College van Bestuur.

Mocht er verschil van inzicht zijn over de aanbevelingen uit het PRC-rapport – de onderzoeksgroep kan een fundamenteel verschil van inzicht hebben over de opmerkingen van de PRC of met het standpunt van hun bestuur – dan ontstaat er intern een bestuurlijke discussie. De conclusie kan zijn dat de betreffende adviezen van de PRC niet worden opgevolgd of naar een later tijdstip worden verschoven. De betreffende voorzitter van de PRC wordt daarbij overigens nooit geconsulteerd.

De instellingen en de aanbevelingen van de PRC

Externe evaluaties zijn een regulier onderdeel geworden van het bestuur en beleid van de instellingen, zowel inhoudelijk als financieel. De aanbeveling van een PRC worden serieus genomen.

De besluitvorming over het al dan niet navolgen van de aanbevelingen is een eigen aangelegenheid van de instellingen zelf en valt onder de eigen verantwoordelijkheid.

Om inzicht te krijgen in het interne besluitvormingsproces zijn de bestuursstandpunten van de instellingen opgevraagd. Bij KNAW en NWO was dat geen enkel probleem. De bestuursstandpunten van de KNAW staan zelfs op het web. De besluitvorming binnen deze twee instellingen is goed te traceren. Bij de KNAW worden aanbevelingen puntsgewijs becommentarieerd door de instituutdirectie en wetenschapcommissie. Uiteindelijk komt het bestuur van de KNAW met een reactie op het rapport waarin zij op beide reacties ingaat en een standpunt formuleert.

Bij de NWO loopt het iets anders: Het bestuur reageert eerst en vraagt dan het betreffende instituut hoe men denkt om te gaan met de aanbevelingen. Het bestuur van het instituut reageert daar dan weer op.

Bij de universiteiten gaf de vraag naar de bestuursstandpunten buitengewoon veel problemen. De bestuurlijke reacties zijn dan ook niet compleet. Bij de universiteiten is de bestuurlijke structuur ook anders: College van Bestuur, een faculteitsbestuur en mogelijk nog een eigen instituutsbestuur. De standpunten van faculteiten en instituten zijn meestal wel aanwezig en vormen het basismateriaal voor het interview met de directeuren onderzoek.

De beschikbare standpunten van de Colleges van Bestuur van de universiteiten variëren sterk in omvang: van een korte mededeling dat het betreffende rapport geen aanleiding gaf tot uitgebreide maatregelen tot omvangrijke verbeter- en heroriëntatieplannen.

Concrete gevolgen

Goede beoordelingen worden door meerdere besturen van de eigen instelling gehonoreerd met een extra financiële injectie. Een score 3 zit in de gevarezone: een goed gesprek of drastische maatregelen van hogerhand behoren dan tot de mogelijkheden. Er is dan vaak al sprake van - reeds ingezette - reorganisatie of samenvoeging.

Het is geen automatisme dat een goede beoordeling en hartelijke aanbevelingen het bestuur inspireren tot een financiële impuls. Dat hangt van het eigen universitaire zwaartepuntenbeleid of het eigen het beleid van NWO en KNAW af. Het ontvangen van een financiële premie is vaak het gevolg van goed gedrag binnen het vigerende lokale universitaire beleid: het werken aan gezamenlijke onderzoeksprojecten binnen een faculteit, aan lokale onderzoeksscholen en aan interdisciplinair lokaal onderzoek van meerdere disciplines en faculteiten.

In het universitaire beleid worden jaarlijks de eigen lokale speerpunten geformuleerd. Deze actuele aandachtsvelden komen voort uit de (internationale)ontwikkeling van een discipline en/of zijn het gevolg van actuele en politieke aandachtsvelden. Deze 'universitaire zwaartepunten' worden in de universiteitsraden democratisch bekrachtigd.

Overweging

Zowel uit de informatie van de directeuren van de onderzoeksgroepen als uit het schriftelijke materiaal van de instellingen is voldoende informatie verkregen om te

kunnen concluderen dat de instellingen serieus omgaan met de aanbevelingen van de PRC. Deze conclusie komt overeen met die van de Commissie Dynamisering uit 2006¹². Op basis van dat onderzoek bij de instellingen werd geconcludeerd dat er goed rekening gehouden wordt met de resultaten van de onderzoeksvisitaties.

Inzicht in en verantwoording van wetenschappelijk onderzoek kan beter. Veel zelfevaluatie-rapporten zijn te vinden op het web. De bestuursstandpunten van de KNAW ook. De universiteiten en NWO zouden dat kunnen navolgen. Openbaarheid van de uiteindelijke standpunten en inzicht in de (eventuele) navolging van de aanbevelingen van de PRC is onmisbaar voor de verantwoording van publiek verschaft gelden. Daar waar kwaliteit te wensen overlaat worden maatregelen inzichtelijk, daar waar sprake is van excellentie en/of gewenste groei kan de motivering voor beloning verantwoord worden.

De besturen van de instellingen zouden verplicht de Peer Review Committees moeten rapporteren over de besluiten die zij hebben genomen naar aanleiding van de externe evaluatie van het onderzoek binnen hun instelling

4. Conclusies

Het stelsel

Evaluatie van wetenschappelijk onderzoek aan de Nederlandse universiteiten en de KNAW en NWO verloopt goed. De afspraken uit 2001 zijn nog actueel en blijken werkbaar.

Het uitgangspunt in het stelsel dat evalueren zou moeten leiden tot het verbeteren van management en leiderschap binnen de instellingen is zinvol, maar is een zaak van de instellingen zelve. Een Peer Review Committee is ook niet de aangewezen instantie om deze aspecten te beoordelen. De Peer Review Committee dient wel te oordelen over de levensvatbaarheid van de betreffende onderzoeksgroep en dient daartoe over enige relevante gegevens te beschikken.

Daarnaast kan worden geconcludeerd dat de Peer Review Committees hun werk serieus en op verantwoorde wijze verrichten. Ze opereren onafhankelijk en worden op zorgvuldige wijze geselecteerd. De rapportages zijn kwalitatief en overwogen. Zij achten zichzelf de experts op het betreffende vakgebied en dus in staat dit werk te verrichten. Hun kwaliteit en expertise maar ook hun werkwijze en rapportages worden breed gewaardeerd door betrokken directeuren van onderzoeksgroepen.

Het Standaard Evaluatie Protocol

Het Standaard Evaluatie Protocol behoeft geen ingrijpende wijzigingen. Het protocol wordt op hoofdlijnen gewaardeerd. Het is een goede leidraad om de beoordeling te verrichten, het geeft een goed kader om de zelfevaluaties te structureren.

Suggesties voor verbeteringen en aanvullingen:

- Het protocol kan sterk vereenvoudigd worden, met name – zoals ook hiervoor vermeld – de organisatorische component. De veelheid aan informatie die in de zelfevaluatierapporten gevraagd wordt over management en leiderschap is onnodig en wordt nauwelijks gebruikt door een PRC. Voor PRC's is het zicht in hoofdlijnen op de continuïteit van het onderzoek in termen van bemensing, apparatuur en financiën wel van belang.
- De PRC's zouden hun oordeel helderder en meer overwogen kunnen motiveren als de onderzoeksgroep haar doelstelling en streven voor de toekomst in de zelfevaluatierapporten exacter zou aangeven. Bij de beschrijving van de ambitie zou een zelf geformuleerd referentiepunt een duidelijker basis vormen voor het oordeel.

- Er is feitelijk geen aanleiding de beoordeling in cijfers te wijzigen, mits de PRC's goede argumentatie blijven leveren voor hun cijfermatig oordeel, vooral op het niveau van de afzonderlijke programma's. Zowel excellent onderzoek (meerdere scores vijf) als onderzoek beneden de maat (meerdere scores lager dan drie) blijven met een vijfpunts-schaal goed inzichtelijk. Te overwegen valt om daar waar de kwaliteit te wensen overlaat (meerdere scores drie of lager) te adviseren een mid-term evaluatie mede door externe vakgenoten te laten beoordelen.
- De erosie van de scores. Zoals al eerder vermeld in 'Trust but Verify' wordt geadviseerd dat de instellingen het realistisch gemiddelde, een drie, als uitgangspunt nemen. Zij kunnen dan ook hun PRC's nadrukkelijk wijzen op het nivellerend effect van hoge cijfers.
- Er is een terechte toenemende aandacht bij de instellingen om getalenteerde wetenschappelijke onderzoekers aan te trekken. De expertise die nodig is om een onderzoeksgroep tot excellentie te brengen danwel deze kwalificatie te behouden zou tot uitdrukking zou moeten komen in de beoordeling van onderzoeksgroepen en dient dus nadrukkelijk als onderwerp opgenomen te worden in het SEP.

De navolging van de aanbevelingen uit het PRC-rapport

Regelmatig evalueren wordt door onderzoekers en instellingen belangrijk en nuttig gevonden. Evalueren is inmiddels onlosmakelijk verbonden met de beleidscyclus van een instelling. Een externe evaluatie eens in de zes jaar en een eigen mid-term evaluatie elke drie jaar is volgens de geïnterviewde betrokkenen noodzakelijk om het onderzoek regelmatig zowel inhoudelijk als organisatorisch scherp te bezien. Dat gebeurt dus ook. De oordelen en aanbevelingen van een externe commissie van vakgenoten met postuur geeft gewicht aan de – in vele gevallen al aanwezige – eigen analyse en aanpak, verdiept en geeft soms nieuwe impulsen.

Op basis van de gesprekken met de directeuren onderzoek is te concluderen dat de aanbevelingen van de PRC worden opgevolgd of serieus bediscussieerd binnen de instellingen.

De organisatorische en inhoudelijke bemerkingen van een PRC over een onderzoeksgroep of instituut behoeven uiteraard niet altijd te sporen met de doelstellingen en programmering van universiteiten of van NWO en KNAW. Dat is immers een verantwoordelijkheid van de instellingen zelf.

Mocht de vraag rijzen naar een meer uitgebreide en diepgaande analyse van de wijze waarop de instellingen intern omgaan met de aanbevelingen van een Peer Review Committee dan ligt een nader onderzoek naar de interne besluitvorming gekoppeld aan een analyse van het stelsel van onderzoeksfinanciering in de rede. Op basis van de resultaten van dit onderzoek kan vertrouwd worden op het zelfsturend vermogen van de instellingen.

Aggregatieniveau

De huidige praktijk een eigen eenheid van evaluatie te kiezen is goed en onontkoombaar. In toenemende mate wordt onderzoek immers interdisciplinair en in meerdere organisatorische verbanden beoefend. Hoewel de mogelijkheid tot een landelijk inzicht in een vakgebied en een mogelijke onderlinge vergelijking dan kan ontbreken, blijft een vrije keuze de voorkeur verdienen.

Voor grote nationale investeringen, maar ook voor de toekenning van gelden voor onderzoek, wordt een landelijk overzicht van onderzoek van een discipline en inzicht in afstemming en samenwerking van instellingen binnen een vakgebied echter onontbeerlijk geacht¹³.

Aan de vraag naar een landelijk inzicht in een discipline kan mogelijk op een andere wijze worden tegemoetgekomen. De beroepsgroepen zouden deze taak op zich kunnen nemen.

Ook gaf een aantal voorzitters van PRC's aan wel geïnteresseerd te zijn deze (mogelijk) nieuwe taak nader te exploreren.

Wetenschappelijke en maatschappelijke kwaliteit

Bij de omschrijving van de wetenschappelijke kwaliteit zou zeker ook de bijdrage die het onderzoek levert aan de verdere ontwikkeling van de discipline opgenomen moeten worden. Te denken valt aan een beschrijving van de eigen onderzoeksniche in het internationale veld, of de bijdrage die de onderzoeksgroep leverde aan de voortgang van het betreffende wetenschappelijk onderzoek in die discipline de afgelopen zes jaar.

Het criterium 'relevance' in het SEP verdient nadere aandacht. Bij de beoordeling van wetenschappelijke relevantie is een bibliometrische analyse van publicaties voor veel disciplines een handzaam instrument. De humaniora, de technische en toegepaste wetenschappen en een aantal gammawetenschappen komen met dit instrument alleen niet goed over het voetlicht. Zo werd recentelijk in het rapport 'Duurzame Geesteswetenschappen'¹⁴ aanbevolen dat de KNAW het voortouw

zou kunnen nemen in de ontwikkeling van een helder en adequaat systeem van kwaliteitsindicatoren voor de geesteswetenschappen. Wellicht kan de KNAW dit verbreden in de richting van de overige genoemde wetenschappen.

Hoewel de discussie over ‘maatschappelijke relevantie’ of ‘valorisatie’ of ‘accounting’ een hoge actualiteitswaarde heeft blijkt uit de interviews dat niet alle onderzoekers en leden van Peer Review Committees dit onderwerp met dezelfde kracht onderschrijven. Eigenlijk is alleen de onderzoeksgroep zelf in staat aan te geven waar hun onderzoek waarde heeft of in de toekomst kan hebben voor de samenleving.

De beschrijving en beoordeling van maatschappelijke relevantie vereist – naast een uitgebreidere passage in het nieuwe SEP – nog een andere benadering. Gerichte gesprekken met wetenschappelijke onderzoekers uit verschillende disciplines blijken hun verbeelding te prikkelen om de maatschappelijke relevantie van hun onderzoek in het vervolg beter weer te geven. Deze procesmatige aanpak biedt goede kansen voor een betere omschrijving van maatschappelijke relevantie.

Het woud aan evaluaties

Hoewel nooit uitdrukkelijk bedoeld door de ontwerpers van dit stelsel is er een ‘woud aan evaluaties’ ontstaan. De veelheid aan soorten evaluaties en aanvragen voor financiering blijkt een groot probleem. De evaluaties en accreditatie van bijvoorbeeld de onderzoeksscholen (KNAW) of het onderwijs (NVAO) kennen in hun protocollen veel criteria die qua strekking lijken op die in het SEP, maar net weer anders uitgewerkt. Door de relatieve krimp van de eerste geldstroom zijn onderzoeksgroepen meer gedwongen elders financiering te verkrijgen. De vele protocollen en voorschriften maken dat dergelijke aanvragen veel tijd kosten. Tijd die niet besteed kan worden aan onderzoek. Het is zeer belangrijk dat VSNU, KNAW en NWO deze problematiek in eigen kring en met andere subsidieverleners – waaronder ook de rijksoverheid en de EU – gaan bespreken. Wellicht behoort een modulaire aanpak tot de mogelijkheden.

Vervolg

Het is zinvol over vijf jaar het onderzoek onder Peer Review Committees en directeuren van een onderzoeksgroep te herhalen. Een nieuwe tranche evaluatierapporten is dan beschikbaar, het nieuwe SEP zal zijn ingevoerd, en de effecten van de aanbevelingen uit ‘Trust but Verify’ en dit rapport ‘E.va.lu.e.ren’ kunnen dan worden gezien.

Literatuur

¹ Meta Evaluatie Commissie Kwaliteitszorg Wetenschappelijk Onderzoek, *Trust but Verify*, Amsterdam 2007

² Werkgroep Kwaliteitszorg Wetenschappelijk Onderzoek en standpuntbepaling KNAW, NOW en VSNU, *Kwaliteit verplicht. Naar een nieuw stelsel van kwaliteitszorg voor het wetenschappelijk onderzoek*, Amsterdam 2001

³ VSNU, NWO, KNAW: *Standard Evaluation Protocol 2003-2009 for Public Research Organisations*, 2003

⁴ Innovatieplatform: *Kennisinvesteringsagenda 2006-2016*, den Haag 2006

⁵ Adviesraad voor het wetenschap en technologiebeleid: *Alfa en Gamma stralen*, Den Haag 2007

⁶ VSNU: *Met deze wetenschap, Branchejaarverslag van de universiteiten*, den Haag 2008

⁷ Jaarverslagen 2007 van verschillende universiteiten

⁸ J.C. Stoof: *It's the quality, stupid....., rede ter gelegenheid van de opening van het academisch jaar 2008-2009 in Utrecht*, 1 september 2008

⁹ KNAW: *Judging Research on its Merits*, Amsterdam 2005

¹⁰ Spaapen, J., H. Dijkstra, F. Wamelink: *Evaluating Research in Context*, den Haag 2007.

¹¹ KNAW: *Kwaliteitszorg in de wetenschap*, Amsterdam 2008

¹² Commissie Dynamisering: *Investeren in dynamiek*, 20062

¹³ Commissie Nationale Roadmap Grootschalige Onderzoeksfaciliteiten: *Nederlandse Roadmap Grootschalige Onderzoeksfaciliteiten*, Den Haag 2008

¹⁴ Commissie Nationaal Plan Toekomst Geesteswetenschappen, *Duurzame Geesteswetenschappen*, Amsterdam 2008

Bijlage 1

Leden Meta Evaluatie Commissie

De Meta Evaluatie Commissie is als volgt samengesteld:

Prof. dr. C.P. Veerman, voorzitter

Prof. dr. H.P.M. Adriaansens

Prof. dr. J.J. Engelen

Mw. prof. dr. ir. L.O. Fresco

Dr. mr. P.W. Kwant

Mw. prof. dr. P.L. Meurs

Prof. dr. A.J.B. Zehnder

Het secretariaat van de commissie zetelt bij de KNAW.

De secretaris van de commissie is mevrouw dr. P.M. Renou

Het secretariaat wordt gevoerd door mevrouw dr. E.M. van Duuren.

Bijlage 2

Bezochte onderzoeksinstellingen:

Universiteiten:

- Universiteit van Amsterdam
- Vrije Universiteit
- Universiteit Utrecht
- Universiteit Leiden
- Radboud Universiteit
- Universiteit Maastricht
- Wageningen Universiteit
- Technische Universiteit Delft
- Technische Universiteit Eindhoven
- Universiteit van Tilburg
- Rijksuniversiteit Groningen

KNAW:

- NIOO: Nederlands Instituut voor Ecologie
- Meertens Instituut

NWO:

- NIOZ: Koninklijk Nederlands Instituut voor Zeeonderzoek
- SRON: Netherlands Institute for space Research

Bijlage 3

Aandachtspunten bij het gesprek met voorzitters van PRCs.

- 1 Kwaliteit en onafhankelijkheid externe evaluatiecommissies
 - Hoe beoordeelt u de kwaliteit en onafhankelijkheid van uw evaluatiecommissie?
 - Hoe was de werkwijze en hoe kwam u tot commissiebreed gedragen oordeelsvorming ten aanzien van de (toelichting op) scores?

- 2 Bestuurlijk vervolg uitkomsten onderzoeksevaluaties
 - Hebt u een beeld van de wijze waarop bestuurlijk met de onderzoeksevaluatie is/wordt omgegaan; zo ja, hoe zou u dat willen kenschetsen?
 - Hoe kijkt u terug op het bestuurlijke traject voorafgaand aan het definitief vaststellen van het evaluatierapport (werving commissieleden, opstellen conceptrapport, voor feitelijke onjuistheden voorleggen aan beoordeelde onderzoekseenheid, eventuele interactie tussen beoordeelde onderzoekseenheid met 'hogere gremia', verwerking commentaar op conceptrapport e.d.)?

- 3 Hanteren van een ijkpunt
 - In het SEP staat niets vermeld over het hanteren van een ijkpunt. In de praktijk zie je dan ook vrijwel nooit dat in zelfevaluatierapporten of in externe evaluatierapporten wordt vermeld waartegen de beoordeling in een internationale context wordt afgezet. Hebt u het ontbreken van een dergelijk ijkpunt gemist en hoe bent u daar in de praktijk mee omgegaan?
 - Wat zou binnen de door u beoordeelde onderzoek als ijkpunt gehanteerd kunnen worden?

- 4 Beoordeling van maatschappelijke relevantie
 - Een van de items van de beoordeling door de evaluatiecommissie is de maatschappelijke relevantie van het onderzoek. Vormde de beoordeling van dit item al dan niet een probleem?
 - Commissieleden worden primair gekozen op basis van hun wetenschappelijke reputatie. Er zijn mensen die om die reden pleiten voor het toevoegen een lid aan de commissie die de niet-specifiek wetenschappelijke aspecten

van onderzoek (bijvoorbeeld de maatschappelijke relevantie of de collectievorming in het geval van een instituut waarvan de bibliotheek een essentieel onderdeel van het wezen van het instituut vormt) beoordeelt. Wat vindt u van die suggestie?

5 Beoordeling één onderzoeksinstituut versus beoordeling discipline

- Het SEP laat instellingen vrij om zelf het aggregatieniveau van beoordeling te kiezen. Praktijk is dat beoordelingen in alle soorten en maten plaatsvinden. Wat vindt u hiervan?
- In het geval u één onderzoeksinstituut hebt beoordeeld; had u inzicht in de plaatsbepaling van het onderzoeksinstituut in de nationale en internationale context?

6 Evaluatielast

- Er wordt weleens gesproken over een ‘woud aan evaluaties’. Hoe kijkt u aan tegen de hoeveelheid en aard van de beoordelingen waarmee onderzoekers te maken hebben?
- In het geval dat u de hoeveelheid en aard van de beoordelingen waarmee onderzoekers te maken hebben als een last beoordeelt: zijn er naar uw mening mogelijkheden om deze last te verminderen?

7 Standaard Evaluatie Protocol (SEP):

- Hoe beoordeelt u de door het SEP voorgeschreven procedure bij een externe onderzoeksevaluatie?
- Hebt u suggesties voor verbetering van het SEP?

Bijlage 4

Aandachtspunten bij het gesprek met de onderzoeksinstellingen.

- 1 Kwaliteit en onafhankelijkheid externe evaluatiecommissie
 - Wat was de procedure bij de totstandkoming van de evaluatiecommissie? Wie droeg de kandidaten voor en wie stelde de evaluatiecommissie uiteindelijk vast? Wat was uw eigen betrokkenheid bij de samenstelling van de evaluatiecommissie?
 - Hoe kijkt u terug op de evaluatiecommissie in termen van kwaliteit en onafhankelijkheid?
 - Bestond de mogelijkheid om in het conceptstadium van het rapport nog commentaar te geven en wat deed de evaluatiecommissie met dit commentaar?
 - Was er sprake van een ‘extra’ protocol dat uw instelling de PRC heeft doen toekomen?
 - Is u bekend of er mondeling extra vragen aan de evaluatiecommissie zijn gesteld?

- 2 Het resultaat van de externe evaluatie
 - Wat is uw oordeel over de uitkomst van de externe evaluatie?
 - de beoordeling en advisering op de inhoud en het programma?
 - de beoordeling en advisering op organisatorische aspecten?
 - Wat is het oordeel van het College van Bestuur, respectievelijk Algemeen Bestuur KNAW/NWO op de uitkomst van de externe evaluatie?
 - Zijn er organisatorische en/of wetenschapsinhoudelijke wijzigingen doorgevoerd als uitvloeisel van het evaluatierapport? Zo ja, welke? Zo nee, waarom niet?
 - Heeft de externe beoordeling de beeldvorming over uw onderzoek in de wetenschappelijke wereld goed of slecht gedaan of speelde de externe beoordeling geen rol van betekenis?

- 3 Ijkpunt
 - In het SEP staat niets vermeld over het hanteren van een ijkpunt. In de praktijk zie je dan ook vrijwel nooit dat in zelfevaluatierapporten of in externe evaluatierapporten wordt vermeld waartegen de beoordeling in een

nationale dan wel internationale context wordt afgezet. In hoeverre is het naar uw mening mogelijk om ten aanzien van uw eigen instituut een ijkpunt te hanteren?

- Indien het mogelijk is om ten aanzien van uw instituut een ijkpunt te hanteren: Wat is het ijkpunt?

4 Maatschappelijke relevantie

- Een van de onderwerpen voor beoordeling door de evaluatiecommissie is de maatschappelijke relevantie van het onderzoek. Dit blijkt in de praktijk een lastig item te zijn. Bent u er in geslaagd dat helder weer te geven in het zelfevaluatie rapport?
- Is de evaluatiecommissie er bij de beoordeling van uw instituut naar uw oordeel in geslaagd om het aspect van maatschappelijke relevantie op waarde te beoordelen?
- Wat is naar uw mening de meest geschikte wijze waarop de maatschappelijke relevantie van uw eigen instituut beoordeeld zou kunnen worden?

5 Interdisciplinariteit en samenwerking

- Zou u het onderzoek willen kenschetsen als interdisciplinair?
Zo ja, is daar een institutionele vorm aan gegeven?
Zo ja, vloeien uit een deze interdisciplinaire benadering van uw onderzoek voor- of nadelen voort bij een externe beoordeling?
- Kunt u aangeven met welke onderzoeksinstituten in Nederland u samenwerkt? Hoe is die samenwerking tot stand gekomen?

6 Aggregatieniveau van de evaluatie

- Het SEP laat instellingen vrij om zelf het aggregatieniveau van beoordeling te kiezen. Praktijk is dat beoordelingen in alle soorten en maten plaatsvinden (zowel landelijke als lokale beoordelingen, zowel disciplinaire als multidisciplinaire beoordelingen). Wat vindt u hiervan?
- Is in het geval van de beoordeling van uw instituut naar uw mening gekozen voor het juiste aggregatieniveau van beoordeling?

7 Evaluatielast

- Er wordt weleens gesproken over een ‘woud aan evaluaties’.
Hoe kijkt u aan tegen de hoeveelheid en aard van de beoordelingen waarmee onderzoekers te maken hebben?

- In het geval dat u de hoeveelheid en aard van de beoordelingen waarmee onderzoekers te maken hebben als een last beoordeelt: zijn er naar uw mening mogelijkheden om deze last te verminderen?
- Hoe kijkt u in dit verband aan tegen de gedachte om het onderwijs en onderzoek in één keer te evalueren? In het geval u deze gedachte niet op voorhand uitsluit, aan welke voorwaarden zou dan tenminste voldaan moeten worden om over te gaan tot een dergelijke gezamenlijke evaluatie?

8 Evaluatiecyclus

- Het SEP hanteert een cyclus van zes jaar, met eens in de drie jaar een interne mid-term review en eens in de zes jaar een externe evaluatie. Hanteert u deze cyclus in de praktijk? In het geval u deze cyclus hanteert: Vindt u deze ook zinvol of doet u het ‘omdat het moet’?
- Op welke wijze houdt u bij het schrijven van het zelfevaluatierapport rekening met de uitkomsten van de vorige evaluatie (bv in de vorm van een apart hoofdstuk waarin wordt aangegeven wat met de aanbevelingen van de vorige evaluatiecommissie is gedaan)?

9 Scores

- De MEC stelde in haar eerste rapport ‘Trust but Verify’ vast dat er een erosie van de scores plaats vindt. Dit verschijnsel is na het verschijnen van het MEC-rapport alleen nog maar toegenomen. Het lijkt erop dat de 5-punt-schaal in de praktijk bijna een 3-punt schaal is geworden. Hoe kijkt u aan tegen deze erosie van de scores?
- Hoe staat u tegenover het afschaffen van de scores? Wat zou een alternatief beoordelingssysteem kunnen zijn?

10 Standaard Evaluatie Protocol (SEP)

- Hoe beoordeelt u de door het SEP voorgeschreven procedure bij een externe onderzoeksevaluatie?
- Hebt u suggesties voor verbetering van het SEP?

Bijlage 5

Briefadvies van de Meta Evaluatie Commissie aan de opdrachtgevers

Bezoekadres:
Het Trippenhuis
Kloveniersburgwal 29

Postadres:
Postbus 19121
1000 GC Amsterdam

T (020) 5510700
F (020) 6204941

Meta Evaluatie Commissie

Ons kenmerk

Doorkiesnr. (020) 5510787

E-mail Marieke.renou@bureau.knaw.nl

Onderwerp

Amsterdam, 16 december 2008

Aan president van de KNAW, Prof. Dr. R. H. Dijkgraaf
Aan de voorzitter van NWO, Prof. Dr. P. N. Nijkamp
Aan de voorzitter van de VSNU, Dr. S. Noorda

Geachte heren,

De Meta Evaluatie Commissie heeft zich de afgelopen jaren verdiept in de externe evaluatie van het wetenschappelijk onderzoek in Nederland. Zij bracht in 2007 het rapport 'Trust but Verify' uit en verrichtte nader onderzoek naar de werking van het stelsel en het Standaard Evaluatie Protocol.

Zij concludeert dat:

- het stelsel van evalueren – zoals vastgelegd in 'Kwaliteit verplicht' in 2001– in procedurele zin adequaat blijkt te werken;

- het Standaard Evaluatie Protocol juist wordt gebruikt en – grosso modo – voldoet aan de verwachtingen;
- de instellingen serieus omgaan met de conclusies en aanbevelingen van de Peer Review Committees.

De Meta Evaluatie Commissie heeft daarmee voldaan aan de haar – in het instellingsbesluit van november 2004 – opgedragen taken en verzoekt de KNAW, NWO en VSNU de Commissie derhalve te ontbinden.

Zij doet de aanbeveling de volgende punten aan een nadere beschouwing te onderwerpen:

- Er zijn verschillende instanties die het wetenschappelijk onderzoek evalueren of beoordelen. De daarbij voorgeschreven protocollen zijn zeer divers van karakter. Dat leidt tot vergaande bureaucrativering bij de onderzoeksgroepen. Vereenvoudiging moet hierbij mogelijk zijn.
- De externe evaluaties zouden bij voorkeur disciplinebreed georganiseerd moeten worden.
- De Peer Review Committees dienen verplicht een rapportage te ontvangen waarin de besluiten die de instellingen nemen naar aanleiding van hun rapport wordt verwoord.
- Het Standaard Evaluatie Protocol kan op een aantal onderdelen aangescherpt en aangevuld worden.

De Commissie doet de aanbeveling over vijf jaar het onderzoek onder onderzoeksgroepen en Peer Review Committees met een nieuw in te stellen commissie te herhalen. In januari 2009 zult u de rapportage ontvangen die is opgesteld door het secretariaat van onze commissie naar aanleiding van bezoeken aan de diverse instellingen en gesprekken met voorzitters van Peer Review Committees en onderzoeksdirecteuren. Wij hopen u in deze naar genoegen te hebben geadviseerd.

Mocht er behoefte bestaan aan een nader contact met de Commissie dan ben ik vanzelfsprekend bereid met u van gedachten te wisselen.

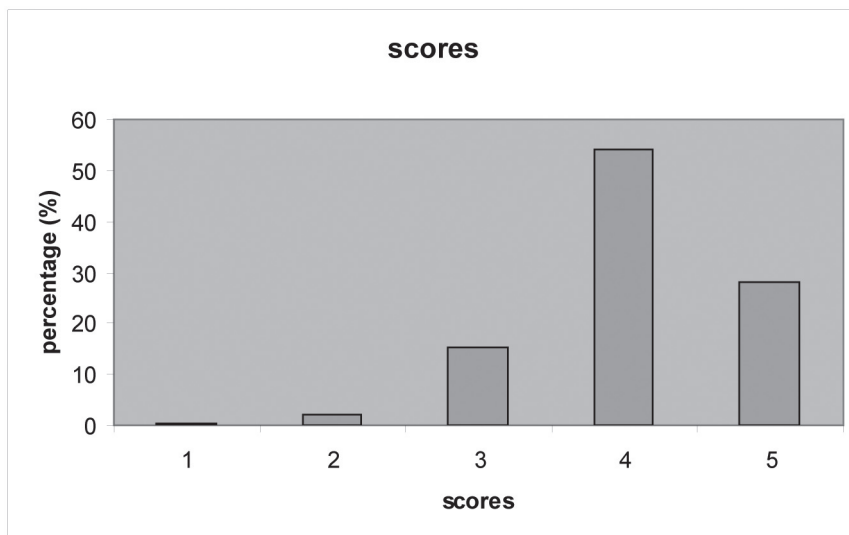
Met een vriendelijke groet,

Prof. dr. C.P. Veerman,
 Voorzitter van de Meta Evaluatie Commissie Kwaliteitszorg Wetenschappelijk
 Onderzoek

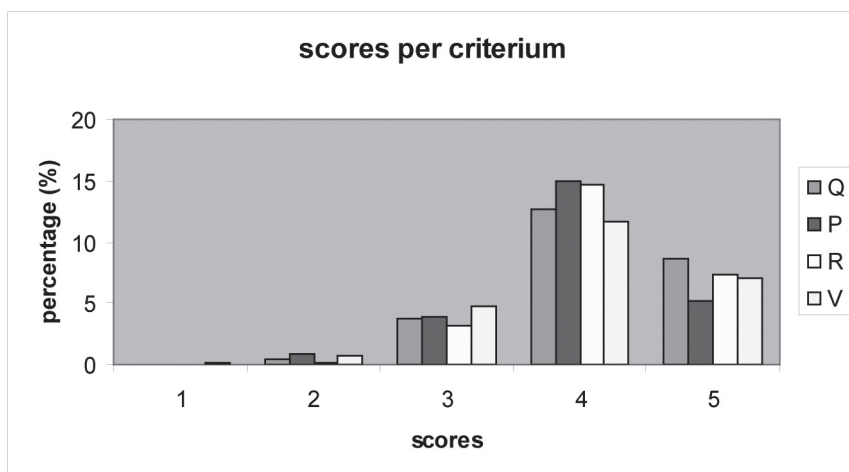
Bijlage 6

Overzicht van de scores, gegeven op programmaniveau:

Percentages van de op programmaniveau gegeven scores (N=137) in de evaluatierapporten bestudeerd voor de disciplines Taal en Cultuur, Biologie, Techniek en Psychologie.



Percentages van de op programmaniveau gegeven scores (N=137) in de evaluatierapporten bestudeerd voor de disciplines Taal en Cultuur, Biologie, Techniek en Psychologie, nu gesplitst naar de criteria als gebruikt in het SEP (Q: Quality; P: Productivity; R: Relevance; V: Vitality and feasibility).



Bureau Meta Evaluatie Commissie,

Drs. N.R.J. Deen

Dr. E. van Duuren

Dr. P.M. Renou (eindredactie)

