



## Het beoordelen van onderzoek

**Auteur(s):**

Groot, T.L.C.M.

*De auteur is hoogleraar management accounting bij de afdeling accounting, faculteit der economische wetenschappen en bedrijfskunde aan de Vrije Universiteit Amsterdam. [tgroot@feweb.vu.nl](mailto:tgroot@feweb.vu.nl)*

**Verschenen in:**

ESB, 88e jaargang, nr. 4422, pagina 628, 19 december 2003

**Rubriek:**

Monitor

**Trefwoord(en):**

onderzoek

*Welke methode is het meest geschikt om onderzoeksprestaties te meten en welke factoren bevorderen een hoge kwaliteit en productiviteit van onderzoeksgroepen?*

Sinds de invoering van de voorwaardelijke financiering van wetenschappelijk onderzoek in 1983 is de belangstelling voor kwaliteitsmeting en -beoordeling van onderzoek sterk toegenomen. In de afgelopen jaren zijn twee manieren van prestatiemeting vooral toegepast: de bibliometrische methode en de methode van de collegiale toetsing (ook wel "peer review" genoemd).

Dit artikel beoogt twee vragen te beantwoorden. Allereerst is er de vraag in hoeverre de bibliometrische methode en collegiale toetsing tot dezelfde beoordelingen leiden. Ten tweede wordt de vraag beantwoord wat de kenmerken zijn van onderzoeksgroepen die een hoge onderzoekskwaliteit en -productiviteit halen. Om deze vragen te beantwoorden analyseren we de resultaten van de twee onderzoeksbeoordelingen en de bibliometrische gegevens die door de onderzoeksgroepen zijn verstrekt. De tweede beoordelingsronde is gebruikt als een replicatie van de eerste.

Bij de bibliometrische methode maakt men gebruik van kwantitatieve metingen van prestaties (aantal tijdschriftartikelen, boeken en andere uitingen van wetenschappelijke prestaties), veelal gewogen voor kwaliteit op basis van (gecorrigeerde) impactscores of op basis van in onderling overleg vastgestelde rangorde van tijdschriften en uitgevers<sup>1</sup>. Bij de methode van collegiale toetsing wordt een groep toonaangevende experts gevraagd een oordeel te geven over de kwaliteit van onderzoeksprestaties.

Beide methodes hebben voor- en nadelen. De bibliometrische methode beperkt zich tot in druk verschenen publicaties, is sterk afhankelijk van de gekozen methode van kwaliteitsbeoordeling van journals en uitgevers en is gevoelig voor 'publicatiegewoonten' in verschillende disciplines. De uitkomst van de methode van collegiale toetsing kan sterk worden beïnvloed door de samenstelling en werkwijze van de groep van beoordelaars. In het algemeen wordt echter verwacht dat collegiale toetsing meer mogelijkheden biedt alle relevante aspecten van de onderzoeksprestaties van onderzoeksgroepen in hun onderlinge samenhang en op een evenwichtige wijze te beoordelen. Het nadeel van collegiale toetsing is echter dat het een duur systeem is dat om die reden niet frequent kan worden toegepast.

In de economische wetenschappen bestaan beide evaluatiemethoden reeds enige tijd naast elkaar. Sinds 1991 wordt een lijst gehanteerd waarin tijdschriften van een kwaliteitsniveau zijn voorzien. Deze lijst maakt het mogelijk bibliometrische gegevens op te stellen.

---

### Methode

De onderzoeksgroepen in de economische wetenschappen (met inbegrip van bedrijfswetenschappen) zijn tweemaal door collegiale toetsing beoordeeld: de eerste keer in 1995 en de tweede keer in 2001<sup>2</sup>. De eerste beoordeling richtte zich op de publicaties in de vijfjaarsperiode 1990-1994, de tweede op de zesjaarsperiode 1995-2000. De collegiale toetsingen werden uitgevoerd door verschillende commissies, bestaande uit vooraanstaande buitenlandse onderzoekers in de economie inclusief econometrie en bedrijfskunde.

#### *De onderzoeksbeoordelingen*

De tweede beoordelingsronde is geen exacte replicatie van de eerste evaluatie, omdat de werkwijze op onderdelen verschillend was. De eerste commissie bestreek de disciplines econometrie, economie en bedrijfseconomie. In de tweede beoordelingsronde werd gewerkt met twee commissies: één voor economie (inclusief econometrie en bedrijfseconomie) en één voor bedrijfskunde. De evaluatiecommissie in 1995 beoordeelde negentig onderzoeksprogramma's in de economie, terwijl in 2001 door de economiecommissie zestig onderzoeksprogramma's en door de bedrijfskundecommissie negentien programma's werden beoordeeld. De procedure was in beide jaren sterk vergelijkbaar.

De commissies ontvingen voorafgaande aan hun beoordeling van elke onderzoeksgroep een zelfstudie. In deze zelfstudie werd een beschrijving van het programma gegeven, alsmede een lijst met publicaties, een overzicht van de inzet van personeel en overdrukken van de vijf beste artikelen. De lijst van publicaties bevatte dissertaties, wetenschappelijke publicaties (artikelen, boeken en bijdragen in boeken) en professionele publicaties. In de categorie 'wetenschappelijke publicaties' werden de publicaties ingedeeld naar kwaliteitsniveau op basis van de vsnu-classificatie: zes kwaliteitsniveaus van tijdschriften ten behoeve van het wegen van artikelen en vijf kwaliteitsniveaus van uitgevers voor het classificeren van boeken en bijdragen in boeken. De eerste commissie liet weten dat zij wel naar de aldus geclassificeerde informatie heeft gekeken, maar dat zij die niet direct heeft gebruikt.

In de tweede ronde is de vsnu-classificatie niet langer gebruikt, maar is een minder stringente indeling toegepast. Deze indeling maakt onderscheid tussen dissertaties, internationale publicaties (artikelen, boeken, bijdragen in boeken en "conference proceedings"), Nederlandse publicaties (artikelen, boeken en bijdragen in boeken) en professionele publicaties.

### Dimensies

Alle drie de commissies gaven hun oordeel met een vijf-puntsschaal op vier dimensies: wetenschappelijke kwaliteit (kwaliteit van productie in relatie tot soortgelijke onderzoeksgroepen in binnen- en buitenland), wetenschappelijke productiviteit (aantal publicaties per groep, gegeven de kwaliteit en omvang van de groep), relevantie (voor wetenschap en praktijk) en levensvatbaarheid (de langetermijnoverlevingskansen van de onderzoeksgroep gegeven de nationale en internationale competitie). Deze vier dimensies geven de commissies de mogelijkheid verschillende aspecten van wetenschappelijke productie afzonderlijk te beoordelen. De laatste twee commissies hebben een kleine verandering aangebracht in de definitie van de dimensie relevantie. In 1995 was gekozen voor een brede definitie: zowel wetenschappelijke als maatschappelijke relevantie. In de 2001-evaluatie is gekozen voor een beperkte interpretatie, namelijk alleen maatschappelijke relevantie voor professie en beleid.

### Collegiale toetsing en bibliometrie

De scores op de vier prestatiedimensies van elk van de drie commissies blijken sterk met elkaar samen te hangen en laden hoog op één gemeenschappelijke factor die een groot deel van de totale variantie verklaart (zie [tabel 1](#)). Eén prestatiedimensie wijkt duidelijk af: in de 2001-evaluaties correleert 'relevantie' minder goed met de gemeenschappelijke factor (zie de factorladingen). Dit lijkt het effect te zijn van de gewijzigde definitie van de term 'relevantie'. In het algemeen suggereren de resultaten in [tabel 1](#) dat de vier kwaliteitsdimensies inherent sterk met elkaar zijn verweven of dat de commissies niet in staat zijn een duidelijk onderscheid tussen deze dimensies te maken.

**Tabel 1. Principale Componenten Analyse op commissieoordelen**

	evaluatie	evaluatie 2001	
	1995	economie	bedrijfskunde
gemeenschappelijke factor			
eigenwaarde	3,161	2,350	2,374
verklaarde variantie	79%	58%	59%
prestatiedimensies (factorladingen)			
kwaliteit	0,921	0,838	0,899
productiviteit	0,889	0,825	0,832
relevantie	0,853	0,418	0,282
levensvatbaarheid	0,892	0,891	0,892

De resultaten van de principale componentenanalyses geven de mogelijkheid de vier afzonderlijke prestatiedimensies samen te voegen in een totaalscore van de commissieoordelen. Hiertoe hebben we het gemiddelde genomen van de scores op de vier dimensies. De totaalscore blijkt redelijk betrouwbaar. Uit [tabel 2](#) blijkt voorts dat de eindbeoordeling van de toenmalige commissie redelijk goed kan worden voorspeld door het aantal artikelen in internationale toptijdschriften (categorie a) en internationale zeer goede tijdschriften (categorie b).

**Tabel 2. ols-regressies met als afhankelijke variabele de totaalbeoordeling van de evaluatiecommissies in 1995 en in 2001**

bibliom. variabelen	evaluatie 1995 / b	bibliometrische variabelen	evaluatie 2001 / b
artikelen A	0,359***	internationale artikelen	0,904***
artikelen B	0,510***	internationale boeken/bijdragen	-0,329**
artikelen F	0,099	internationale proceedings	0,354**
boeken C	0,131	Nederlandse artikelen	0,292**
boeken E	-0,83*	professionele publicaties	-0,484***
F-waarde	19,857	F-waarde	11,212
adj. R2	0,51	adj. R2	0,615
n	89	n	79

\*\*\*: p<0,01; \*\*: p<0,05; \*: p<0,10.

Alhoewel voor de evaluatie in 2001 een andere indeling van publicatiecategorieën is gebruikt, blijkt uit [tabel 2](#) dat vooral het aantal internationale artikelen een goede voorspeller is voor de waardering van de commissies, gevolgd door internationale proceedings en Nederlandse artikelen. Tegelijkertijd zijn er ook categorieën met een tegengesteld teken: het aantal internationale boeken(bijdragen) en

professionele publicaties blijkt negatief samen te hangen met het totaaloordeel van de commissies. Ook hierin ontdekken we parallellen met 1995: toen bestond er een negatief verband tussen het aantal boeken van de laagste categorie (de zogenaamde professionele uitgaven) en het eindoordeel van de commissie.

### Kwaliteit en productiviteit

In dit gedeelte van het onderzoek proberen we de tweede vraag te beantwoorden: welke factoren bevorderen een hoge kwaliteit en productiviteit van onderzoeksgroepen? Een mogelijke verklarende factor kan de omvang van de onderzoeksgroep zijn: hoe meer onderzoekers in een groep werken, des te meer kansen er zijn dat onderzoekers elkaar stimuleren en helpen. Bovendien, hoe groter een onderzoeksgroep, des te meer middelen er zijn om ook ondersteunende activiteiten mogelijk te maken, zoals het organiseren van seminars en het bezoeken van congressen.

Daarnaast kan de samenstelling van de onderzoeksgroep belangrijk zijn. Zo ligt het voor de hand dat een relatief groot aantal promovendi (AIO's) op het totale personeel de totale kwaliteit en productiviteit van de groep stimuleert. Aio's zijn voor een groot deel van hun tijd vrijgesteld voor het doen van onderzoek, terwijl hen een grote beloning wacht in de vorm van een promotie. Een soortgelijke vrijstelling is er ook voor personeel dat betaald wordt uit de tweede geldstroom: onderzoekers betaald door het de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) hebben eigen fondsen en werken aan een duidelijk gedefinieerde onderzoeksopdracht. We veronderstellen derhalve ook dat het aandeel tweede geldstroom personeel in de onderzoeksgroep de kwaliteit en productiviteit van het programma verhoogt.

Als indicator van kwaliteit nemen we de totaalbeoordelingen van de commissies. De productiviteit hebben we gedefinieerd als het aantal geproduceerde outputs per eenheid input. Als outputindicatoren gebruiken we de in [tabel 2](#) geïdentificeerde belangrijke outputcategorieën. Voor de evaluatie van 1995 zijn dit het aantal artikelen in de categorieën a en b. Voor de evaluaties in 2001 gebruiken we het aantal internationale artikelen, het aantal internationale proceedings en het aantal Nederlandse artikelen. Als inputs gebruiken we het aantal fulltime equivalenten promovendi en overige wetenschappers van elke onderzoeksgroep. Op basis van deze outputs en inputs hebben we per evaluatieronde de relatieve efficiëntiescores van alle onderzoeksgroepen berekend door toepassing van Data Envelopment Analyse (dea). Om de verklarende factor omvang niet bij voorbaat in de dea-score te verwerken zijn de 'constant returns to scale'-modellen gebruikt. De dea-efficiëntiescores correleren hoog met de productiviteitsscores van de onderzoekscommissies<sup>3</sup> Op deze wijze hebben we twee van elkaar onafhankelijke endogene variabelen geconstrueerd: één voor kwaliteit en één voor productiviteit.

De relatie tussen kenmerken van de onderzoeksgroepen en kwaliteit van onderzoeksproductie is bepaald met OLS-regressies. We hebben de dea-scores in twee groepen verdeeld: de 2 decielen meest efficiënte programma's en de 8 decielen minst efficiënte programma's<sup>4</sup> De relatie tussen productiviteit en organisatiekenmerken is met logistische regressie geanalyseerd.

Tenslotte veronderstellen we dat ook de discipline waarin het onderzoek wordt uitgevoerd bepalend is voor de kwaliteit en productiviteit. Ten tijde van de eerste evaluatie werd verwacht dat onderzoekers in econometrie en operations research meer publicatiekansen hebben dan groepen in de economie, terwijl deze weer meer publicatiekansen hebben dan disciplines in de bedrijfseconomie en bedrijfskunde<sup>5</sup> De resultaten staan weergegeven in [tabel 3](#).

**Tabel 3. De relatie tussen organisatiekenmerken van onderzoeksgroepen en de door hen geleverde onderzoekskwaliteit (ols) en onderzoeksproductiviteit (Logistische Regressie), evaluaties in 1995 en 2001**

exogene variabelen	evaluatie 1995		evaluaties 2001	
	kwaliteit1 â	productiviteit2 â	kwaliteit1 â	productiviteit2 â
constante	-	2,351**	-	0,194
totale fte staf	0,461***	-0,057**	0,425***	-0,059*
aio's (in % van fte)	0,173 *	1,669	-0,005	-2,889
2e geldstroom (in % van fte)	-0,106	0,76	0,217*	3,408
3e geldstroom (in % van fte)	0,018	-2,721	-0,112	2,105
discipline1	0,258***	0,821**	-0,049	0,199
	F=8,639***	modelc2=14,9**	F=3,657***	model c2=13,0**
	adj. R2=0,3	McFadden 0,166	adj. R2=0,15	McFadden 0,189
	N=89	tp=28,8**; N=90	N=74	tp=21,4**; N=75

\*\*\*: p<0,01; \*\*: p<0,05; \*: p<0,10.

1. ordening is als volgt: (1) bedrijfseconomie/bedrijfskunde; (2) algemene economie; (3) econometrie en mathematische beslissonderzoek, operations research.

In 1995 behaalden grotere onderzoeksgroepen, groepen met relatief veel promovendi en groepen die in meer analytische disciplines opereren inderdaad een hogere kwaliteit. Publicatiekansen lijken structureel scheef te zijn verdeeld over de disciplines.

Verrassend is voorts de rol van omvang in de productiviteit van onderzoeksgroepen. Terwijl grotere groepen hogere kwaliteit leveren,

blijken ze tevens minder productief. Naarmate de onderzoeksgroep meer mensen telt, lijkt het voor de leiding moeilijker de productiviteit van elke medewerker te beheersen. Groei van de onderzoeksgroep leidt kennelijk ook tot het ontstaan van beheersingsproblemen.

Replicatie van het onderzoek in 2001 wijst uit dat de omvang van de groep nog steeds positief correleert met kwaliteit en negatief met productiviteit. Het lijkt derhalve dat dit een redelijk persistent gegeven is. Daartegenover staat dat de invloed van de discipline is verdwenen. Kennelijk is het in 1995 gevonden verband niet structureel, maar tijdelijk van aard.

In de evaluatie van 1995 zijn vooral onderzoeksgroepen in het domein van de bedrijfseconomie kritisch beoordeeld. De nieuwe resultaten wijzen erop dat de bedrijfseconomische onderzoeksgroepen een inhaalslag hebben gemaakt, waarmee ze hun structurele handicap in publicatiemogelijkheden logenstraffen.

Een derde opvallend resultaat is de invloed van de samenstelling van de onderzoeksgroepen. Nu leidt niet het percentage AIO's mede tot hogere kwaliteit, maar het percentage tweedegeldstroompersoneel. Dit kan te maken hebben met het toegenomen aandeel dat de tweede geldstroom in de financiering van wetenschappelijk onderzoek in de loop der tijd heeft gekregen.

## Conclusie

Ons onderzoek heeft laten zien dat de evaluatiecommissies weliswaar op vier verschillende dimensies rapporteren, maar dat deze dimensies een sterke onderlinge samenhang vertonen. De beoordeling van de commissies kan redelijk goed op basis van bibliometrische informatie worden voorspeld. Naast de hier geanalyseerde scores zijn evaluatiecommissies door collegiale toetsing ook in staat andere aspecten in de beschouwing te betrekken en hierover genuanceerd te oordelen. Een combinatie van collegiale toetsing en bibliometrische informatie lijkt dan ook de beste benadering voor de beoordeling van onderzoeksprestaties te zijn.

Tom Groot

---

1 Commissie Herziening Outputmeting Economie, *Towards a new system of output measurement in Dutch economics*, final report, VSNU, Utrecht, 1999.

2 VSNU (Vereniging van Samenwerkende Nederlandse Universiteiten), *Quality assessment of research in economics* (PU/130316), Utrecht, 1995; VSNU, *Assessment of research quality in business administration and management* (PU/130350), Utrecht, 2002; VSNU, *Assessment of research quality in economics* (PU/130347), Utrecht, 2002.

3 De correlatiecoëfficiënt is 0,449 voor  $p < 0,05$ .

4 De DEA-scores laten OLS niet toe, omdat zij niet normaal zijn verdeeld. In het algemeen vertonen DEA-scores een tweetoppige verdeling.

5 De Cronbach alpha is 0,85 voor 1995, 0,75 voor economie 2001 en 0,74 voor bedrijfskunde 2001. 3. VSNU, *Quality assessment of research in economics* (PU/130316), Utrecht, 1995, blz. 6; E. Sterken en P.S.H. Leeflang, Een expertlijst van tijdschriften in de bedrijfswetenschappen, *Maandblad voor Accountancy en Bedrijfseconomie*, jrg. 74, nr. 4, blz. 73-85.