



Twintig jaar onderzoeks- evaluatie

Leonie van Drooge, Stefan de Jong¹,
Marieke Faber en Don Westerheijden²

Nederland heeft een lange traditie van kwaliteitszorg van het academische onderzoek. Deze Feiten en Cijfers geeft een overzicht van de ontwikkeling van het systeem van kwaliteitszorg en biedt inzicht in de resultaten van de onderzoeksevaluaties.

1. Inleiding

Wat heeft twintig jaar kwaliteitszorg van het academische onderzoek opgeleverd? Die vraag is aan de orde in deze Feiten en Cijfers. Het document geeft voor het eerst sinds 1993 zowel een overzicht van de ontwikkeling van het systeem als van de resultaten van de evaluaties. Het biedt inzicht in de manier waarop universiteiten en onderzoeksinstituten de kwaliteitszorg van het onderzoek uitvoeren.

Aan de orde komen een beschrijving van het huidige systeem en een korte schets van de ontwikkeling die tot dit systeem heeft geleid. Cijfers over aantallen evaluaties en toegekende oordelen bieden inzicht in de resultaten; een samenvatting van studies naar gebruikers geeft weer hoe de protocollen in de praktijk worden ervaren. Een karakterisering van een aantal buitenlandse systemen dient om het Nederlandse systeem in een internationaal perspectief te plaatsen.

Inhoud

Inleiding	1
Het huidige systeem van institutionele onderzoeksevaluatie: SEP 2009-2015	2
Voorgeschiedenis en wettelijk kader	3
VSNU-protocollen 1993, 1994 en 1998	4
Standaard Evaluatie Protocollen 2003-2009 en 2009-2015	4
Twintig jaar evaluatie in Nederland: grote overeenkomsten, kleine verschillen	5
Twintig jaar evaluatie in Nederland: verzameling en ontsluiting van gegevens	6
Aantallen evaluaties	7
Scores	9
Buitenland	13
Evaluatie in de praktijk: ervaringen, gebruik en follow-up	16
Samenvatting	17
Geraadpleegde bronnen en literatuur	18

Het Rathenau Instituut stimuleert de publieke en politieke meningsvorming over wetenschap en technologie. Daartoe doet het instituut onderzoek naar de organisatie en ontwikkeling van het wetenschapssysteem, publiceert het over maatschappelijke effecten van nieuwe technologieën, en organiseert het debatten over vraagstukken en dilemma's op het gebied van wetenschap en technologie.

1 Leonie van Drooge en Stefan de Jong zijn werkzaam als onderzoeker bij het Rathenau Instituut. Deze publicatie is tot stand gekomen met medewerking van Jasper Deuten, Catherine Chiong Meza en Barend van der Meulen van het Rathenau Instituut.

2 Marieke Faber en Don Westerheijden zijn werkzaam als onderzoeker bij CHEPS (Center for Higher Education Policy Studies), Universiteit Twente.

2 Twintig jaar onderzoeksevaluatie

De belangrijkste conclusies zijn:

- Nederland heeft een lang bestaand en stabiel systeem van kwaliteitszorg van academisch onderzoek, ook in vergelijking met andere landen. Het ontbrak echter tot heden aan een overzicht van de uitgevoerde evaluaties en bestuursstandpunten.
- Er is een grote variatie in reikwijdte van de evaluaties, variërend van hele disciplines, via hele disciplines minus een enkele instelling, of een combinatie van disciplines binnen een instelling, tot een enkel centrum of een enkele leerstoelgroep. Het is hierdoor niet mogelijk de kwaliteit van het onderzoek van deze evaluaties op systematische wijze te vergelijken.
- De score voor de kwaliteit van het onderzoek is in de afgelopen twintig jaar toegenomen. Momenteel scoort vrijwel al het onderzoek op alle aspecten als ten minste internationaal concurrerend. Hierdoor zijn er nauwelijks verschillen waarneembaar tussen de scores van onderzoekseenheden.
- Het Nederlandse systeem verschilt aanzienlijk van andere nationale systemen. Het ontbreekt in Nederland aan een nationale doelstelling, aan op voorhand bepaalde consequenties en aan een centrale organisatie die verantwoordelijk is voor het systeem. De doelstellingen zijn op instellingsniveau gedefinieerd en het zijn de instellingen zelf die de verantwoordelijkheid dragen voor de evaluaties en de consequenties bepalen.

2. Het huidige systeem van institutionele onderzoeksevaluatie: SEP 2009-2015

Het Standard Evaluation Protocol (SEP) 2009-2015³ beschrijft het huidige systeem voor onderzoeksevaluatie. Het gaat hierbij om institutionele evaluatie, dat wil zeggen om evaluatie van onderzoekseenheden aan universiteiten, inclusief de universitaire medische centra, en aan de instituten van de Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO) en de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW).

Er is een vast protocol voor evaluatie:

- Al het onderzoek dient in een periode van zes jaar eenmaal extern te worden geëvalueerd.
- Een interne mid-term evaluatie na drie jaar heeft als doel de follow-up van de externe evaluatie te monitoren.
- Evaluaties vinden plaats op twee niveaus: dat van de individuele onderzoekseenheid (groep, programma) en dat van het overkoepelende onderzoeksinstituut als geheel.
 - De criteria zijn:
 - wetenschappelijke kwaliteit;
 - wetenschappelijke productiviteit;
 - maatschappelijke relevantie;
 - levensvatbaarheid.
 - De doelstellingen zijn:
 - de kwaliteit van het onderzoek te verbeteren;
 - de besteding van de publieke middelen te verantwoorden aan het eigen bestuur, financiers, de overheid en de maatschappij.

Het bestuur van de instelling is opdrachtgever.

- Het bestuur van de instelling bepaalt welk instituut wanneer wordt beoordeeld;
- het stelt de *terms of reference* (TOR) voor iedere evaluatie vast;
- het stelt de *peer review committee* (PRC, verder te noemen evaluatiecommissie) samen en benoemt de leden.

De instelling presenteert het onderzoek in een zelfevaluatiedocument.

- De geëvalueerde eenheid / het instituut beschrijft de missie, doelen en context van het onderzoek in een zelfevaluatie;

3 VSNU, NWO, KNAW (2009). *Standard Evaluation Protocol 2009-2015*. Amsterdam: KNAW.

- deze bevat ook een beschrijving van de wetenschappelijke kwaliteit en relevantie, de maatschappelijke relevantie en het toekomstperspectief;
- verder een strategie voor de toekomst gebaseerd op een sterkte-zwakteanalyse, de zogenaamde SWOT (Strengths, Weaknesses, Opportunities, Threats);
- alsmede een overzicht van inputgegevens (middelen, personeelssamenstelling) en relevante outputgegevens.

De evaluatiecommissie formuleert een oordeel.

- De basis van het oordeel van de evaluatiecommissie zijn het zelfevaluatierapport en een ontmoeting met vertegenwoordigers van het instituut en de eenheden, veelal via een *site visit*;
- de evaluatiecommissie beschrijft het oordeel in een evaluatierapport;
- het oordeel over de onderzoekseenheid behelst elk van de vier criteria en betreft zowel prestaties van de afgelopen periode als plannen voor de toekomst;
- dit oordeel dient te bestaan uit een (kwalitatieve) samenvatting van belangrijkste bevindingen, alsmede uit een score waarvoor een vijfpuntsschaal is ontwikkeld;
- het oordeel over het instituut als geheel behelst beleid en management, is gericht op de toekomst en dient minimaal te bestaan uit een kwalitatieve toelichting.

Het bestuur van de instelling rondt de evaluatie af.

- Het instellingsbestuur ontvangt het evaluatierapport;
- na overleg met het instituut bepaalt het bestuur een standpunt over de oordelen en adviezen van de evaluatiecommissie;
- het evaluatierapport en het standpunt van het bestuur dienen openbaar gemaakt te worden.

Het SEP 2009-2015 beschrijft ook de meta-evaluatie van het protocol, dus de wijze waarop de deelnemende organisaties verantwoording afleggen over het juiste gebruik van het protocol.

- Ieder instellingsbestuur maakt een overzicht bekend van aanstaande evaluaties.
- Iedere instelling biedt in het jaarverslag een overzicht van evaluaties die zijn afgerond en van de standpunten van het bestuur.
- KNAW, NWO en de Vereniging van Universiteiten (VSNU) zorgen er gezamenlijk voor dat in 2013 een onafhankelijke expertcommissie het SEP 2009-2015 evalueert; dit betreft niet alleen het gebruik van het protocol, maar ook de gevolgen van de evaluaties op het beleid van de instellingen. De resultaten worden openbaar gemaakt.

Het SEP 2009-2015 staat niet op zichzelf. Nederland kent een traditie van dertig jaar institutionele onderzoeksevaluatie.

3. Voorgeschiedenis en wettelijk kader

Tot 1982 ligt de verantwoordelijkheid voor de kwaliteitszorg van het universitaire onderzoek bij de faculteiten. Facultaire wetenschapscommissies beoordelen het onderzoek op wetenschappelijke merites, haalbaarheid en samenstelling. De kritiek op het systeem is dat er geen inzicht is in de onderzoeksinspanningen en dat er geen verantwoording wordt afgelegd aan de financier, het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW)⁴. Dit verandert in 1982 met de introductie van de Voorwaardelijke Financiering (VF), een procedure waarbij beoordelingscommissies met externe experts onderzoeksprogramma's beoordelen.

In 1985 introduceert de minister van OCW een nieuwe bestuursfilosofie voor universiteiten⁵. Deze is gebaseerd op autonomie en zelfregulering: in ruil voor grotere autonomie moeten universiteiten kunnen laten zien dat ze kwaliteit leveren. In 1992 wordt dit wettelijk verankerd en

4 Tot 2003 was de naam Ministerie van Onderwijs, Cultuur en Wetenschappen, afgekort als OC&W. De actuele naam OCW (2013) wordt in deze publicatie gehanteerd.

5 Tweede Kamer der Staten-Generaal (vergaderjaar 1985-1986). *Beleidsnota Hoger Onderwijs Autonomie en Kwaliteit*, dossier 19 253, nrs. 1-2.

4 Twintig jaar onderzoeksevaluatie

krijgen universiteiten de verplichting zorg te dragen voor de totstandkoming van een regelmatige beoordeling van de kwaliteit van de werkzaamheden van de instelling. Voor de tekst in de Wet Hoger Onderwijs en Wetenschappelijk Onderzoek, zie Box 1.

Box 1: Wet Hoger Onderwijs en Wetenschappelijk Onderzoek

Het instellingsbestuur (...) draagt er zorg voor dat, zo veel mogelijk in samenwerking met andere instellingen, wordt voorzien in een regelmatige beoordeling, mede door onafhankelijke deskundigen, van de kwaliteit van de werkzaamheden van de instelling. (...) Voor zover die beoordeling mede geschiedt door onafhankelijke deskundigen, zijn de uitkomsten daarvan openbaar. (Wet Hoger Onderwijs en Wetenschappelijk Onderzoek, artikel 1.18).

Onze minister kan aan de bekostiging van onderzoek aan universiteiten voorwaarden verbinden, verband houdend met de kwaliteitszorg. (Wet Hoger Onderwijs en Wetenschappelijk Onderzoek, artikel 2.5, lid 2).

4. VSNU-protocollen 1993, 1994 en 1998⁶

Als gevolg van de nieuwe wetgeving ontwikkelt de VSNU begin jaren negentig een landelijk stelsel voor onderzoeksevaluatie. De VSNU doet dit in overleg met NWO en KNAW. De kern van het stelsel is periodieke, disciplinegebonden beoordeling van al het onderzoek aan de universiteiten door internationaal samengestelde beoordelingscommissies.

In februari 1993 wordt het eerste algemene protocol voor de kwaliteitszorg van academisch onderzoek vastgesteld. In een proefronde vinden evaluaties plaats van onderzoek in de werktuigbouwkunde, biologie, psychologie en historische wetenschappen. Naar aanleiding van de ervaringen met deze evaluaties stelt de VSNU het protocol op enkele onderdelen bij. In de vier volgende jaren wordt al het overige universitaire onderzoek beoordeeld met het VSNU-protocol 1994. Na een evaluatie van het protocol besluit de VSNU een nieuwe evaluatieronde te organiseren. In het VSNU-protocol 1998 blijven de doelen en criteria in grote lijnen onveranderd.

5. Standaard Evaluatie Protocollen 2003-2009⁷ en 2009-2015

In 1999 richten VSNU, NWO en KNAW de werkgroep Kwaliteitszorg Wetenschappelijk Onderzoek op. Deze werkgroep publiceert in 2000 een rapport⁸ over de contouren van een nieuw landelijk stelsel van kwaliteitszorg. Dit rapport vormt de basis voor het Standard Evaluation Protocol (SEP).

Het SEP 2003-2009 kent enkele belangrijke veranderingen ten opzichte van de VSNU-protocollen. Niet langer is de VSNU verantwoordelijk voor het organiseren van de evaluatie; de onderzoeksinstellingen zijn dat zelf. Daarnaast wordt het systeem van vergelijkende beoordelingen op nationaal niveau losgelaten. Dit lost een belangrijk knelpunt op. Te vaak gaat het in disciplinebrede beoordelingen om een vergelijking van 'steenkolens met eieren', waardoor de waarde van het oordeel voor het veld ontoereikend is⁹.

De Meta Evaluatie Commissie Kwaliteitszorg Wetenschappelijk Onderzoek (MEC) evalueert in 2007¹⁰ en 2009¹¹ het SEP 2003-2009. Hoewel overwegend positief, zijn er twee belangrijke punten van kritiek. Het is onduidelijk wat er binnen de instellingen gebeurt met de resultaten van de evaluaties en de scores zijn aan erosie onderhevig. Het SEP 2009-2015 is in grote lijnen een voortzetting van het SEP 2003-2009. Er zijn kleine wijzigingen, onder andere naar aanleiding van de kritiekpunten van de MEC.

6 VSNU (1993). *Quality Assessment of Research – protocol 1993*. Utrecht: VSNU; VSNU (1994). *Quality Assessment of Research – protocol 1994*. Utrecht: VSNU; VSNU (1998). *Protocol 1998*. In: *Series Assessment of Research Quality*. Utrecht: VSNU.

7 VSNU, NWO, KNAW (2003). *Standard Evaluation Protocol 2003-2009*. Utrecht: VSNU.

8 Werkgroep Kwaliteitszorg Wetenschappelijk Onderzoek (2001). *Kwaliteit verplicht. Naar een nieuw stelsel van kwaliteitszorg voor het wetenschappelijk onderzoek*. Amsterdam: KNAW.

9 Zie noot 8, p. 41.

10 Meta Evaluatie Commissie Kwaliteitszorg Wetenschappelijk Onderzoek (2007). *Trust but Verify*. Amsterdam: KNAW.

11 Meta Evaluatie Commissie Kwaliteitszorg Wetenschappelijk Onderzoek (2009). *E-VA-LU-E-REN. Het beoordelen van wetenschappelijk onderzoek in de praktijk*. Amsterdam: KNAW.

6. Twintig jaar evaluatie in Nederland: grote overeenkomsten, kleine verschillen

De protocollen die in Nederland in de afgelopen twintig jaar zijn gebruikt, verschillen allemaal van elkaar. Toch zijn de overeenkomsten groot en is er sprake van continuïteit. Zie Box 2 voor een overzicht van de overeenkomsten in de evaluatieprotocollen.

Box 2: overeenkomsten in VSNU- en SEP-evaluatieprotocollen

De VSNU- en SEP-evaluatieprotocollen kennen de volgende overeenkomsten:

- Er is een nationaal systeem voor de periodieke beoordeling van al het academisch onderzoek.
- De doelen zijn kwaliteitsverbetering en verantwoording afleggen.
- Beoordeling vindt plaats op het niveau van de onderzoekseenheid.
- De vier beoordelingscriteria zijn:
 - wetenschappelijke kwaliteit;
 - wetenschappelijke productiviteit;
 - relevantie;
 - levensvatbaarheid.
- Wetenschappelijke peers vormen een oordeel, mede op basis van informatie die de eenheid aandraagt.
- Het bestuur van de instelling is de ontvanger van het rapport en dient een reactie te geven.

Als de protocollen nader worden bezien, valt vooral een aantal verschillen op tussen de VSNU-protocollen enerzijds en de SEP-protocollen anderzijds. De belangrijkste verschillen staan in Box 3 vermeld.

Box 3: verschil SEP-protocollen ten opzichte van VSNU-protocollen

De SEP-protocollen kennen enkele verschillen ten opzichte van de VSNU-protocollen:

- De disciplinebrede beoordeling is losgelaten; het is niet langer verplicht gezamenlijke evaluaties te organiseren met alle instellingen waar onderzoek plaatsvindt in de betreffende discipline.
- De (neven)doelstelling dat de overheid de evaluatie als verkenning van een discipline kan gebruiken is vervallen.
- Het protocol geldt ook voor evaluatie het onderzoek van de KNAW- en NWO-instituten¹².
- De verantwoordelijkheid voor het systeem ligt volledig bij de instellingen en daarmee heeft iedere instelling de vrijheid om eigen keuzes te maken aangaande:
 - de planning van de evaluaties;
 - de formulering van de evaluatie-opdracht;
 - de samenstelling en benoeming van de evaluatiecommissie;
 - de reactie op het rapport van de evaluatiecommissie en het verbinden van consequenties aan de evaluatie.
- De verplichting om het rapport aan het Ministerie van OCW te sturen is vervallen¹³.

¹² Ook andere instellingen maken gebruik van onderdelen van het SEP voor evaluatie van het onderzoek. In deze Feiten en Cijfers staan alleen gegevens vermeld over eenheden die binnen het dan geldende protocol vallen.

¹³ SEP 2003-2009 kent nog wel de verplichting om de resultaten aan het Ministerie van OCW te rapporteren; SEP 2009-2015 schrijft alleen voor de resultaten publiek te maken, bij voorkeur via de website van de instelling.

Tabel 1 Betekenis van het oordeel in de verschillende protocollen

Score	VSNU 1994 en 1998	SEP 2003-2009	SEP 2009-2015
5	Excellent (excellent)	Excellent international leader; most likely important and substantial impact	Excellent world leading; has important and substantial impact
4	Goed (good)	Very good internationally competitive and national leader; expected to make significant contribution	Very good internationally competitive and nationally leading; makes a significant contribution
3	Volgende/gemiddeld (satisfactory/average)	Good internationally visible and national player; will probably make valuable contribution	Good internationally visible and nationally competitive; makes a valuable contribution
2	Onvoldoende (unsatisfactory)	Satisfactory nationally visible; solid, not exciting, will add to understanding	Satisfactory nationally visible; solid, not exciting, adds to understanding
1	Slecht (poor)	Unsatisfactory neither solid, nor exciting; not worth pursuing	Unsatisfactory neither solid, nor exciting

Bron: VSNU (1994), VSNU (1998), VSNU, NWO, KNAW (2003), VSNU, NWO, KNAW (2009)

Rathenau Instituut

De betekenis van de oordelen op de vijfpuntsschaal is een aantal keren gewijzigd (Tabel 1). Waren in de VSNU-protocollen 1994 en 1998 de oordelen 1 en 2 onvoldoende, in het SEP is alleen 1 nog onvoldoende. Nieuw in het SEP is het oordeel *very good*, tussen *good* en *excellent* in. En daar waar in de VSNU-protocollen voor de toelichting een enkel woord volstaat, worden in de SEP-protocollen de toelichtingen steeds nauwkeuriger omschreven. In SEP 2003-2009 staan termen van verwachting (*most likely, will probably*); in SEP 2009-2015 staan scherper geformuleerde constatering (*has, makes*). Deze wijzigingen zijn een reactie op de kritiek van de MEC dat er een erosie van scores optreedt.

7. Twintig jaar evaluatie in Nederland: verzameling en ontsluiting van gegevens

Met de ontwikkeling van de PER Base¹⁴ heeft het Center for Higher Education Policy Studies (CHEPS) de resultaten van twintig jaar onderzoeksevaluatie ontsloten. PER Base is een database met gegevens van alle bekende evaluaties vanaf 1993¹⁵. Het betreft institutionele onderzoeksevaluatie, volgens de VSNU- en SEP-protocollen. Een overzicht van rapporten is digitaal beschikbaar op www.rathenau.nl.

Het bleek niet eenvoudig om een goed overzicht te krijgen van alle onderzoeksevaluaties van de afgelopen twintig jaar. Er is een aantal onzekerheden. Zo is het niet duidelijk welke evaluaties er zijn geweest. In evaluatierapporten die gepubliceerd zijn tot en met 2000 staat een overzicht van de reeds gepubliceerde evaluaties. De rapporten die in deze overzichten staan vermeld, zijn opgenomen in de PER Base. Dit betekent dat de database compleet is wat betreft de evaluaties met de protocollen VSNU 1993 en VSNU 1994. Voor het VSNU-protocol 1998 is dit niet met zekerheid vast te stellen.

Het ontbreekt sinds de introductie van het SEP aan een centraal overzicht van uitgevoerde en geplande evaluaties. Sommige instellingen negeren de verplichting om aanstaande evaluaties bekend en resultaten openbaar te maken. Ook is er geen voorgeschreven ontvanger meer van de rapporten.

Daarnaast is het niet duidelijk welk deel van het onderzoek is geëvalueerd. Niet al het onderzoek is eenmaal per protocol geëvalueerd. Zo zijn natuurkunde, sterrenkunde en een deel van de landbouwwetenschappen niet geëvalueerd met het VSNU-protocol 1998. In hoeverre alle disciplines en alle eenheden met de SEP-protocollen zijn geëvalueerd, is door de diversiteit aan evaluaties niet na te gaan.

¹⁴ PER Base is ontwikkeld door Center for Higher Education Policy Studies (CHEPS, Universiteit Twente). Voor de ontwikkeling van PER Base heeft CHEPS financiering ontvangen van het Ministerie van OCW in het kader van het project CHERPA. CHEPS is verantwoordelijk voor de gegevens over evaluaties tot en met 2009; het Rathenau Instituut over de gegevens van evaluaties vanaf 2009.

¹⁵ Het betreft alleen gegevens van instellingen die binnen de definitie van het betreffende protocol vallen (VSNU-protocollen: universiteiten; SEP: universiteiten, KNAW- en NWO-instituten).

Bovendien ontbreekt het aan inzicht in de omvang van het geëvalueerde onderzoek in fte. In sommige rapporten staan hierover geen gegevens vermeld, in andere staat de totale omvang (som over de jaren) vermeld, een omvang per jaar per categorie personeel, of een omvang in een bepaald jaar. Het is hierdoor niet vast te stellen of al het onderzoek dat binnen een geëvalueerde eenheid valt, ook daadwerkelijk is beoordeeld.

Ten slotte varieert de reikwijdte van de evaluatierapporten in grote mate, vooral na invoering van het SEP 2003-2009. Dit belemmert het zicht op de eenheden die zijn geëvalueerd. Zo is het vanaf het SEP niet meer verplicht als discipline gezamenlijk op te trekken. Er vinden nog steeds evaluaties plaats met alle instellingen gezamenlijk; in sommige disciplines organiseert vrijwel iedere universiteit een eigenstandige evaluatie, zoals bij natuurkunde (zeven rapporten over universitaire groepen en drie over NWO-instituten in de periode 2003-2009); en er komen mengvormen voor, waarbij een aantal universiteiten gezamenlijk optrekt en een of enkele een eigenstandige evaluatie organiseren, zoals filosofie (2003-2009), psychologie (2009-2015) en rechtsgeleerdheid (2003-2009).

Ook de disciplinaire samenstelling van het geëvalueerde verandert met de introductie van het SEP. Er vinden evaluaties plaats van meerdere disciplines tegelijk, zoals de evaluatie van *science and technology* (Rijksuniversiteit Groningen, 2005) of van sociale wetenschappen (Vrije Universiteit, 2008). Maar vooral vinden er veel evaluaties plaats van subdisciplines, zoals de evaluaties van een centrum voor gender en diversiteit (Universiteit Maastricht, 2005), een leerstoel meteorologie en luchtkwaliteit (Wageningen UR, 2004) of een instituut voor sociaal-cultureel onderzoek (Radboud Universiteit Nijmegen, 2006).

Door de vele veranderingen in de afgelopen twintig jaar wat betreft naamgeving van groepen, omvang en samenstelling, is het niet mogelijk om de onderzoekseenheid als uitgangspunt te nemen. Slechts in een enkel geval is een eenheid over de jaren te volgen. Zo komt de groep Information Processing and Task Performance (Rijksuniversiteit Groningen, psychologie) in alle evaluaties van de afgelopen twintig jaar voor. Dit is echter een uitzondering.

8. Aantallen evaluaties

Twintig jaar institutionele onderzoeksevaluatie levert 222 evaluatierapporten op met oordelen over 4765 eenheden¹⁶. Figuur 1 toont de aantallen evaluatierapporten per jaar; Figuur 2 toont de aantallen geëvalueerde eenheden (onderzoeksgroepen, programma's, afdelingen) per jaar.

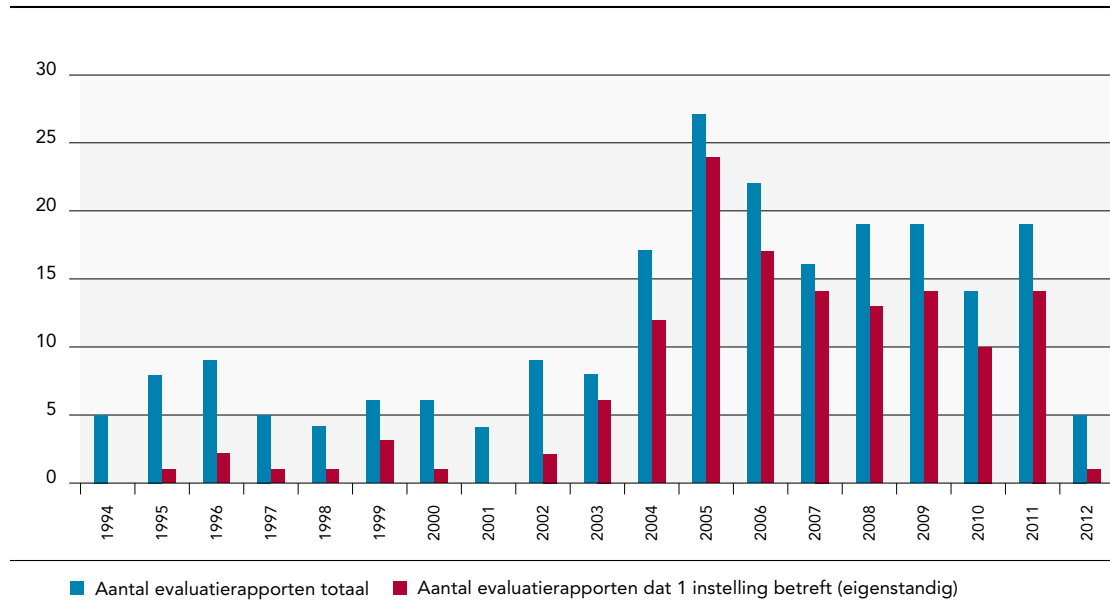
Met de introductie van het SEP in 2003 neemt het aantal evaluatierapporten per jaar fors toe. Deze stijging is voor een groot deel toe te schrijven aan de toename van het aantal eigenstandige evaluaties. Een eigenstandige evaluatie betreft één instelling (universiteit of instituut), en kan een of meerdere eenheden binnen die instelling betreffen. Deze stijging van eigenstandige evaluaties is conform de wijziging in het protocol: de disciplinebrede beoordeling is losgelaten; de besturen van de instellingen worden verantwoordelijk en de KNAW- en NWO-instituten nemen ook deel aan de evaluaties.

Figuur 1 toont, naast het totaal aantal evaluatierapporten, ook het aantal eigenstandige evaluatierapporten per jaar. In totaal zijn dat er 136 van 222 voor de periode 1994-2012. Te zien is dat er voor 2003 ook al enige eigenstandige evaluaties plaatsvonden. Sommige disciplines komen namelijk slechts aan een enkele instelling voor, zoals luchtvaarttechniek of mariene technologie.

Gemiddeld worden in een evaluatierapport 21 eenheden beoordeeld. Er zijn 25 rapporten bekend waarin één eenheid is beoordeeld. Verreweg het omvangrijkste rapport is dat van geneeskunde uit 1994, met oordelen over 572 eenheden. Daarna volgt een rapport over chemie uit 1996, met oordelen over 162 eenheden.

¹⁶ Alleen die eenheden zijn opgenomen in de database, die op het moment van evaluatie deel uitmaakten van het protocol. Dit betekent dat gegevens over TNO-instituten die met het VSNU-protocol zijn geëvalueerd, of het BSIK-programma NGI dat met het SEP is geëvalueerd, buiten beschouwing zijn gelaten.

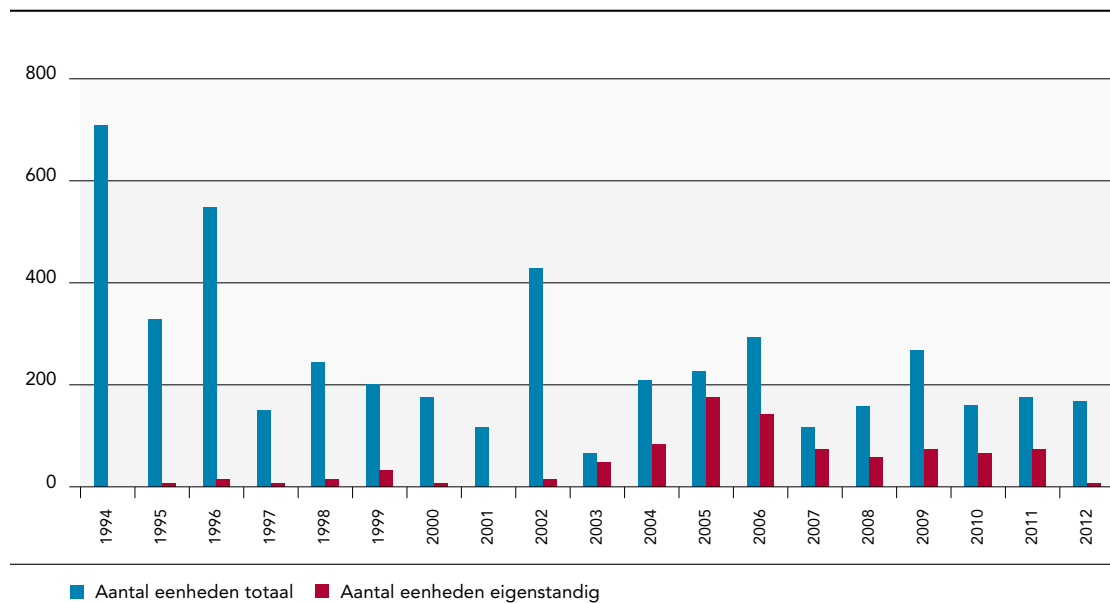
Figuur 1 Aantal evaluatierapporten per jaar



Bron: PER Base (CHEPS en Rathenau Instituut)

Rathenau Instituut

Figuur 2 Aantal geëvalueerde eenheden per jaar



Bron: PER Base (CHEPS en Rathenau Instituut)

Rathenau Instituut

9. Scores

De PER Base bevat de scores voor de vier criteria voor elke eenheid op de vijfpuntsschaal.

De analyse hieronder betreft deze scores. De evaluatiecommissies beschrijven ook kwalitatieve oordelen in de rapporten. Die oordelen, over 4765 eenheden maal vier criteria, zijn in deze analyse buiten beschouwing gelaten. In de analyse richten we ons op de scores op de vijfpuntsschaal. In de praktijk wordt juist ook aan deze scores een groot belang gehecht.

Een aandachtspunt bij de interpretatie van de scores is de toekenning ervan in de praktijk. De scores zijn volgens de protocollen een code voor een oordeel. In de praktijk blijkt een aanzienlijke hoeveelheid scores voor te komen anders dan een geheel getal tussen 1 en 5. Het gebruik van gebroken cijfers tussen 1 en 5 (zoals 4,5 of 4,13) suggereert dat de evaluatiecommissie de score niet interpreteert als code, maar als glijdende schaal. In de analyse van de scores is uitgegaan van deze praktijk.

Een aantal opmerkingen over de analyse van de scores. De eerste vier evaluaties - werktuigbouwkunde, biologie, psychologie en historische wetenschappen - zijn buiten beschouwing gelaten bij de analyse van de scores. Het VSNU-protocol 1993, dat bij deze evaluaties is gebruikt, hanteert namelijk een driepuntsschaal.

De analyse betreft alleen scores die binnen de systematiek van het dan geldende protocol te passen zijn¹⁷. Dit betekent dat het rapport geneeskunde 1994 weliswaar is opgenomen in de database, maar dat er geen analyse is gemaakt van de scores. In deze evaluatie zijn er slechts twee scores per groep gegeven en deze scores zijn niet eenduidig te vertalen naar de voorgeschreven criteria.

De score 'niet van toepassing' of 'niet beoordeelbaar' is buiten beschouwing gelaten in de analyse. Dit betreft 13 procent van de gevallen, variërend per criterium van 10 procent (wetenschappelijke kwaliteit) tot 16 procent (wetenschappelijke productiviteit). De implicatie is dat er voor ieder criterium een ander aantal scores staat geregistreerd.

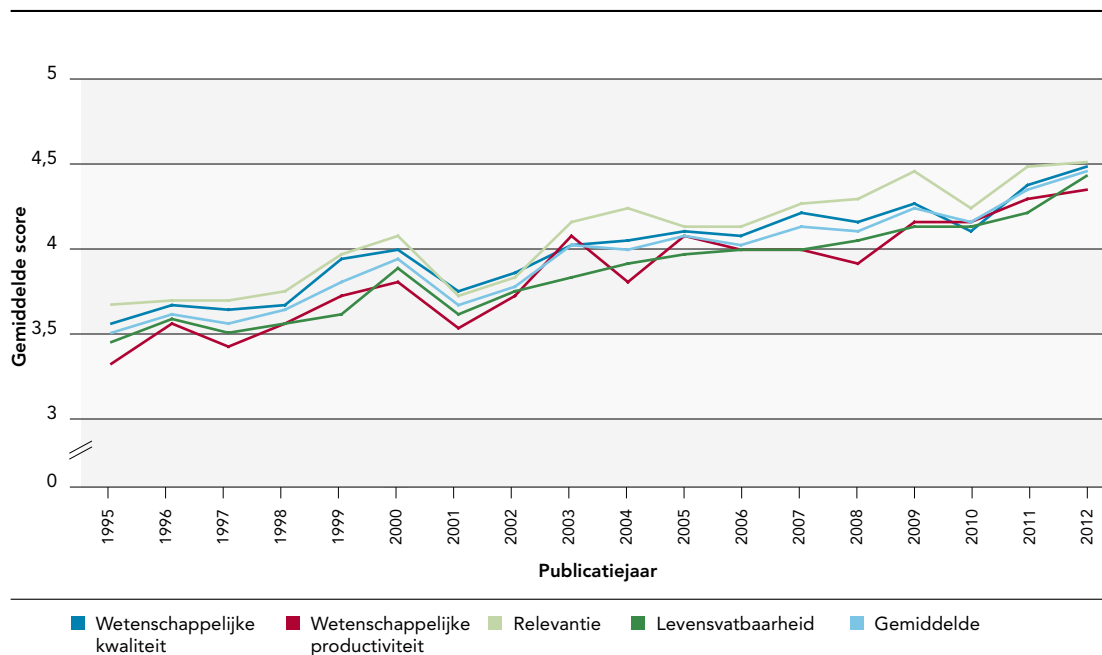
De wijdverspreide indruk dat de scores toenemen, wordt bevestigd door de gegevens in Figuur 3 en Tabel 2. Hierin staan de gemiddelde scores voor elk van de criteria per jaar en het gemiddelde voor de vier criteria per jaar (Figuur 3) en per protocol (Tabel 2). De score stijgt op elk van de vier criteria. Uit nadere analyse blijkt dat de spreiding van de scores klein is en afneemt met de opeenvolgende protocollen. Na de introductie van SEP 2009-2015 is de meest voorkomende score over alle criteria een 5, oftewel *world leading*. Bij de beschrijving van de protocollen is reeds vermeld dat de betekenis van de scores enkele malen is gewijzigd (Tabel 1).

Als gevolg hiervan zou bij gelijkblijvende kwaliteit de score dalen. Als we de omschrijving van de scores in tabel 1 serieus nemen, betekent de trend dat de stijging nog sterker is dan uit deze tabel en figuur blijkt.

De verwachting dat de stijging van de scores is toe te schrijven aan de toename van eigenstandige evaluaties (waarbij slechts één instelling is betrokken) wordt ontkracht in Tabel 3. De verschillen tussen beide soorten evaluaties zijn gering. De indruk bestond dat eigenstandige evaluaties veel meer op maat zouden zijn en dus tot hogere scores zouden leiden.

¹⁷ In enkele gevallen wordt in de rapporten volstaan met een woordelijke kwalificatie per onderdeel. Alleen wanneer deze kwalificatie de score systematiek volgt, en dus eenduidig te vertalen is naar een score, is dit gedaan. Een voorbeeld is het rapport van de evaluatie van het LUMC in 2006.

Figuur 3 Gemiddelde score per criterium per jaar



Bron: PER Base (CHEPS en Rathenau Instituut)

Rathenau Instituut

Tabel 2 Gemiddelde scores per eenheid, per criterium en per protocol¹⁸

	Wetenschappelijke kwaliteit	Wetenschappelijke productiviteit	Relevantie	Levensvatbaarheid
VSNU 1994	3,65 (n=1179)	3,47 (n=988)	3,70 (n=1066)	3,53 (n=1014)
VSNU 1998	3,88 (n=1009)	3,70 (n=852)	3,93 (n=1009)	3,72 (n=970)
SEP 2003-2009	4,14 (n=1205)	4,03 (n=1201)	4,23 (n=1217)	4,03 (n=1166)
SEP 2009-2015	4,39 (n=385)	4,31 (n=384)	4,48 (n=387)	4,28 (n=379)

Bron: PER Base (CHEPS en Rathenau Instituut)

Rathenau Instituut

Tabel 3 Gemiddelde score per eenheid per protocol uitgesplitst naar eigenstandige of gezamenlijke evaluatie¹⁷

	Eigenstandige evaluatie	Gezamenlijke evaluatie
VSNU 1994	3,46 (n=49)	3,58 (n=1138)
VSNU 1998	3,80 (n=39)	3,80 (n=975)
SEP 2003-2009	4,16 (n=565)	4,05 (n=663)
SEP 2009-2015	4,37 (n=132)	4,37 (n=256)

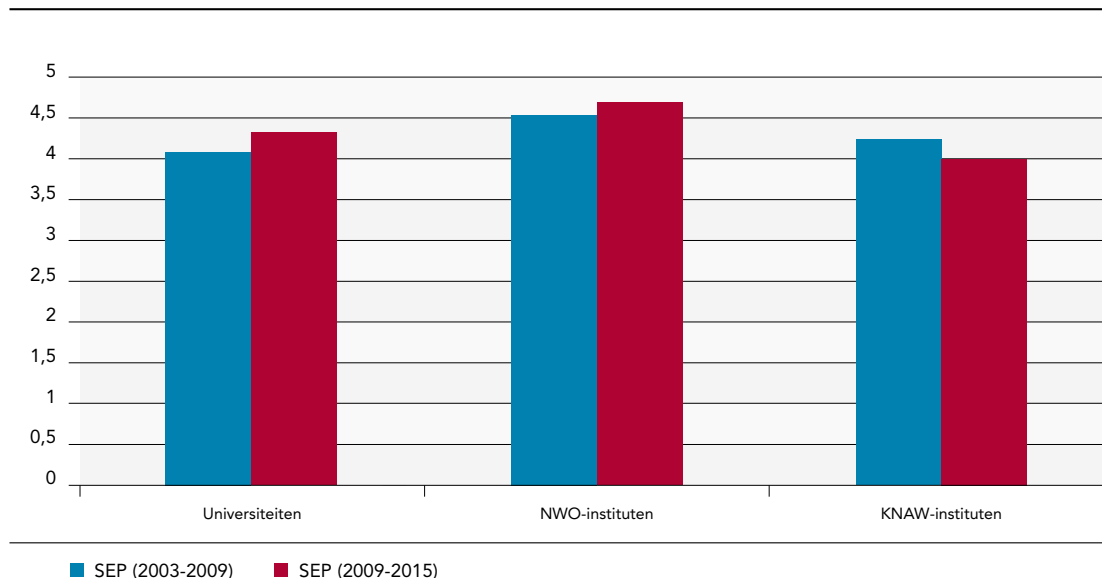
Bron: PER Base (CHEPS en Rathenau Instituut)

Rathenau Instituut

18 Evaluaties waarvan het niet duidelijk is welk protocol is gebruikt zijn buiten beschouwing gelaten.

Sinds de invoering van het SEP vallen ook de NWO- en KNAW-instituten binnen de protocollen. Deze onderzoeksinstituten worden in vrijwel alle gevallen eigenstandig geëvalueerd. Figuur 4 laat zien wat de gemiddelde scores zijn voor de KNAW- en de NWO-instituten in vergelijking met de universiteiten. De NWO-instituten, die alle met beide protocollen zijn geëvalueerd, scoren zeer hoog. De gemiddelde score voor de KNAW-instituten lijkt af te nemen. Hierbij moet worden aangetekend dat pas een klein deel van de KNAW-instituten, en dan vooral de kleinere instituten, is beoordeeld met SEP 2009-2015.

Figuur 4 Gemiddelde score voor de universiteiten, NWO- en KNAW-instituten



Bron: PER Base (CHEPS en Rathenau Instituut)

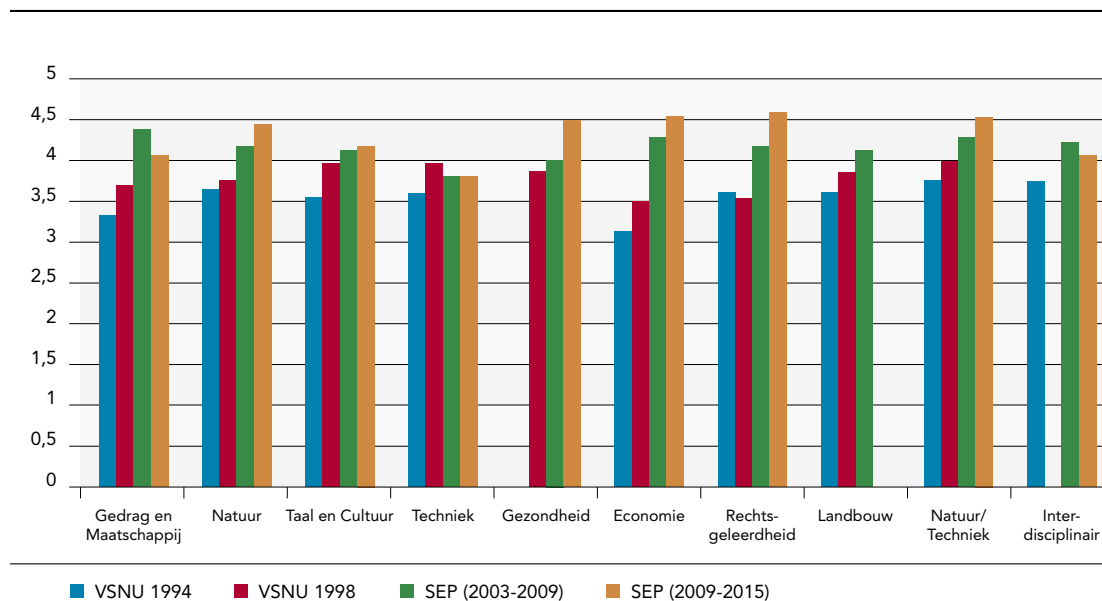
Rathenau Instituut

Figuur 5 biedt een overzicht van de gemiddelde score per HOOP-gebied¹⁹; figuur 6 toont het aantal eenheden per HOOP-gebied die voor tenminste één criterium een score toegekend hebben gekregen. In alle HOOP-gebieden is de gemiddelde score gestegen sinds de eerste evaluaties; in sommige gebieden zelfs met meer dan 1. Het enige gebied waar de stijging kleiner is, is Techniek. In sommige gebieden is er in een bepaalde periode een lichte daling waarneembaar. Zo is in HOOP-gebied Gedrag en Maatschappij, waar vrijwel alle eenheden zijn geëvalueerd met SEP 2009-2015, de gemiddelde score gedaald ten opzichte van de evaluaties met het vorige protocol. Ook de score voor de interdisciplinaire evaluaties daalt licht, maar hierbij moet worden aangetekend dat het slechts om een beperkt aantal eenheden gaat. In het algemeen kan gesteld worden dat de stijging van scores niet is toe te schrijven aan een bepaald gebied.²⁰

¹⁹ Evaluaties waarvan het niet duidelijk is welk protocol is gebruikt zijn buiten beschouwing gelaten.

²⁰ HOOP gebieden zijn wetenschapsgebieden zoals onderscheiden in het Hoger Onderwijs en Onderzoek Plan van OCW. Voor HOOP-gebieden per rapport: zie het digitale overzicht. (www.rathenau.nl) Onder HOOP-gebied Natuur/Techniek vallen evaluatierapporten van disciplines die zowel onder Natuur als Techniek vallen: natuurkunde, sterrenkunde, scheikunde, wiskunde, informatica.

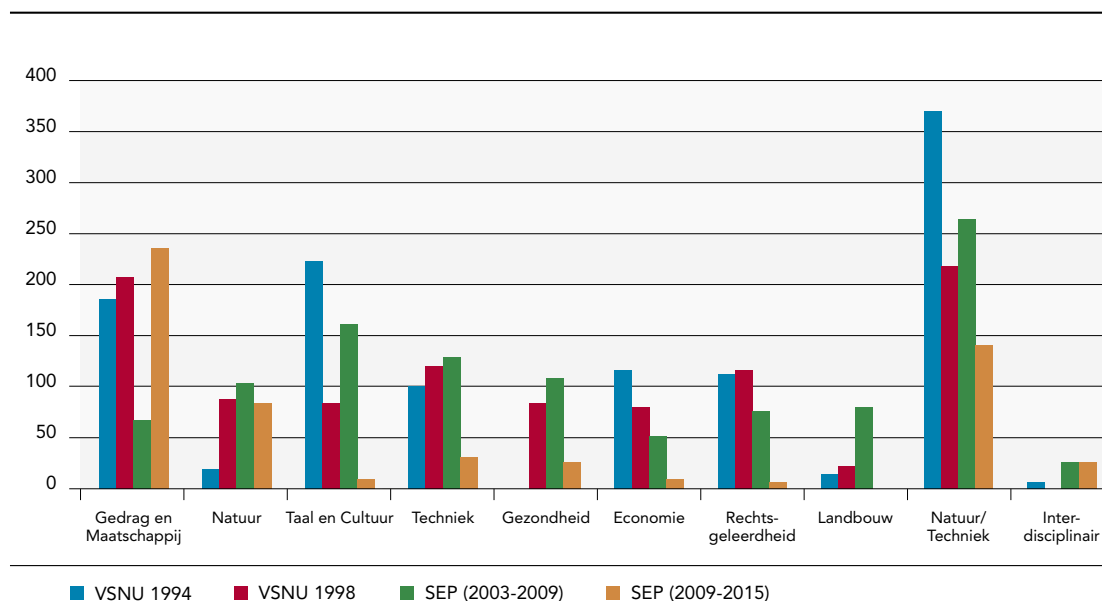
Figuur 5 Gemiddelde score per HOOP-gebied per protocol



Bron: PER Base (CHEPS en Rathenau Instituut)

Rathenau Instituut

Figuur 6 Aantal geëvalueerde eenheden per HOOP-gebied en per protocol



Bron: PER Base (CHEPS en Rathenau Instituut)

Rathenau Instituut

10. Buitenland

Om de ervaring in Nederland in een breder perspectief te plaatsen, is een vergelijking gemaakt met de evaluatiesystemen in een aantal andere landen met een hoogwaardig wetenschapssysteem. Tabel 4 toont een overzicht van nationale systemen voor institutionele evaluatie. Het gaat hierbij dus niet om evaluatie van financieringsinstrumenten of programma's, maar om evaluatie van onderzoekseenheden.

Uit deze vergelijking komt naar voren dat er elders andere keuzen worden gemaakt. Zo kennen Duitsland en de Verenigde Staten geen nationaal systeem. Verschillende Duitse deelstaten hebben weliswaar een eigen systeem van institutionele evaluatie, maar die zijn weinig stabiel. Op federaal niveau bestaat een dergelijk systeem niet. In de Verenigde Staten is onlangs *STAR metrics* geïntroduceerd, ter verantwoording van door de overheid gefinancierd onderzoek. De doelstelling is echter afwijkend. Doel is in eerste instantie de impact van het financieringsinstrument op de werkgelegenheid vast te stellen.

Landen die wel een nationaal systeem kennen, maken andere keuzen dan Nederland²¹. Zo is er in die landen een organisatie (bijv. nationaal agentschap of ministerie) verantwoordelijk voor de evaluatie. Nederland is het enige land waar de instellingen zelf de volledige verantwoordelijkheid dragen.

Daarnaast zijn er in die landen doelen op nationaal niveau geformuleerd: de internationale positie versterken, de kwaliteit verhogen tot wereldniveau. Nederland is het enige land waar de doelstelling van kwaliteitsverbetering op het niveau van de onderzoekseenheid is gedefinieerd.

Box 4: Doelen van evaluatie

Er zijn verschillende doelen voor evaluatie: evaluatie ter verantwoording, ter beloning en ter verbetering. Bij elk van deze doelen horen geëigende vragen en uitkomsten.

Verantwoorden

Kern van deze evaluatie is de vraag of de middelen juist worden besteed en of aan de vastgestelde procedures is voldaan. Het is van belang dat duidelijk is wat de eisen zijn. Een belangrijk deel van de evidentie kan door anderen dan de direct betrokkenen worden verzameld: maken van overzichten van middelen en resultaten en beschrijven van procedures. Het oordeel komt in de kern neer op: voldoende/onvoldoende.

Rangorde aanbrenge en belonen

Kern van deze vorm van evaluatie is de vraag wie of wat de beste is en in aanmerking komt voor (meer) beloning. Het is van belang dat er overeenstemming is over de criteria en over de weging daarvan. Het oordeel kan een absolute rangorde opleveren – dus een overzicht van beste naar slechtste – of een relatieve rangorde, waarbij hetzelfde oordeel aan meerdere geëvalueerden gegeven kan worden.

Verbeteren

Kern van deze vorm is te bepalen of de performance zo goed mogelijk bijdraagt aan de beoogde missie of doelen, en mogelijkheden tot verbetering vaststellen. Het vraagt van de geëvalueerde eenheid het vermogen tot zelfreflectie. Het oordeel behelst een diagnose van de actuele stand van zaken en aanbevelingen voor de toekomst. Het oordeel is deels beschrijvend en deels voorschrijvend.

21 Het betreft Engeland, Australië, Spanje, Italië, Frankrijk, Noorwegen, Zweden en Denemarken.

Tabel 4 Internationale vergelijking van nationale systemen voor institutionele evaluatie

Land	Sinds	Verantwoordelijke organisatie	Methode evaluatie
Engeland ^{22,23}	1986	Higher Education Funding Council for England (HEFCE)	Peer score over voorgedragen prestaties van individuele onderzoekers (rangorde)
Spanje	1989	Agentschap ANEP	Beoordeling van output van individuele onderzoekers door expert panel
Nederland	1993	VSNU / organisaties	Peer score / oordeel van voorgedragen onderzoeksprestaties van groepen en instituten
Noorwegen	2005	Ministerie van Onderwijs en Onderzoek	Metten van geregistreerde onderzoeksoutput via prestatie-indicatoren (deels disciplinair vastgesteld)
Denemarken ²⁴	2006	Agentschap voor Wetenschap, Technologie en Innovatie	Nationale evaluaties van onderdelen van het wetenschapssysteem (financierings-instrumenten, disciplines, programma's, systeemonderdelen) door peers; mede gebaseerd op zelfevaluatie; binnen door minister vastgesteld kader
Frankrijk ²⁵	2006	Agentschap AERES	Zelfevaluatie en visitatie door peers
Australië	2008	Australian Research Council (ARC)	Disciplinaire nationale reviews op basis van indicatoren en peer review
Zweden	2008	Ministerie van Onderwijs en Onderzoek, ondersteund door Vetenskapsrådet (research council)	Prestatie-indicatoren: output en externe financiering
Italië	2009	Agentschap ANVUR	Peer score van prestatie-indicatoren
Duitsland		Geen nationaal systeem	
Verenigde Staten ²⁶		Geen nationaal systeem	

22 HEFCE (2010). *Guide to funding*. http://www.hefce.ac.uk/media/hefce1/pubs/hefce/2010/1024/10_24.pdf (geraadpleegd 21-11-2012).

23 Voor Engeland; voor Schotland, Wales en Noord-Ierland zijn het zusterorganisaties. De systematiek is dezelfde.

24 Danish Agency for Science, Technology and Innovation. *Research Evaluation Guidelines*.

<http://en.fi.dk/research/research-evaluation/framework-and-methods/action-plan-for-research-evaluation/Research%20Evaluation%20Guidelines.pdf> (geraadpleegd 21-11-2012).

25 LERU (2012) *Research universities and research assessment*. Leuven: LERU.

26 <https://www.starmetrics.nih.gov/> (geraadpleegd 26-11-2012).

Land	Nationale doelstelling van het systeem	Doelstelling en consequentie van evaluaties binnen het systeem
Engeland	Versterken van de internationale positie van het onderzoek door stimuleren van toponderzoek	Rangorde aanbrengen en belonen: prestatiefinanciering: verdelen van 70% van de eerste geldstroom (onderdeel <i>lumpsum</i>)
Spanje	Verhogen van de kwaliteit van onderzoek tot wereldniveau door het vergroten van de onderzoeksinspanning/-output	Rangorde aanbrengen en belonen: toewijzen (of weigeren) van 6-jarige onderzoeksinstellingen
Nederland	Geen	Kwaliteitsverbetering en verantwoording: de consequentie is niet omschreven; het is aan de instellingsbesturen zelf om dit te bepalen
Noorwegen	Verhogen van onderzoeksactiviteit en stimuleren van excellentie	Rangorde aanbrengen en belonen: prestatiefinanciering: herverdeling van een deel van de eerste geldstroom (minder dan 15%; onderdeel <i>lumpsum</i>)
Denemarken	Verantwoording van de nationale investeringen in onderzoek en verbetering van het systeem zodat de investeringen tot excellentie leiden	Rangorde aanbrengen en belonen: prestatiefinanciering: herverdeling van een toenemend deel van de financiering (onbekend in hoeverre dit gebeurt)
Frankrijk ²⁵	Excellentie identificeren en ranking; daarnaast inzicht in kwaliteit voor onderzoeksgroepen en instituten (verbetering) en de overheid (strategische beslissingen); informeren van studenten, bedrijven en de maatschappij (verantwoording)	Rangorde aanbrengen en belonen: prestatiefinanciering: herverdeling financiering door ministerie en binnen de instelling
Australië	Identificatie van excellentie en nieuwe gebieden; internationale benchmark; incentives creëren om onderzoekskwaliteit te verbeteren	Rangorde aanbrengen en belonen: prestatiefinanciering (voorgenomen): herverdeling <i>lumpsum</i> financiering, alsmede door het verhogen van nationale budget voor onderzoek
Zweden	Strategisch universitair management gericht op het verhogen van de onderzoekskwaliteit te stimuleren	Rangorde aanbrengen en belonen: prestatiefinanciering: verdelen van 25% van <i>lumpsum</i>
Italië	Identificatie en stimulering van kwaliteit	Rangorde aanbrengen en belonen: prestatiefinanciering: het verdelen van 5% <i>lumpsum</i> , alsmede het ontmoedigen van te geringe individuele prestatie
Duitsland		
Verenigde Staten		

Bovendien zijn er in andere landen aan het doel van de evaluatie middelen verbonden om het doel te bereiken; er zijn voorgeschreven consequenties. In de meeste gevallen gaat het erom kwaliteit of excellentie te stimuleren door middel van een vorm van prestatiefinanciering: herverdeling van een deel van de lumpsum of al dan niet individuele aanstellingen toekennen. In Nederland ontbreekt het aan voorgeschreven consequenties of middelen om dat doel te bereiken. Ter verduidelijking in Box 4 een toelichting op de verschillende doelen van evaluatie.

De implicatie is dat de oordelen in die landen een zeer belangrijke rol spelen, want er zijn vergaande consequenties aan verbonden. De evaluatie moet duidelijk maken wie of wat goed genoeg is om financiering te ontvangen en wie of wat beter is en dus meer verdient. Dit verklaart de veelgebruikte methode van nauwkeurig gedefinieerde prestatietingen.

Ook in Nederland wordt gebruikgemaakt van scores. Dit suggereert dat het van belang is om een rangorde vast te stellen. Dat is echter niet het geval; er is geen noodzaak een rangorde vast te stellen om algemene doelstellingen van kwaliteitsverbetering en verantwoording te bereiken.

11. Evaluatie in de praktijk: ervaringen, gebruik en follow-up

In Nederland is in de afgelopen jaren een aantal studies²⁷ gedaan naar de praktijk van de institutionele evaluaties. Het gebruik van de protocollen en de gevolgen van de evaluatie stonden centraal. Uit deze studies komt het volgende naar voren:

- het systeem van periodieke evaluatie, van een zelfevaluatierapport schrijven, van peer review en van een managementgesprek tussen eenheid en instellingsbestuur wordt in het algemeen gewaardeerd als managementinstrument door bestuurders, decanen en groepsleiders;
- resultaten van evaluaties kunnen een rol spelen bij bestuursbesluiten (zoals een eenheid opheffen of de financiering verminderen), maar zijn nooit de enige overweging om tot een besluit te komen;
- aan lage scores zijn vaak directe consequenties verbonden, dat kan zijn de dwingende opdracht te verbeteren en hiervoor kunnen middelen beschikbaar worden gesteld;
- hoge scores worden nauwelijks direct beloond, daarvoor ontbreekt het aan financiële middelen;
- hoog scorende eenheden worden wel indirect beloond, hun wervingskracht (zoals voor externe financiering of voor personeel) neemt toe;
- het eindoordeel van de evaluatiecommissie en de cijfermatige conclusie zijn niet altijd eenduidig af te leiden uit de argumentaties;
- betrokkenen ervaren een grote administratieve last;
- lang niet alle instellingen geven volledige openbaarheid.

Uit de diverse onderzoeken in Nederland blijkt dus een waardering voor evaluatie als managementinstrument. Ook blijkt een waardering voor de autonome beslissingen die bestuurders kunnen nemen naar aanleiding van de uitkomsten van de evaluatie. In termen van doelstellingen van het SEP gaat het over kwaliteitsverbetering en over verantwoording afleggen aan het instellingsbestuur. Betrokkenen lijken de doelstelling om verantwoording af te leggen over de besteding van de publieke middelen aan financiers, de overheid en de maatschappij te negeren. Het gebrek aan openbaarheid versterkt deze indruk.

In vergelijking met andere landen is er weinig kritiek op de oordelen; noch op de gebruikte indicatoren, noch op de uiteindelijke oordelen. Aan de andere kant wordt de administratieve last als zwaar ervaren.

27 Meta Evaluatie Commissie Kwaliteitszorg Wetenschappelijk Onderzoek (2007). *Trust but Verify*. Amsterdam: KNAW; Meta Evaluatie Commissie Kwaliteitszorg Wetenschappelijk Onderzoek (2009). *E-VA-LU-E-REN. Het beoordelen van wetenschappelijk onderzoek in de praktijk*. Amsterdam: KNAW; Ben Jongbloed & Barend van der Meulen (2006). *De follow-up van onderzoeksvisities. Onderzoek in opdracht van de Commissie Dynamisering*. Enschede: CHEPS; Barend van der Meulen (2007). 'Interfering Governance and Emerging Centres of Control'. In: Whitley and Glaser (eds), (2007). *The Changing Governance of the Sciences, Sociology of the Sciences Yearbook 26*. Springer. pp. 191-2004; Leonie van Drooge, Stefan de Jong, Jos de Jonge (2012). *Focusgroepen SEP. Verslag van twee bijeenkomsten op 26 september 2012*. Den Haag: Rathenau Instituut.

12. Samenvatting

Twintig jaar institutionele evaluatie levert het volgende beeld op:

Nederland heeft een lang bestaand en stabiel systeem van kwaliteitszorg van academisch onderzoek. Zeker in vergelijking met een aantal andere landen, waar dergelijke systemen afwezig zijn of pas afgelopen decennium zijn ingevoerd.

Het Nederlandse systeem verschilt aanzienlijk van dat in andere landen. In tegenstelling tot die systemen ontbreekt het in Nederland aan een nationale doelstelling, aan een centrale organisatie die verantwoordelijkheid draagt en aan voorgeschreven consequenties verbonden aan de uitkomst van een evaluatie. De Nederlandse situatie is uniek: de verantwoordelijkheid ligt volledig bij de instellingen. De gebruikers waarderen het systeem als managementinstrument en waarderen de autonome beslissingen die bestuurders kunnen nemen naar aanleiding van de uitkomsten van de evaluatie.

De instellingen hebben gebruikgemaakt van de mogelijkheid om zelf te bepalen welke eenheden worden geëvalueerd. Er is een grote variatie in reikwijdte van de evaluatie, variërend van nationale evaluaties van hele disciplines, via hele disciplines minus een enkele instelling, of een combinatie van disciplines binnen een instelling, tot een enkel centrum of een enkele leerstoelgroep. Een totaal overzicht op evaluaties ontbrak tot nu toe. Het overzicht op de bestuursstandpunten en de gevolgen ontbreekt nog steeds. De vergelijkbaarheid van de evaluaties is uitermate klein.

De score op alle criteria - wetenschappelijke kwaliteit en productiviteit, impact en perspectief - is in de afgelopen twintig jaar toegenomen. Momenteel scoort vrijwel al het onderzoek als ten minste internationaal concurrerend. Hierdoor zijn er nauwelijks kwaliteitsverschillen waarneembaar tussen onderzoekseenheden.

Box 5: Gebruikte afkortingen

HOOP	Wetenschapsgebieden onderscheiden in het Hoger Onderwijs en Onderzoek Plan van OCW
KNAW	Koninklijke Nederlandse Akademie van Wetenschappen
MEC	Meta Evaluatie Commissie Kwaliteitszorg Wetenschappelijk Onderzoek
NWO	Nederlandse organisatie voor Wetenschappelijk Onderzoek
OCW	Ministerie van Onderwijs, Cultuur en Wetenschap (vanaf 2003)
OC&W	Ministerie van Onderwijs, Cultuur en Wetenschappen (tot 2003)
PRC	peer review committee
SEP	Standard Evaluation Protocol
SWOT	sterkte - zwakte analyse (strengths, weaknesses, opportunities, threats)
TOR	terms of reference
VF	Voorwaardelijke Financiering
VSNU	Vereniging van Universiteiten

13. Geraadpleegde bronnen en literatuur

CHEPS en Rathenau Instituut: PER Base

Danish Agency for Science, Technology and Innovation. *Research Evaluation Guidelines*. <http://en.fi.dk/research/research-evaluation/framework-and-methods/action-plan-for-research-evaluation/Research%20Evaluation%20Guidelines.pdf> (geraadpleegd 21-11-2012).

Leonie van Drooge, Stefan de Jong, Jos de Jonge (2012). *Focusgroepen SEP. Verslag van twee bijeenkomsten op 26 september 2012*. Den Haag: Rathenau Instituut.

European Commission (2010). *Assessing Europe's University-Based Research*. Expert Group on Assessment of University-Based Research.

HEFCE (2010). *Guide to funding*. http://www.hefce.ac.uk/media/hefce1/pubs/hefce/2010/1024/10_24.pdf (geraadpleegd 21-11-2012).

Ben Jongbloed & Barend van der Meulen (2006). *De follow-up van onderzoeksvisitaties. Onderzoek in opdracht van de Commissie Dynamisering*. Enschede: CHEPS.

LERU (2012). *Research universities and research assessment*. Leuven: LERU

Meta Evaluatie Commissie Kwaliteitszorg Wetenschappelijk Onderzoek (2007). *Trust but Verify*. Amsterdam: KNAW.

Meta Evaluatie Commissie Kwaliteitszorg Wetenschappelijk Onderzoek (2009). *E-VA-LU-E-REN. Het beoordelen van wetenschappelijk onderzoek in de praktijk*. Amsterdam: KNAW.

Barend van der Meulen (2007). 'Interfering Governance and Emerging Centres of Control'. In: Whitley and Glaser (eds), (2007). *The Changing Governance of the Sciences, Sociology of the Sciences Yearbook 26*. Springer. pp. 191-2004.

<https://www.starmetrics.nih.gov/> (geraadpleegd 26-11-2012).

Tweede Kamer der Staten-Generaal (vergaderjaar 1985-1986). *Beleidsnota Hoger Onderwijs Autonomie en Kwaliteit*, dossier 19 253, nrs. 1-2.

VSNU (1993). *Quality Assessment of Research – protocol 1993*. Utrecht: VSNU.

VSNU (1994). *Quality Assessment of Research – protocol 1994*. Utrecht: VSNU.

VSNU (1998). Protocol 1998. In: *Series Assessment of Research Quality*. Utrecht: VSNU.

VSNU, NWO, KNAW (2003). *Standard Evaluation Protocol 2003-2009*. Utrecht: VSNU.

VSNU, NWO, KNAW (2009). *Standard Evaluation Protocol 2009-2015*. Amsterdam: KNAW.

Werkgroep Kwaliteitszorg Wetenschappelijk Onderzoek (2001). *Kwaliteit verplicht. Naar een nieuw stelsel van kwaliteitszorg voor het wetenschappelijk onderzoek*. Amsterdam: KNAW.

Wet Hoger Onderwijs en Wetenschappelijk Onderzoek.

Ten geleide

Dit is de achtste aflevering van de Science System Assessment-reeks Feiten & Cijfers. Deze Feiten en Cijfers geeft een overzicht van de ontwikkeling van het systeem van kwaliteitszorg van het academische onderzoek. De cijfers zijn bij verschillende bronnen verzameld.

Voor nadere informatie over deze publicatie kunt u contact opnemen met de auteurs: Drs. Leonie van Drooge (l.vandrooge@rathenau.nl), Stefan de Jong MsC (s.dejong@rathenau.nl) of met het hoofd van de afdeling Science System Assessment, dr. Barend van der Meulen (b.vandermeulen@rathenau.nl).

Colofon:

© Rathenau Instituut, Den Haag
Mei 2013

Rathenau Instituut
Postbus 93566
2509 CJ Den Haag
Telefoon: 070-3421542
Website: www.rathenau.nl

Deze publicatie kan als volgt worden aangehaald:
Leonie van Drooge, Stefan de Jong et al (2013):
Twintig jaar onderzoeksevaluatie,
Feiten & Cijfers 8. Den Haag: Rathenau Instituut.

Vereenvoudiging en/of openbaarmaking door middel van druk, fotokopie of welke wijze dan ook is toegestaan voor niet-commerciële doeleinden en met adequate bronvermelding. Voor alle andere doeleinden is toestemming van de uitgever vereist.



Rathenau Instituut