

Heeft het wegen van wetenschappelijke publikaties zin?

Rita Poppen en Frans Stokman*

Summary

Is it useful to weigh scientific publications?

In this paper, scientific productivity of academic institutions, research groups, and researchers in Sociology is measured by counting the numbers of scientific publications. We investigate whether the use of weight factors for length, quality, and multiple authorship has any effect on productivity ranking. We conclude that this is not the case for length and multiple authorship, but that weighting for quality matters. At the level of academic institutions and research groups, this effect is largest for publications in the Dutch language; at the level of the individual author, weighting for quality has the strongest effect for articles in journals.

1. Inleiding en centrale vraagstelling

De invoering van periodieke beoordeling van onderzoeksgroepen via het stelsel van de Voorwaardelijke Financiering in 1984 en het gebruik van publikatietellingen door enkele landelijke visitatiecommissies heeft het vraagstuk van de meting van wetenschappelijke produktiviteit ook in Nederland een directe beleidsrelevantie gegeven. In dit artikel gaan wij niet in op de vraag of wetenschappelijke beoordelingen moeten worden gebaseerd op 'peer review', citatieanalyse, het (al dan niet gewogen) tellen van publikaties of een combinatie van deze drie. Ons standpunt is dat alle drie elementen hun eigen waarde hebben in de beoordeling van wetenschapsgroepen en personen. Uiteindelijke beoorde-

* Werkzaam bij het Interuniversitair centrum voor Theorievorming en Methodenontwikkeling in de sociologie (ICS), RUG. Dit artikel is een vervolg op de gelijknamige scriptie van R.S. Poppen in het kader van haar doctoraalstudie.

lingen kunnen slechts op 'peer reviews' zijn gebaseerd, maar kwantitatieve wetenschapsindicatoren kunnen bijdragen tot goede 'peer reviews' en kunnen deze in elk geval onderbouwen. Dit geldt zeker voor vrij massale beoordelingen over een langere tijd, zoals dat bij de VFO-beoordelingen van de KNAW het geval is. Daarbij verwijst de citatieanalyse meer naar de 'impact' van onderzoeksgroepen en personen (indien althans een groot aantal valkuilen worden vermeden), terwijl publikatietellingen meer verwijzen naar de onderzoeksproductiviteit (Moed & Van Raan, 1987; Moed, 1985).

De weerstand tegen het tellen van publikaties is vooral gebaseerd op het idee dat een kleine publikatie in een onbeduidend medium even zwaar weegt als een publikatie in een vooraanstaand internationaal tijdschrift of een boek bij een gerenommeerde uitgever. Die weerstand is begrijpelijk en terecht. Toen het Nederlandse wetenschapsbeleid in het begin van de jaren tachtig meer output georiënteerd werd, ging dat om deze reden gepaard met stringenter criteria welke publikaties wel en welke niet mochten worden opgenomen in de rubriek van wetenschappelijke publikaties van de wetenschappelijke verslagen van de Nederlandse universiteiten. Ontegenzeggelijk is de kwaliteit van de wetenschappelijke verslagen van de Nederlandse universiteiten sinds die tijd aanzienlijk vooruitgegaan (maar nog lang niet perfect) en zijn met name veel publikaties in het grijze circuit, waarvan de waarde moeilijk vast te stellen was, uit deze rubriek verdwenen. De vraag die wij in dit artikel willen beantwoorden is nu, of weging naar omvang, kwaliteit en meervoudig auteurschap binnen deze categorie van publikaties leidt tot omvangrijke verschuivingen in rangordes van instellingen, onderzoeksgroepen en individuen naar produktiviteit. Deze vraag zal worden beantwoord voor de discipline Sociologie, waar de heterogeniteit van publikaties groter mag worden geacht dan voor meer gevestigde wetenschapsgebieden als de beta-wetenschappen en psychologie.

De onderzoeksvraag heeft grote beleidsrelevantie. Indien ongewogen tellingen dezelfde rangordes opleveren dan gewogen tellingen, kunnen ingewikkelde weegprocedures achterwege blijven en geven de overzichten in de wetenschappelijke verslagen een redelijk beeld van de produktiviteit van een onderzoeksgroep. Daarnaast heeft de vraag echter ook wetenschappelijke waarde, omdat de zinvolheid van weging bij indexconstructie in sociaal-wetenschappelijk onderzoek (bijv. bij regressie- en factoranalyse) reeds lang betwijfeld wordt (Lawshe & Schucker, 1959; Schmidt, 1971; Wainer, 1976; Nijssen & Pijl, 1977): keuzes van verschillende weegfactoren (waaronder gelijke gewichten) leiden niet tot wezenlijk andere resultaten en zijn derhalve overbodig. Indien dat ook geldt voor de sommatie van wetenschappelijke publikaties, dan is het weinig zinvol nieuwe weegmethoden te bedenken. Een voorbeeld hiervan is de

intentie van Kapteyn en Wansbeek (1985) om publikaties zo te wegen dat zij optimaal het aantal te verwachten citaties voorspellen.

De opbouw van dit artikel is als volgt: na een beschrijving van de data en de globale onderzoeksopzet wordt nader ingegaan op de gebruikte weegfactoren en het beoordelaarsonderzoek dat nodig was voor de vaststelling van de kwaliteit van de Nederlandstalige tijdschriften en uitgevers. Vervolgens wordt het effect van de verschillende combinaties van weegfactoren op de produktiviteit op drie niveaus gerapporteerd: op het niveau van de studierichtingen, de vakgroepen en individuele onderzoekers. Ten slotte volgen conclusies en aanbevelingen voor verder onderzoek.

2. De data

Om de centrale vraagstelling te beantwoorden zijn uit een in opbouw zijnde database voor wetenschappelijke publikaties in de disciplines sociologie, politologie en bestuurskunde over de periode 1983-1991 de gegevens omtrent de in totaal 2456 wetenschappelijke publikaties in de Nederlandse sociologie van 1983-1986 voor de analyses gebruikt.¹ De gegevens van de publikaties zijn overgenomen uit de wetenschappelijke verslagen van Nederlandse universiteiten.

Alleen de titels uit de rubrieken 'dissertaties' en 'wetenschappelijke publikaties' zijn in de database opgenomen. Vakpublikaties, boekbesprekingen en abstracts worden niet tot de categorie van wetenschappelijke publikaties gerekend. De database is een relationele database en bevat alle beschikbare informatie over een publikatie. Zo wordt per publikatie aangegeven hoeveel pagina's ze omvat, in welk tijdschrift ze is gepubliceerd of door welke uitgever ze is uitgegeven. Ook kan er worden gevonden welke auteurs er bij de publikatie horen. Door de relationele opbouw van de database kan zonder verlies van informatie de database worden geanalyseerd op verschillende niveaus, zoals die van de publikatie zelf, de auteur, het medium, de onderzoeksgroep et cetera. Voor meer informatie over de opbouw en de inhoud van de database wordt verwezen naar Stokman et al. (1989).

3. Onderzoeksopzet en weegfactoren

Voordat de reeks publikatietellingen kon plaatsvinden werden eerst de weegfactoren voor omvang, meervoudig auteurschap en kwaliteit geformuleerd.

3.1. *Weging naar omvang*

In de scientometrische literatuur wordt aan de omvang van een publikatie doorgaans weinig aandacht besteed. Becker en Van der Merwe (1980) hebben een beoordelingsformulier voor wetenschappelijke publikaties ontwikkeld, waarin ook de omvang van de publikatie aandacht krijgt. Hiervoor hanteren zij het begrip norm-pagina. Hun methode is erg omslachtig en tijdrovend. Zie ook Van Oostende (1990).

In onderhavig onderzoek zijn de weegfactoren voor weging naar omvang afgeleid van de classificatie die Van Heeringen (1983) gebruikte in zijn studie binnen natuurkunde, scheikunde en economie. Hij gaat uit van 4 klassen: de weegfactor 1 voor publikaties tot en met 50 pagina's, 2 voor 51-100 pagina's, 3 voor 101-200 en 4 voor meer dan 201 pagina's. Binnen sociologie bestaat er echter nogal wat differentiatie onder de 50 en boven de 201 pagina's, zodat het ons juist voorkwam om de schaal in die twee gebieden verder te differentiëren. De weegfactoren zijn als volgt gekozen:

<i>aantal pagina's</i>	<i>weegfactor</i>
<= 12	1
13 - 25	2
26 - 50	3
51 - 100	4
101 - 200	5
201 - 400	6
401 - 800	7
>= 801	8

In de wetenschappelijke verslagen worden publikaties onder redactie afzonderlijk vermeld naast de eventuele eigen bijdragen in de reader. Bovenstaande scores worden door twee gedeeld indien het gaat om een redacteurschap.

3.2. *Weging naar meervoudig auteurschap*

Voor het meten van wetenschappelijke produktiviteit op het individuele niveau is het van belang om rekening te houden met meervoudig auteurschap. Op het niveau van vakgroepen en instellingen is dit belang minder, omdat weinig samenwerking tussen verschillende onderzoeksgroepen en instellingen plaatsvindt (slechts 76 van de 2456 ofwel 3,1% van de publikaties bleken geschreven te zijn door auteurs die aan verschillende sociologische vakgroepen verbonden

waren). In het onderhavige onderzoek is gekozen om bij weging naar meervoudig auteurschap de score van de publikatie gelijkelijk te verdelen over de bijbehorende auteurs. Is dus een artikel geschreven door drie personen, dan krijgt ieder een derde deel van de score ($1/3$ in het geval van een ongewogen publikatie).

Deze verdeelsleutel is de meest eenvoudige die denkbaar is. In andere onderzoeken worden ook wel andere verdeelsleutels gebruikt, maar de daarbij onderliggende gedachten zijn meestal niet zo duidelijk. Ook wordt er wel rekening mee gehouden of de auteurs al of niet alfabetisch worden gepresenteerd bij een publikatie (zie Becker & Van der Merwe, 1980; Van Heeringen, 1983 en Van Oostende, 1990). In onderhavig onderzoek wordt geen rekening gehouden met een eventuele alfabetische volgorde van de auteurs, omdat dit in de meeste gevallen eveneens een omstreden zaak is. In geval van een alfabetische volgorde kan het wel degelijk zo zijn dat de belangrijkste auteur voorop staat (zie ook Hofstee, 1988).

Bij dissertaties wordt ook aan de promotores een deel van de publikatie toegewezen. Hier krijgt de promovendus $3/4$ deel van de publikatiescore, terwijl de bijbehorende promotoren het resterende $1/4$ moeten delen.

3.3. Weging naar kwaliteit

Knorr et al. (1979) geven in hun werk aan dat de meeste wetenschapsonderzoekers het idee van 'assessment by peers' als de meest valide methode voor het meten van kwaliteit aanhangen, maar dat toch vanwege de uitvoerbaarheid tellingen van publikaties in de meeste wetenschapsonderzoekingen worden gebruikt. In de praktijk wordt 'assessment by peers' dan ook zelden voor scientometrische toepassingen gebruikt (zie voorts Becker & Van der Merwe, 1980 en Van Spiegel, 1985).

Een veelgebruikte methode om de kwaliteit van een publikatie te meten is de citatie-analyse (zie bijvoorbeeld Moed & Van Raan, 1987 en Peritz, 1983): hoe vaker een publikatie door een andere auteur wordt aangehaald, des te beter is het artikel. Hierbij moet worden opgemerkt dat citatie-analyse verwijst naar de impact van een publicatie, hetgeen wellicht wel gecorreleerd is met kwaliteit maar daar niet gelijk aan kan worden gesteld. Daarnaast zijn aan citatie-analyse een aantal andere problemen verbonden. Van Spiegel (1985) bespreekt ze uitgebreid. Het meest problematisch aspect is het optreden van zogenaamde citatieclusters/cirkels. Dit verschijnsel houdt in dat er groepen wetenschappers zijn die elkaar steeds weer blijken te citeren. Hagendijk en Prins (1984) hebben in hun onderzoek aan kunnen tonen dat ook in de Nederlandse sociologiebeoefening citeernetwerken aanwezig zijn. Deze citeernetwerken zijn een algemeen

voorkomend verschijnsel en daarom is citatie-analyse niet zo geschikt om kwaliteit van publikaties te onderzoeken. Het is wel een goed instrument om samenwerkingsverbanden en schoolvorming op te sporen.

Een andere manier om de kwaliteit van een publikatie te onderzoeken is aan de hand van de kwaliteit van het medium waarin de publikatie is opgenomen, i.e. het tijdschrift of de uitgever. Dit wordt onder meer gedaan door Van Heeringen (1983), Van Loon et al. (1988) en Van Oostende (1990). Ook bij de differentiatie van tijdschriften naar kwaliteit wordt vaak citatie-analyse gebruikt: tijdschriften die vaak worden geciteerd hebben een hoge status, tijdschriften die zelden geciteerd worden een lage. Op deze wijze wordt ook de impactscore van tijdschriften door de Social Science Citation Index (SSCI) vastgesteld.

Een andere methode om de kwaliteit van een tijdschrift te onderzoeken is door middel van tijdschriftenquêtes (Nederhof, 1988). Aan deskundigen wordt gevraagd wat ze van de kwaliteit van een bepaald tijdschrift vinden. Deze methode is goed bruikbaar mits het aantal te beoordelen media niet te groot is en over een te groot wetenschapsgebied gespreid is.

In het hier gerapporteerde onderzoek is gekozen om de kwaliteit van een publikatie af te leiden van het bijbehorende medium. Wegens de problemen met de vaststelling van kwaliteit van media aan de hand van citatieanalyse is gekozen voor een niet al te ver gedifferentieerde kwaliteitsscore, namelijk een driedeling in wetenschappelijk gerenommeerde media, media met een wetenschappelijke status en media zonder wetenschappelijke status. Deze score is deels bepaald door beoordelaars, deels door gebruik te maken van de SSCI. Buitenlandse tijdschriften die in de SSCI zijn opgenomen zijn alle tot de categorie van wetenschappelijk gerenommeerde tijdschriften gerekend. Andere buitenlandse tijdschriften zijn beschouwd als media zonder wetenschappelijke status. Omdat voor uitgevers een dergelijke bron niet aanwezig is, hebben wij buitenlandse boeken bij weging naar kwaliteit helaas buiten beschouwing moeten laten. Voor Nederlandse tijdschriften, die nauwelijks worden gedekt door het SSCI, en voor Nederlandse uitgevers is een beoordelaarsonderzoek uitgevoerd, geïnspireerd door Nederhof (1988). Er zijn voor het beoordelaarsonderzoek vijf deskundigen op het gebied van de sociologie benaderd met het verzoek om aan het beoordelaarsonderzoek deel te nemen. Deze deskundigen zijn allen hoogleraar en kunnen gezien worden als representanten van verschillende stromingen binnen de Nederlandse sociologie. Vier van hen hebben uiteindelijk aan het onderzoek deelgenomen. Aan de respondenten is gevraagd om twee beoordelingsformulieren in te vullen. Een formulier bevatte een lijst met uitgevers, het andere een lijst met tijdschriften. Nederlandse sociologen maken voor het publiceren gebruik van talloze tijdschriften en uitgevers, zodat selectie van de media vooraf onontkoombaar was, want het zou te veel vragen van de deskundi-

gen om over alle uitgevers en niet in de SSCI opgevoerde tijdschriften in de database een oordeel te vellen. Daarom is er zo verantwoord mogelijk een voorselectie gemaakt. Om de Nederlandstalige wetenschappelijke toptijdschriften te kunnen onderscheiden is in het beoordelingsformulier een selectie *Nederlandse tijdschriften* opgenomen. Van de Nederlandstalige tijdschriften uit de database is nagegaan of ze zijn opgenomen in de – zeer uitgebreide – collectie van het Sociologisch Instituut in Groningen. Zo ja, dan zijn ze opgenomen. Zo kon worden voorkomen dat elk onbeduidend orgaan moest worden beoordeeld. De voorselectie van *uitgevers* was niet eenvoudig, want de database bevat veel vage en onbekende uitgevers. Om het overzichtelijk te houden is ervoor gekozen om uitsluitend Nederlandse uitgevers in het beoordelingsonderzoek op te nemen die met een ISBN-nummer geregistreerd staan in The Publishers' International Directory With ISBN Index.² De resultaten van de indelingen van de beoordelaars zijn te vinden in bijlage 1 voor de tijdschriften en bijlage 2 voor de uitgevers.

Aan de categorie van de wetenschappelijke gerenommeerde media is de weegfactor 4 toegekend, aan die van de media met wetenschappelijke status de weegfactor 3 en aan de overige media de weegfactor 1. Deze weegfactoren zijn zo gekozen omdat het onderscheid tussen media met en zonder wetenschappelijke status van groter belang is dan dat tussen gerenommeerd of niet. Het gaat immers om de rubriek van wetenschappelijke publikaties in de jaarverslagen.

Dissertaties hebben onafhankelijk van de betreffende uitgever in dit onderzoek altijd de weegfactor 4 toegekend gekregen.

3.4. Onderzoeksopzet

Na de vaststelling van alle weegfactoren konden de publikatietellingen worden verricht op de drie reeds genoemde niveaus: de wetenschappelijke instelling, de vakgroep en de individuele auteur. De gewogen tellingen worden steeds vergeleken met de ongewogen telling over *dezelfde* publikaties om het effect van de weging eenduidig vast te stellen. De afzonderlijke weging naar omvang en kwaliteit is op alle niveaus toegepast. Op individueel niveau is bovendien gewogen naar meervoudig auteurschap. Uit voorgaande analyses was reeds naar voren gekomen dat weging naar meervoudig auteurschap op het niveau van instelling of vakgroep niet veel effect te zien gaf. De reden hiervoor is dat co-auteurschap zich veelal binnen de instellingen en ook binnen de vakgroepen afspeelt, of er wordt samengewerkt met personen buiten de universitaire instellingen, maar co-auteurschap tussen vakgroepen is niet zo gebruikelijk.

Naast de afzonderlijke wegingen zijn ook combinaties van weegfactoren toegepast. Op alle niveaus werden de effecten van de combinatie van omvang en

kwaliteit onderzocht, terwijl op individueel niveau ook de combinatie van omvang, kwaliteit en meervoudig auteurschap werd onderzocht. De afzonderlijke weegfactoren worden in geval van combinatie met elkaar vermenigvuldigd.

Om op instellings- en vakgroepsniveau vergelijkbare produktiviteitscijfers te krijgen zijn de tellingen gerelateerd aan de bijbehorende hoeveelheid fte (full-time-equivalent) voor onderzoek. Zouden alleen de ruwe uitkomsten worden onderzocht, dan zou de interpretatie bijna onmogelijk zijn. De grootste vakgroep met veel publikaties zou ook na weging steeds de grootste zijn. Door de ruwe uitkomsten steeds te delen door de omvang fte, worden de vakgroepen en instellingen onderling beter vergelijkbaar, zodat het berekenen van rangcorrelatiecoëfficiënten zinvol wordt. De omvang van de fte per vakgroep is verkregen uit de wetenschappelijke verslagen. De omvang van de fte per instelling is verkregen door aggregatie over de vakgroepen.

De rapportage van de resultaten van het onderzoek geschiedt zodanig dat de identiteit van de instellingen, vakgroepen en individuele onderzoekers onbekend blijft. Omdat weegfactoren steeds slechts voor een deel van de publikaties konden worden vastgesteld, zouden overzichten met rangordes van instellingen, vakgroepen en onderzoekers steeds betrekking hebben op deelverzamelingen van publikaties. Dergelijke deeloverzichten zijn ongeschikt om de totale produktiviteit van instellingen, vakgroepen en onderzoekers te beoordelen. Wel zullen eind 1992, in een aparte bijdrage aan Mens en Maatschappij, verantwoorde rangscores per instelling, vakgroep en onderzoekers op basis van meer recente gegevens in de database worden gepubliceerd, waarbij rekening gehouden wordt met de bevindingen in dit artikel.

4. De effecten van weegfactoren

In deze paragraaf wordt verslag gedaan van de effecten van weging naar omvang, weging naar kwaliteit, weging naar meervoudig auteurschap en enkele combinaties van deze weegfactoren. Door de weegfactoren per niveau eerst onafhankelijk te onderzoeken op hun effecten is het mogelijk om nog van zoveel mogelijk informatie gebruik te maken. We hoeven dan bijvoorbeeld bij het wegen naar omvang alleen die publikaties buiten beschouwing te laten waarvoor een ontbrekende waarde geldt voor omvang. Als het medium van dezelfde publikatie wel bekend is, dan kan de betreffende publikatie wel worden opgenomen bij de publikatie-telling waarbij het gaat om weging naar kwaliteit. Een reden hiervoor is dat we te maken hebben met veel ontbrekende waarden in de database. Het aantal analyse-eenheden kan dus per vorm van weging verschillen.

Steeds is er naast een gewogen publikatie-telling een ongewogen telling over dezelfde publikaties verricht. Dat maakt het mogelijk om de effecten van de weging aan te tonen met behulp van correlatiecoëfficiënten.

In tabel 1 worden eerst de resultaten op instellingsniveau gegeven.

Tabel 1. Correlatie-coëfficiënten tussen ongewogen en gewogen tellingen op instellingsniveau

Gewogen naar	Ongewogen			
	pearson's r	R ²	spearman's rho	N
Omvang	0,94	0,88	0,88	8
Kwaliteit-artikelen	0,99	0,81	0,90	8
Kwaliteit-ned. publ.	0,80	0,64	0,48	8
Omvang-kwaliteit-art.	0,98	0,96	0,83	8
Omvang-kwaliteit-ned. publ.	0,82	0,67	0,67	8

Weging naar omvang blijkt op instellingsniveau niet veel effect te hebben in vergelijking met een ongewogen optelling. Pearson's rho bedraagt 0,94 en de gemeenschappelijke variantie (R^2) – het gemeenschappelijke dat de ongewogen telling en de gewogen telling meten – bedraagt 0,88 (0,94 in het kwadraat). De rangcorrelatie is hoog te noemen. Dit betekent dat de instellingen na weging nauwelijks van plaats zijn gewisseld wat rangnummer betreft. Lange en korte publikaties zijn blijkbaar evenredig over de instellingen verdeeld. Afzonderlijk wegen naar omvang geeft dus weinig effect.

Bij het wegen naar kwaliteit zien we bij de artikelen ongeveer hetzelfde, ook hier heeft afzonderlijk wegen niet veel effect. Bij het wegen naar kwaliteit bij Nederlandse publikaties zien we de zaak meer veranderen. De gemeenschappelijke variantie is nu slechts 0,64. Het effect van de weging naar kwaliteit op de rangorde is hier opmerkelijk. Spearman's rho bedraagt slechts 0,48. Dit geeft aan dat de rangorde behoorlijk verandert. Om enkele sprekende voorbeelden te noemen: de instellingen die ongewogen op de tweede en derde plaats stonden, zakten elk drie rangordes naar beneden en komen zo op de vijfde en zesde plaats terecht! Een mogelijke verklaring hiervoor kan zijn dat de tijdschriften-artikelen met een hoge status evenredig over de instellingen zijn verdeeld, maar dat de Nederlandse boeken en rapporten met een lage status bij enkele instellingen zijn geconcentreerd. Bij de combinatie omvang-kwaliteit blijkt dat weging nauwelijks effect heeft op artikelen. Bij de Nederlandse publikaties geeft

de combinatie omvang-kwaliteit eveneens een veel lagere gemeenschappelijke variantie en een lage Spearmans rho van 0,67. Na weging is de rangorde beduidend anders dan voor de weging. We kunnen op instellingsniveau concluderen dat weging naar kwaliteit voor Nederlandse publikaties zin heeft.

Tabel 2. Correlatie-coëfficiënten tussen ongewogen en gewogen tellingen op vakgroepsniveau

Gewogen naar	Ongewogen			
	pearson's r	R ²	spearman's rho	N
Omvang	0,95	0,90	0,94	47
Kwaliteit-artikelen	0,85	0,72	0,89	43
Kwaliteit-ned. publ.	0,89	0,79	0,79	47
Omvang-kwaliteit-art.	0,86	0,74	0,88	43
Omvang-kwaliteit-ned. publ.	0,88	0,77	0,72	47

Tabel 2 laat zien dat op vakgroepsniveau weging naar omvang eveneens weinig effect heeft. Weging naar kwaliteit geeft wel meer effect te zien, zowel voor tijdschriftartikelen als voor Nederlandse publikaties. Het effect van deze laatste is echter minder dan op instellingsniveau.

De combinaties van weegfactoren blijken op vakgroepsniveau iets meer effect te hebben dan op instellingsniveau – tenminste wat de pearsons rho betreft. De rangorde van de vakgroepen verandert wel, maar niet beduidend. Toch komt een rangorde coëfficiënt van 0,72 bij de combinatie van alle weegfactoren erop neer dat zowel in de bovenste 10 als in de onderste 10 vakgroepen vier van de 10 vakgroepen uit de top respectievelijk de onderkant verdwijnen na weging. Dit houdt ons inziens in dat ook op vakgroepniveau geen beleidsmaatregelen op ongewogen publikatietellingen mogen worden gebaseerd.

Tabel 3. *Correlatie-coëfficiënten tussen ongewogen en gewogen tellingen op individueel niveau*

Gewogen naar	Ongewogen			
	pearson's r	R ²	spearman's rho	N
Omvang	0,92	0,85	0,81	937
Kwaliteit-artikelen	0,89	0,79	0,80	527
Kwaliteit-ned. publ.	0,93	0,86	0,78	922
Meervoudig auteurschap	0,96	0,92	0,86	1135
Omvang-kwaliteit*	0,79	0,62	0,73	494
MVAuteurschap-artikelen				
Omvang-kwaliteit*	0,82	0,67	0,71	770
MVAuteurschap-ned. publ.				

Zoals uit tabel 3 blijkt, is op het niveau van het individu – de auteur – het effect van weging naar omvang nog steeds niet sterk, alhoewel spearman's rho nu wel daalt tot 0,81. Opmerkelijk is dat weging naar kwaliteit op individueel niveau het grootste effect veroorzaakt bij de artikelen, terwijl op instellings- en vak-groepsniveau de Nederlandse publikaties gevoeliger blijken voor wegen naar kwaliteit. Het is mogelijk dat de goede tijdschriftartikelen geconcentreerd zijn bij enkele personen.

Op het individuele niveau blijkt dat weging naar alle drie aspecten wel degelijk een behoorlijk effect heeft. Dit geldt zowel voor artikelen als voor Nederlandse publikaties. In beide gevallen is Spearman's rho slechts iets hoger dan 0,70 en is de gemeenschappelijke variantie met ongewogen tellingen slechts in de zestig procent. Om een indruk te geven welk effect dit heeft op de top: voor zowel de artikelen als de Nederlandse publikaties geldt dat van de 20 sociologen met de meeste publicaties er 8 uit de top-20 verdwijnen na weging naar omvang, kwaliteit en meervoudig auteurschap.

5. Conclusies en aanbevelingen voor verder onderzoek

Voor het niveau van de individuele onderzoeker maken elk van de weegfactoren afzonderlijk weinig uit, maar in interactie met elkaar hebben zij een substantiële invloed op de rangorde tussen wetenschappelijke onderzoekers. Om-

dat juist de interactie tussen alle drie weegfactoren invloed uitoefent, dienen alle weegfactoren op het individuele niveau te worden meegenomen.

Op het geaggregeerde niveau van de vakgroep is weging naar omvang onbelangrijk, maar heeft de weging naar kwaliteit wel degelijk een substantiële invloed op de rangorde tussen de vakgroepen. Dit geldt zowel voor de artikelen als voor de Nederlandse publikaties. Op instellingsniveau speelt alleen de weging naar kwaliteit een belangrijke rol bij de Nederlandse publikaties. De verklaring moet worden gezocht in de verschillende eisen die door de verschillende instellingen worden gesteld aan publikaties die voor opname in de rubriek wetenschappelijke publikaties mogen worden opgenomen. Juist bij Nederlandse publikaties is dit demarkatiecriterium boterzacht. Zo kan het voorkomen dat het bij een ongewogen telling lijkt alsof een te onderzoeken eenheid een enorme hoeveelheid publikaties heeft voortgebracht, terwijl na weging naar kwaliteit naar voren komt dat veel van die publikaties geen wetenschappelijke status bezitten. De belangrijkste conclusie van dit onderzoek is dan ook: juist doordat er geen uniforme richtlijnen zijn voor het opstellen van het wetenschappelijk verslag is weging naar kwaliteit, met name op instellingsniveau (i.e. op het niveau van studierichtingen) noodzakelijk. Aangezien dat juist die wegingsvorm is die het moeilijkst te kwantificeren is, is het belangrijk om nog even stil te staan bij de validiteit van deze wegingsvariant.

Ten eerste kunnen we ons afvragen of de afstanden tussen de kwaliteitscategorieën (1, 3 en 4) misschien te groot gekozen zijn. Waarschijnlijk is dit niet het geval. De hoogste categorie weegt slechts vier keer zo zwaar als de laagste. Toch zou het zinvol zijn om bij eventuele replicatie van het beoordelaarsonderzoek, van de deskundigen aanvullende informatie te vragen die het mogelijk maakt om de afstanden tussen de categorieën nauwkeuriger vast te stellen.

Ten tweede rijst de vraag of de vier deskundigen in het beoordelaars onderzoek voldoende representatief zijn voor de Nederlandse sociologie. Alhoewel de benaderde deskundigen gewetensvol zijn uitgekozen, blijft de representativiteit een moeilijk punt als gekozen wordt voor een beperkt aantal beoordelaars. Bij een volgend onderzoek moet zeker worden gepoogd om het aantal deskundigen te verhogen. Het Tijdschrift voor Vrouwenstudies bijvoorbeeld bleek bij geen van de beoordelaars bekend, daarom kwam het in de niet wetenschappelijke categorie terecht.

Ondanks voornoemde overwegingen en tekortkomingen mag echter wel worden geconcludeerd dat weging naar kwaliteit betekenis heeft.

Rest nog de vraag of de uitkomsten van het onderhavige onderzoek gegeneraliseerd kunnen worden naar andere disciplines dan de sociologie. We kunnen gerust aannemen dat generalisatie naar andere gamma-disciplines verantwoord is. Er is een enorme overlap wat media betreft. Sociologen publiceren in talloze

tijdschriften behorende bij andere disciplines. Auteurs uit de andere gamma-wetenschappen gebruiken dezelfde media.

De vraag: 'Heeft wegen van wetenschappelijke publicaties zin?' kan worden beantwoord met: 'Ja, als het gaat om weging naar kwaliteit, dan heeft wegen zeker zin!'

Noten

1. Deze database is een uitvloeisel van het werk van de Visitatie-commissie Sociologie, Politologie en Bestuurskunde die op 28 november 1986 werd ingesteld (Van Loon e.a. 1988). De database werd gedefinieerd en wordt opgebouwd binnen het Interuniversitaire Centrum voor Theorievorming en Methodenontwikkeling in de Sociologie (ICS) dankzij een subsidie van het Ministerie van Onderwijs en Wetenschappen.
2. Voor een uitvoerige beschrijving van het beoordelaarsonderzoek wordt verwezen naar Hoofdstuk 4 van R.S. Poppen (1990).

Literatuur

- Becker, H.A. & C. van de Merwe (1980). Beoordeling van publikaties in de (sociale) wetenschappen, een praktisch voorstel. *Universiteit en Hogeschool*, jrg. 27, no. 1, p. 3-29.
- Buijtendorp, T., Voor het eerst dringt een vrouw door tot economen top-30. *NRC-Handelsblad*, 6 december 1990.
- Claudy, J.G. (1972). A comparison of five variable weighting procedures. *Educational and psychological measurement*, 32, p. 311-322.
- Hagendijk, R.P. & A.A.M. Prins (1984). Referenties en Reverences – Onzekerheid, afhankelijkheid en citeernetwerken in de Nederlandse sociologie. *Mens en Maatschappij*, 59, p. 226-280.
- Heeringen, A. van (1983). *Relaties tussen leeftijd, mobiliteit en produktiviteit van wetenschappelijke onderzoekers*. RAWB, 's-Gravenhage.
- Hofstee, W.K.B. (1987). Auteursvolgorden bij wetenschappelijke artikelen, in *Universiteit en Hogeschool. Tijdschrift voor wetenschappelijk onderwijs*, 5, p. 211-213.
- Kapteyn, A., & T.J. Wansbeek (1985). Publikaties en kwaliteitsmeting bij economisch onderzoek. In: Becker, H.A. & A.F.J. van Raan (red.), *Kwaliteit in de wetenschap – Een meetbaar begrip?*, DSWO Press, Leiden, p. 311-320.
- Knorr, K.D. et al. (1979). Individual publication productivity as a social position effect in academic and industrial research units. In: Andrews (ed.), F.M., *Scientific Productivity – The effectiveness of research groups in six countries*. Cambridge University Press/Unesco, Londen/New York/Melbourne/Paris.
- Lawshe, C.H. & R.E. Schucker (1959). The relative efficiency of four test weighting methods in multiple prediction. *Educational and psychological measurement*, 6, p. 376-383.
- Loon, P.J.J.M. van, F.N. Stokman, D. van der Hucht & P. Moorers (1988). Sterkte/zwakteanalyse sociologie, politicologie en bestuurskunde. *Acta Politica*, jrg. 23, 1, p. 21-60.
- Moed, H.F. (1985). (On)mogelijkheden van het gebruik van citatieanalyse in het universitaire onderzoeksbeleid. In: H.A. Becker & A.F.J. van Raan (red.), *Kwaliteit in de wetenschap – Een meetbaar begrip?*, DSWO Press, Leiden, p. 277-293.
- Moed, H.F. & A.F.J. van Raan (1987). Indicatoren van het wetenschappelijk presteren: Toepassing in het universitaire onderzoeksbeleid. In: Vall, M. van der & F.L. Leeuw (red.), *Sociaal Beleidsonderzoek: Differentiatie en ontwikkeling*. 's-Gravenhage, p. 187-203.

- Nederhof, A.J. (1988). *Metingen van kwaliteit van onderzoek van vakgroepen in de antropologie, bestuurskunde en psychologische functieleer: een empirische benadering*. Paper workshop scientometrie van de NSAV op 25 november 1988 te Amsterdam.
- Oostende, P. van (1990). De eerste nieuwe Intermediair Economen top-40. *Intermediair*, jrg. 26, 51.
- Peritz, B.C. (1983). Are methodological papers more cited than theoretical or empirical ones? The case of sociology? *Scientometrics*, vol. 5, no. 4, p. 211-218.
- Pijl, M. & S.J. Nijse (1979). Een empirisch onderzoek naar de bruikbaarheid van vereenvoudigde gewichten in multiple regressieanalyses. *MDN*, 2, 3, p. 99-107.
- Schmidt, F.L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and psychological measurement*, 31, p. 699-714.
- Spiegel, E. van (1985). Wetenschap: Beoordeling en beïnvloeding. In: Becker, H.A. & A.F.J. van Raan (red.), *Kwaliteit in de wetenschap – Een meetbaar begrip?* DSWO Press, Leiden, p. 13-18.
- Stokman, F.N., Popping, R. & E.R. Missoorten, (1988). *Een Relatieve Database voor Wetenschappelijke Productie*. Paper workshop scientometrie van de NSAV op 25 november 1988 te Amsterdam.
- Wainer, H. (1976). Estimating coefficients in linear models: it don't make no nevermind. *Psychological Bulletin*, vol. 83, no. 2, p. 213-217.

Bijlage 1. Categorisering van tijdschriften

Wetenschappelijk gerenommeerde tijdschriften (weegfactor 4)

- alle tijdschriften die door de SSCI worden gedekt en:
- Acta politica
- Mens en maatschappij
- Acta psychologica
- Kennis en methode
- Sociale wetenschappen
- Sociologische gids

Tijdschriften met wetenschappelijke status (weegfactor 3)

- Amsterdams sociologisch tijdschrift
- Antropologische verkenningen
- Beleid en maatschappij
- Economische statistische berichten
- Internationale spectator
- Jeugd en samenleving
- Justitiële verkenningen
- Ned. tijdschrift voor psychologie en haar grensgebieden
- Sociaal maandblad arbeid
- Sociologisch tijdschrift

Tijdschriften zonder wetenschappelijke status (weegfactor 1)

Alle overige tijdschriften

Bijlage 2. Categorisering van Nederlandse uitgevers

Wetenschappelijk gerenommeerde uitgevers (weegfactor 4)

- Kluwer
- Martinus Nijhoff
- Samsom
- Swets & Zeitlinger
- Van Gorcum & Comp
- Wolters Noordhoff
- Boom
- Elsevier
- Reidel
- Stenfert Kroese
- Van Loghum Slaterus

Uitgevers met wetenschappelijke status (weegfactor 3)

- Ambo
- Aula/Het Spectrum
- Couthino
- Dekker en Van der Vecht
- De Walburg Pers
- EJ Brill
- Het Spectrum
- Meulenhoff
- Noord-Hollandse Uitgeversmaatschappij
- Staatsdrukkerij en -uitgeverij
- Stafleu
- SUA
- SUN
- Tilburg University Press
- VUGA

Uitgevers zonder wetenschappelijke status (weegfactor 1)

Alle overige Nederlandse uitgevers