

Ins & outs van benchmarking

A.F.A.Korsten
Aug. 2006

1 Inleiding

Wat benchmarking is

Benchmarking is organisatievergelijking ten aanzien van prestaties en/effecten en/of activiteiten in de vorm van het aanwijzen van best practices, die gericht op verbetering door leren. Benchmarking en verbetermanagement zijn dus gekoppeld.

Benchmarking bij de overheid is in de loop der jaren sterk bepleit. Een voorbeeld. De *commissie- Van Rijn* publiceerde begin 2001 een rapport onder de titel *Investeren in mensen en kwaliteit – De arbeidsmarkt in de collectieve sector*. Daarin pleit de commissie voor meer prestatieafspraken en meer in het algemeen sturen op prestaties, meer inzet van ICT en een beter personeelsbeleid. De commissie meent dat benchmarking bij een *beter management* past.

In de commentaren op het rapport is de wenselijkheid van benchmarking, zelden expliciet bestreden. Gezien het belang van het rapport in de rijksdienst, in 2001 en later, kan het onderwerp benchmarking als *een actueel thema* in de rijksdienst gerekend worden. Die actualiteit geldt ook nog in 2007.

Voor adviseurs is benchmarking zelfs 'handel'. Door te laten zien dat je benchmarking-studies verricht, kan een bureau een moderne indruk maken. Zo kan men wellicht zelfs een concurrentievoorsprong krijgen. De vraag is echter of wat onder de vlag van benchmarking wordt aangeboden ook steeds als *echte* benchmarking moet worden beschouwd en of benchmarking wel zo probleemloos is. Is het wat het schijnt? Belangrijke tijdschriften waarin de discussie over benchmarking zich voltrekt zijn bijvoorbeeld: MAB – maandblad voor Accountancy en Bedrijfseconomie; International Journal of Public Sector Management; Public Administration; Public Administration Review; Public Productivity and Management Review; Accounting, Auditing and Accountability Journal; Managerial Auditing Journal.

Doel

Deze tekst heeft tot doel enig inzicht te verschaffen over wat benchmarking is en niet is, welke valkuilen daarbij optreden, welke functies te onderkennen zijn en wat de grenzen zijn. We geven een algemeen verkennend overzicht en illustreren enkele vraagstukken van benchmarking in de sector riagg's (centra voor geestelijke gezondheidszorg) en verpleeghuizen. Eerst gaan we in op het governance-beeld (lees

bestuursbeeld) achter benchmarking. Blijken zal dat bepaalde vergelijkingen wel nuttig zijn maar niet als benchmarking betiteld mogen worden. En benchmarking is vooral vanuit een perspectief van de 'learning organization' zinvol. Dan moet er wel heel veel oog bestaan voor de meetproblemen en de context.

Indeling

De indeling is als volgt.

- New public management en benchmarking
- Benchmarking: het basisonderwijs als voorbeeld
- Benchmarking: eerste brede verkenning
- Thuiszorginstellingen: mogelijkheden en grenzen van benchmarking
- Progressie in benchmarking: verpleeghuizen als domein
- Benchmarking: Riagg's
- Wat levert de Algemene Rekenkamer aan inzicht over benchmarking
- Bedrijfsvergelijking bij andere organisaties
- Sturing op prestaties en hoe perverse effecten te voorkomen.

2 New public management en benchmarking

Benchmarking is

- *organisatievergelijking* gericht op het vergelijken van werkwijzen en/of prestaties en/of effecten,
- om zo te komen tot een *best practice* en
- *verbetering*.

Denk daarbij aan het vergelijken van prestaties van politiekorpsen, scholen, sociale diensten, ziekenhuizen, gemeentelijke dienstverlening, waterschappen. Deze vergelijking kan bijvoorbeeld duidelijk maken hoeveel opgehelderde misdrijven er in een periode van vijf jaren waren in corps k, x en y. Dat kan vervolgens aanleiding zijn tot een 'best practice'-beeld en vragen waarom het aantal ophelderingen zo verschilt en wat de oorzaak van een gering of een groot aantal ophelderingen is.

Bestuursbeeld

Achter benchmarking gaat een bestuursbeeld van het New Public Management-denken (NPM) schuil. Het governance-beeld achter benchmarking is dat van de *bedrijfsmatige overheid*. Dat NPM-beeld is een eenzijdig beeld. Men kan het als metafoor uitproberen om zo eens van perspectief te wisselen. Dat kan tot inzichten leiden maar verabsoluteren van dit normatief beeld lijkt niet juist. Een politiek bestuur heeft vaak met *andere beoordelingscriteria* te maken dan alleen of primair met bedrijfsmatigheid, bijvoorbeeld met vrijheid, gelijkheid, sociale rechtvaardigheid, duurzaamheid, enz.

'New Public Management seeks to liberate bureaucratic managers by freeing them from the rules and restrictions that government managers face. (...) The liberation version of NPM is often accompanied by benchmarking and performance

standards with the idea that public managers are free to experiment but will be held by a bottom line'.

The Oxford Handbook of Public Management, Oxford UP, 2005: 56.

Vormen

Benchmarking valt vanuit *verschillende doelen* te initiëren en bestaat - mede daardoor - in verschillende vormen. We onderscheiden

- vrijwillige en verplichte benchmarking;
- benchmarking gericht op transparantie, leren of op verantwoording;
- horizontale en verticale, interbestuurlijke benchmarking.

Benchmarking gericht op leren is vaak horizontaal en komt op vrijwillige basis tot stand. De rijksoverheid dacht in 2002 over de introductie van verticale, interbestuurlijke benchmarking gericht op bevordering van verantwoording, waarbij deelname verplicht zou zijn.

Perverteringsgevaar

Benchmarking in de vorm van prestatievergelijking gaat gepaard met discussie over indicatoren van prestaties en perverteringsgevaar (De Bruijn, 2002). Dat gevaar - waarop verderop nog uitvoerig wordt ingegaan - is niet zo groot als vijf organisaties op basis van vrijwilligheid een benchmarkstudie laten verrichten maar veel groter indien benchmarking een interbestuurlijk verantwoordingsmiddel wordt en een toezichthouder oordelen uitspreekt. Er bestaan strategieën om perverse effecten te neutraliseren.

Toelichting op de kwestie van het achterliggend bestuursbeeld

De stroming in het denken over openbaar bestuur, die aandacht vraagt voor een vergelijking van de private en publieke sector en invoering van 'private' manieren van werken bij de overheid en semi-overheid, noemen we *New public management* (NPM). Daarin wordt dus *bedrijfsmatig werken* bepleit. Het gedachtegoed van NPM is in de loop der jaren verder ontwikkeld door auteurs als Osborne & Gaebler en Hood. De overheid zou volgens NPM-denkers meer op output (prestaties) of zelfs outcomes (effecten) moeten sturen, meer klantgerichtheid moeten nastreven, meer intern moeten verzelfstandigen en privatisering nastreven. Marktwerking en concurrentie tussen organisaties bevorderen, past dus bij NPM. Meer nadruk leggen op welke doelen een overheid wil bereiken, en zo mogelijk deze doelen (of doelstelling) in meetbare termen formuleren, is dus iets dat hierbij aansluit. Een voorbeeld: concurrentie tussen ziekenhuizen bevorderen, past binnen NPM-filosofie. Maar dan moet een burger kunnen kiezen tussen ziekenhuizen, moet bekend zijn welk ziekenhuis waarin beter of slechter is, moet goed en slecht duidelijk omschreven en gemeten zijn, en moet een transparant beeld ontstaan over de kwaliteit van een ziekenhuis. Tot 2006 werd daar niet aan voldaan maar intussen is er wel een nieuw zorgstelsel gekomen. Waarom? Burgers kunnen niet kiezen en het beeld van het beste ziekenhuis bestaat niet in voor burgers zichtbare aanduidingen en kwaliteiten,

hoezeer Elseviers Magazine ook anders wil doen geloven met titels op de omslag als 'de beste ziekenhuizen'.

Wie dit leest, zal ervan overtuigd kunnen zijn dat bepaalde middelen van nut zijn om te zien of er ook voortgang en zo mogelijk vooruitgang geboekt wordt op het vlak van doelbereiking, decentralisatie, concurrentie. Een van die middelen is het toepassen van een balanced scorecard-aanpak, en ander het houden van audits en visitaties, en weer een ander benchmarking. Bij diverse van deze en andere methoden is de vraag of ze geschikt zijn om breed toe te passen in de sfeer van het openbaar bestuur. De ervaring is dat toepassing op het eerste gezicht interessant is, en daarom ook bepleit wordt (zie het rapport van de commissie-Van Rijn) maar ook grote problemen geeft en veel discussie oproept. Bij voorbeeld, een balanced score-aanpak blijkt evidente problemen te geven bij toepassing in gemeentelijke sectoren. Begrijpelijk want deze aanpak komt uit de hoek van de bedrijfseconomie en bij een gemeenbestuur telt meer dan doelmatigheid. Dat neemt niet weg dat men 'voor een moment' - als metafoor - het transport van methoden uit de private sfeer naar de publieke sfeer kan uitproberen om te zien wat mogelijk is en wat de grenzen zijn. Als maar in het achterhoofd gehouden worden: de overheid is uiteindelijk geen bedrijf gericht op winst (zoals Tjeenk Willink, vice-president van de Raad van State, onlangs ook nog eens verklaarde). Om met Graham Allison te spreken: de overeenkomst tussen overheid en privaat bedrijf is er maar zit hem meer in de kleine dan in de grote 'dingen' (zoals doelstellingen)!

Toelichting op het gebruik van indicatoren

Lees het volgende stuk uit Binnenlands Bestuur van 7 juni 2002.

Uit Binnenland bestuur 7 juni 2002

Benchmarking

Arno Korsten

Zowel in de rijksdienst als binnen andere overheden geldt benchmarking, dat wil zeggen organisatievergelijking, tegenwoordig als een uitdaging. Pleidooien voor vergelijking komen voort uit een bedrijfsmatig denken. New public management, voor insiders NPM, is de theorie die het bedrijfsmatig denken voedt en daarmee ook benchmarking. Volgens de aanhangers van NPM is outputsturing noodzaak, dus sturing op resultaat. Wie een grotere verkeersveiligheid wil, zorgt dat er drempels komen want dan stuur je op resultaat. Nog beter is het om te bezien of je het aantal slachtoffers verlaagt. Dat heet effectsturing. Een resultaat is dus eigenlijk niet genoeg. Effectsturing, daar gaat het om. Door benchmarking kun je de kwaliteit van werkwijzen, prestaties of effecten vergelijken. Hiervoor dienen meer en meer verdedigers zich aan. Een dezer dagen las ik een pleidooi van een ziekenhuisdirecteur. Die topmanager was het te doen om vergelijking van de ziekenhuisproductie. Door te laten zien hoe lang de wachtlijsten elders en in zijn

eigen organisatie waren, kon hij tonen hoe goed of matig de eigen organisatie was en waar verbetering een optie was. Natuurlijk, er waren nog lange wachtlijsten, zei hij, maar ze waren al korter dan een jaar daarvoor.

Het vergelijken beperkt zich intussen niet tot ziekenhuizen. Met een *kwaliteitskaart* kun je resultaten van zwarte en witte scholen beoordelen. En we kennen analyses van verpleeghuizen, regionale instituten voor geestelijke gezondheidszorg en van sociale diensten. Dat oogt sympathiek. Wie wil niet weten welke de beste en de slechtste is. Vergelijking biedt een stap naar verbetering, dus voor leren. Dat geldt ook voor politiecorpsen. Ik wil best weten hoe het zit met de aangiften van misdrijven en het percentage ophelderingen. Er schijnen teveel politiemensen 'politieman gevulde koek' te heten. Ze vertoeven te lang op kantoor in plaats van dat ze op boevenjacht zijn. Kortom, benchmarking is breed aan de orde en dan laat ook de rijksoverheid zich niet onbetuigd. De Raad voor het Openbaar Bestuur is gevraagd om hierover een advies uit te brengen en vorige week werd op een ministerie van Binnenlandse Zaken al een expertmeeting gehouden. Een goed moment om eens te zien wat de deskundigen te melden hadden. Allereerst dit, deze deskundigen twijfelen niet aan de zinvolheid van organisatievergelijking. Vooral dan kan benchmarking zin hebben *als* organisaties met de vergelijking een leerdoel nastreven en er vrijwillig aan beginnen. Vergelijking blijkt juist dan te kunnen leiden tot verbeteracties. Dat dacht tenminste bijna iedereen hoewel een onderzoek dat de gewenste effecten van benchmarking bewees, nog ontbreekt. De verwachtingen zijn intussen positief. Als blijkt dat de ene sociale dienst gemiddeld weken sneller een uitkering verstrekt dan een andere, dan ligt er een uitdaging tot verbetering. Hoewel, je moet dat verschil wel verklaren. Een langere verstrekkingduur kan namelijk rechtvaardiging vinden in een andere politiek. De tragere sociale dienst biedt misschien meer – politiek gerechtvaardigd - bijstandsmatig werk. Het is dus opletten geblazen. Hoe dan ook, de benchmarkers die de valkuilen weten te omzeilen, kunnen dienstverlening transparanter maken. En dat is, zeker na de verkiezingen, een groot goed. Openheid moet, zegt bijna iedereen. Toch maak ik een kanttekening,. Er zijn namelijk twee soorten benchmarking. Met de vrijwillige benchmarking, waarover ik het had, is niks mis. Maar er zijn ook de interbestuurlijke benchmarkings. Dat zijn stelsels waarbij een hogere overheid met uitvoerende organisaties prestaties afsprekt en vervolgens medewerking aan verplichte benchmarking oplegt. Want wie vergelijkt, zal ook willen straffen of belonen. Als dat gebeurt, gaan partijen zich tactisch gedragen zoals in het hoger onderwijs met rendementen van studentenaantallen – wat uitliep op fraude - al bleek. De kans is aanwezig dat prestatiesystemen perverteren doordat organisaties gaan produceren wat goed scoort maar in plaats van waarmee burgers gediend zijn. Als een Riagg contacten van therapeuten met cliënten gaat tellen, verknippen therapeuten hun ene afspraak gewoon tot twee of drie contacten en tillen zo het systeem (in de wandelgangen geheten: 'fuck the system'). Een klassiek voorbeeld zijn de spijkerproducerende fabrieken in het Rusland van decennia terug. Deze fabrieken werden afgerekend op veel kilo's en daarom produceerden ze grote spijkers. Maar

daaraan was geen behoefte. Ze produceerden geen gedifferentieerd aanbod omdat dat niet scoorde terwijl de burgers juist behoefte had aan een assortiment. Het is dus oppassen geblazen met benchmarking.

Ik ben benieuwd wat er in 2002 uit de informatieonderhandelingen rolt (tussen CDA, VVD en LPF), want Pim Fortuyn was voorstander van zowel minder bureaucratie als prestatiecontracten. Fortuyn was niet van plan te breken met de opkomende tendens onder Paars om prestatieafspraken te maken maar wilde die juist versterken. Dat oogt goed maar aan pervertering en bureaucratisch rompslomp is geen behoefte. De kunst zal zijn om perversiteit en bureaucratie in prestatiesystemen te voorkomen.

3 Benchmarking: het basisonderwijs als voorbeeld

Benchmarking geschiedt *niet* op macro- en niet op microniveau maar *wel* op mesoniveau. Het gaat niet om de beste prestaties of effecten van ziekenhuizen in Nederland in vergelijking met dertig jaar terug (macroperspectief) noch om de beste prestaties van een individuele arts (microperspectief) maar om de prestaties van een ziekenhuis of groepje ziekenhuizen (meso). Denk ook aan benchmarking van de beste gevangenis, het beste museum, de beste sociale dienst, de beste gemeente, de beste basisschool, de beste universiteit. Het kan gaan om prestaties van leerlingen van een school, maar eventueel ook om het beste management (leiding, communicatie, administratie, enz.). Steeds is de entiteit meer dan een individu of een kleine afdeling. Er is sprake van een *zeker aggregatieniveau*.

Benchmarking betekent

- het vergelijken van werkwijzen en/of prestaties en/of effecten van organisaties (en eventueel relaties hiertussen) of subunits,
- gericht op continue verbetering,
- best practice van een organisatie (een school, een cluster opleidingen, een ziekenhuis, een verpleeghuis, een riagg).

Deze benchmark-aanpak, met drie kenmerken, komt uit de private sfeer. Het gaat niet zo maar om vergelijken want vergelijken komt heel vaak voor. Dan zou er geen nieuws onder de zon zijn. Benchmarking is vergelijken *ten dienste van* continue verbetering *op basis van* het verkrijgen van een beeld van 'de' beste praktijk (best practice). Benchmarking in het basisonderwijs zou dus een beeld moeten geven van de beste basisschool. Daarmee komen we direct bij een probleem.

Meetvraagstukken

Benchmarking is mogelijk als meetbaar en onomstreden meetbaar is wat de beste basisschool is. Dat is een *normatieve kwestie* die ons leidt naar de afhankelijke variabele. Is de beste school de school met de beste leerlingen, of de beste leraren, of de beste schoolorganisatie, of het beste strategisch beleid, of de beste communicatie

tussen alle betrokkenen, of die het beste de inspectie ontvangt, of een mix van zwarte en witte leerlingen heeft (multiculturele mix),? En wat is een leerlingprestatie: een gemiddelde score op een Citotoets? Wordt niet naar de spreiding rond het gemiddelde gekeken en de samenstelling van de leerlingpopulatie, bijvoorbeeld gemeten naar het startniveau van de leerlingen? En stel dat we zeggen dat de beste basisschool die school is die de beste leerlingprestaties heeft, dan hebben we een nieuw probleem, want als er *zelfselectie* optreedt doordat gemiddeld goede leerlingen op de school 'afkomen' dan is het onderwijs misschien wel matig tot gemiddeld maar niet excellent, terwijl de school wel de beste school is van een categorie vergeleken scholen (zeg scholen in provincie X).

Kwaliteitskaarten

Er ontstaat ook een ander probleem: moet er niet vergeleken worden op *andere variabelen* dan leerlingprestatie? En is er een mix van variabelen nodig? De discussie over *kwaliteitskaarten in het onderwijs*, zoals gevoerd in het dagblad Trouw en daarbuiten, illustreert vraagstukken van benchmarking. Enfin, deze discussie zetten we hier niet tot het einde door. Als een burger voor een kind die een selectie van een school wil maken, want daartoe dient benchmarking in Groot-Brittannie, dan zal de burger inzicht willen hebben in wat de kwaliteit van ene school verklaart. Waarom is een school de beste school? Dat zou dan op een kwaliteitskaart moeten staan. Is oorzakelijkheid echter zo eenvoudig in enkele getallen te vangen?

Het voorbeeld illustreert dat benchmarking

- normatieve vragen oproept (wat is de maat: de beste leerling, de beste leraar, de beste schoolorganisatie?; is coorrectie voor bepaalde variabelen nodig),
- vragen van bereik oproept (wat valt wel of niet te vergelijken),
- domeinvragen genereert (kan benchmarking alleen worden toegepast waar sprake is van duidelijk omschreven standaardproductie?)
- een vertekeninggevaar geeft (door alleen te kijken naar wat meetbaar is).

Goede condities voor benchmarking

Deze gevaren zijn het geringste als enkele organisaties op basis van vrijwilligheid vanuit eigen doelstellingen en overwegingen, bijvoorbeeld de wil een lerende organisatie te zijn, overgaan tot benchmarking en zelf de condities daarvoor scheppen. Bijvoorbeeld vijf Riagg's besluiten samen tot een benchmark van hun organisaties. Zo gauw sprake is van een stelsel van benchmarking, als rond kwaliteitskaarten, ontstaat discussie over de sturing op grond van die kaarten en discussie een eenzijdige normativiteit achter benchmarking van scholen.

4 Benchmarking: eerste brede verkenning

Benchmarken impliceert het vergelijken van organisaties, of beter gesteld: het is de kunst van het vergelijken (Blank, 1998; Linssen, 1998). Benchmarking is volgens de literatuur *bedrijfsvergelijking* (BMI-Magazine, 1998, nr. 6: 4). Maar dat is geen wet van

Meden en Perzen. Men kan ook landen vergelijken zoals gebeurt bij de *concurrentietoets* die het ministerie van EZ opstelt.

Drie kenmerken

Benchmarking is *meer* dan vergelijking van organisaties want is ook gericht op verbetering van throughput (activiteiten/ productie) en/of output (prestaties) of outcomes (effecten). Benchmarking heeft tot doel dat het moet leiden tot best practices. Er zou dus in deze strenge definitie sprake moeten zijn van drie kenmerken:

- vergelijking tussen organisaties,
- gerichtheid op verbetering van organisaties en
- best practices.

Stel het gaat bij een benchmark om doelmatigheid van instellingen. De best practice is de meest doelmatige instelling. Bij doelmatigheid gaat het dan volgens Jos Blank, die we hier volgen, om de relatie tussen productie en ingezette middelen. Een instelling kan als doelmatig worden gekwalificeerd 'als er geen mogelijkheid is om met dezelfde middelen meer te produceren of eenzelfde productie met minder middelen' (1998: 432). Zo gezien geeft het geheel van doelmatige instellingen aan wat technisch en economisch haalbaar is. De doelmatige instelling vormt, zo gezien, het referentiekader of de benchmark voor instellingen die *niet* als doelmatig beoordeeld kunnen worden. De ondoelmatige instellingen kunnen zich spiegelen aan de doelmatige.

Wat voldoet niet aan de kenmerken

Volgens deze (strenge) omschrijving zijn onderwijsvisitaties bij universiteiten geen benchmarking. Er is bij een visitatie van zeg universitaire bestuurskunde-opleidingen volgens het VSNU-protocol sprake van vergelijking van opleiding k, l, m, n, o, p, enz. Een visitatiecommissie maakt een rangorde van betere en slechtere opleidingen door te wijzen op kwaliteiten op aspect 1, 2, 3, 4 enz. Denk aan rendement, studentsatisfactie, ondersteuning door personeel. Ze komt tot een intersubjectief oordeel. Maar wat de beste praktijk is, in termen van *een combinatie* van beste praktijken op onderdelen *samen* is, wordt *niet* aangegeven. Dat kan ook moeilijk omdat de opleidingen qua *vakinhouden* niet geheel gelijk zijn en een *eigen inbedding* in een universiteit en *missie* hebben.

Deze aanpak is overigens positief en kent een draagvlak, juist omdat er oog is voor verschillen in context. Er is voldaan aan twee van de drie kenmerken. Daarmee hebben we een belangrijk middel van kwaliteitszorg te pakken maar van het vaststellen van de best practice is in strikte zin geen sprake, van de beste opleiding. Toch gebeurt dat door Elsevier? Elsevier publiceert bij tijd en wijle per criterium wel een rangorde. Zo gauw een aantal criteria op een hoop geveegd worden, 'breekt de hel los'. Dan worden opleidingen *ontdaan van hun context*. Het is mede om die reden,

dus het bestaan van een eigen missie per opleiding, dat wel gepleit wordt voor
visitatie afgemeten aan de ambities van een opleiding en een basischecklist. De
vergelijking tussen opleidingen wordt dan losgelaten maar daarvoor komt
beoordeling op het waarmaken van de ambitie in de plaats.

Een dergelijke aanpak is ook gericht op kwaliteit en komt voor bij gemeenten in
Limburg, waar een bestuurskrachtmonitor wordt toegepast. Een
bestuurskrachtmonitor is een zelfevaluatie van de bestuurskracht van een gemeente,
op basis van een eigen ambitieformulering door de gemeenteraad en een checklist,
die wordt uitgewerkt door een visitatiecommissie, waarna het rapport wordt
vastgesteld door een gemeenteraad en visitatiecommissie.

Hype

Benchmarking is in de jaren 1994-1999 uitgegroeid tot een hype. Nieuw is
benchmarking niet want al voor WO II vonden dergelijke vergelijkingen in de
Verenigde Staten plaats. In 1959 werd door Henzel een bijdrage geleverd aan het
'*Handwörterbuch der Sozialwissenschaften*', waarin hij inging op bedrijfsvergelijking
onder de kop 'Betriebsvergleich'. Benchmarking is door Henzel destijds gezien als
een zinvolle verdere ontwikkeling van bedrijfsstatistiek.

Waar benchmarking uit voortkomt?

Benchmarking komt voort uit onzekerheid, meent Tom Groot, hoogleraar
management accounting aan de VU. Een gemeente of een bedrijf weet nooit zeker of
het het wel goed doet, of de organisatie wel doelmatig functioneert. Wijkt mijn
organisatie af van die van anderen? Ongunstig?

Typen benchmarking

Benchmarken op doelmatigheid van instellingen is een mogelijkheid naast
benchmarken op prestaties of iets anders. Dergelijke (doelmatigheids-)
benchmarking, *waarop we verderop ingaan*, gebeurt vaak door de kenmerken van het
productieproces te vergelijken (Blank, 1998: 432). Iets dergelijks is niet geheel nieuw
want in het verleden werden gemeenten van zelfde grootte wel vergeleken per sector
om zo te bezien of een gemeente een politiek gerechtvaardigde of discutabele
extreme positie innam.

Departementale interesse voor benchmarking

Benchmarking van doelmatigheid is een methode om de doelmatigheid van
instellingen te vergelijken en de instellingen de resultaten als spiegel voor te houden.
Door de ministeries van VWS en Financiën is het belang van benchmarking recent
onderkend. Een interdepartementale werkgroep onderzocht de mogelijkheden van
benchmarking in de AWBZ-sector: bij verpleeg- en verzorgingshuizen, thuiszorg en
instellingen voor gehandicapten.

Problemen bij benchmarking

Het beste is het om organisaties van zelfde grootte en functie met elkaar te vergelijken. Dan is men er nog niet. Een van de problemen bij benchmarking is dat er sprake moet zijn van *relevante* gegevens en van *vergelijkbare* gegevens.

Een geheel ander probleem is het gebruiken van benchmarking als excuus om een persoon die verantwoordelijk is voor een afdeling aan de dijk te zetten. Daarvoor is een bedrijfsvergelijking niet, meent prof.dr. Tom Groot (BMI Magazine, 1998, nr. 6). Benchmarking is geschikt om het functioneren van organisaties te verbeteren, *niet* om verantwoordelijkheid te herstellen. Daarvoor bestaan andere aanpakken.

Functies van benchmarking

In algemene zin worden wel de volgende functies van benchmarking onderscheiden.

- a Als er door benchmarking gesignaleerd wordt dat er ergens in het functioneren van een organisatie een achterstand bestaat, kan gepoogd worden die te verhelpen door op zoek te gaan naar andere werkmethoden, die in andere organisaties gebruikt worden.
- b Benchmarking levert zicht op de plaats die een organisatie in een sector of branche inneemt en dat kan een impuls zijn voor de organisatiecultuur ('we doen het goed') of tot verbeteractie leiden.
- c Benchmarking kan bijdragen aan de transparantie van ene organisatie en de verantwoording naar de samenleving.

De verwachting is dat op basis van een benchmark-analyse van doelmatigheid te besluiten is tot het monitoren van de doelmatigheid van voorzieningen: wel of geen voortgang meten. En er zouden door benchmarking meer doelmatigheidsprikkels in de financieringssystematiek kunnen komen. En de bedrijfsvoering kan eventueel worden aangepast. Dat zijn drie functies rond doelmatigheid. Is benchmarking anders gericht, dan komen ook andere functies in beeld.

Risico's van benchmarking

Het toepassen van benchmarking is niet zonder risico's. We noemen er twee: het streven naar best practice maar dan op een voor de betreffende organisatie *irrelevant* punt, en het streven naar het gemiddelde door *kopieergedrag* waardoor een streven naar excellentie uitblijft. We lichten dit toe.

a Een van de risico's is dat de instelling die meedoet aan benchmarking zichzelf wat wijs maakt. Dat gebeurt natuurlijk alleen als de kengetallen niet deugen. Het zegt niets als een instelling de 'best practice' heeft op een irrelevante variabele.

b Een ander risico is dat een organisatie niet de ambitie heeft 'voor de troepen' uit te gaan of blijven lopen maar dat de organisatie kopieergedrag gaat vertonen richting

het gemiddelde. Als we het maar gaan doen zoals de organisatie x, y en z 'dan zitten we goed', luidt de slappe redenering. Wie alleen maar rondkijkt en in de gaten houdt wat anderen doen, loopt nooit voorop (MBI Magazine, 1998, nr. 6).

Literatuur over sturing op prestaties

- Bogt, H. ter, Prestatiegegevens: nuttig voor sturing van gemeentelijke taken, in: Public Controlling, juni 2004, pp. 52-58.
- Groot, H. de, (red.), Benchmarken in de publieke sector, Sdu, Den Haag, 2004 (serie Public controlling, nr. 4).
- Hoogwout, M., Het ontwerpen van overheidsbenchmarks – Een krachtig management- en verantwoordingsinstrument, in: Public Controlling, juni 2004, pp. 32-42.
- Smit, B., T. Tadema-Van Dam, en D. van der Zwaag, Van beheren naar beheersen – Sturen op prestaties, financiën en bezetting bij de gemeente Leiden, in: Public Controlling, juni 2004, pp. 16-28.
- Strikwerda, J., Shared service centers en control in de publieke sector – Hoe verdere decentralisatie leidt tot betere control, in: Public Controlling, juni 2004, pp. 22-27.
- Strikwerda, J., Shared service centers: van kostenbesparing naar waardecreatie, Van Gorcum, Assen, 2003.
- Waal, A. de, en L. Kerklaan, De resultaatgerichte overheid – Op weg naar de prestatiegedreven overheidsorganisatie, Sdu, Den Haag, 2004.

5 Thuiszorginstellingen: mogelijkheden en grenzen van benchmarking

Ik ga nu kijken naar hoe het bij benchmarking toegaat en nemen een voorbeeld. Daarmee wil ik de grenzen van benchmarking illustreren.

Het ministerie van VWS en de Landelijke Vereniging voor Thuiszorg heeft Bakkenist Management Consultants een onderzoek laten doen naar de doelmatigheid. Dit mede met het oogmerk om de productiviteit te vergroten en suggesties te doen voor financieringsrichtlijnen. Op basis van een onderzoek in zeven instellingen voor thuiszorg, die geacht worden een representatieve steekproef te zijn, is de doelmatigheidspositie van thuiszorginstellingen bepaald. Hieruit bleek dat de doelmatigheidspositie van de thuiszorg *niet ongunstig* afstak ten opzichte van andere sectoren (Blank, 1998; Elias e.a., 1997). Tevens bleek dat de doelmatigheid wel kon worden verbeterd, zoals door resultaatgericht management, kwaliteitssystemen, flexibele arbeid, een professionele backoffice en andere wel in de literatuur gesuggereerde maatregelen.

Hoe kwamen de onderzoekers nu tot hun conclusies? Ze verzamelden *kengetallen* over productie, ingezette middelen, proceskenmerken en hun onderlinge verhoudingen. Er is onder meer bezien wat het aandeel uitvoerend en niet-uitvoerend personeel is, en wat het productieve en niet-productieve personeel. Zorguren van wijkverpleging en gezinsverzorging zijn in verband gebracht met het totaal aan personeel of met het budget. Niet-leidinggevenden zijn gerelateerd aan het zorgpersoneel. Dat levert verhoudingscijfers op. Voor *ieder* vergelijkbaar punt voor

de zeven instellingen is een 'best practice' aangeduid met *vet* in de tabel. Het kan een minimum- of een maximumkengetal zijn.

Wat is nu de doelmatige instelling?

De doelmatige instelling bestaat niet. Wel bestaat de *virtuele* doelmatige instelling, opgebouwd uit alle 'best practices'. Zie de volgende tabel.

Voor het aantal zorguren per voltijdse arbeidsplaats (fte) in de wijkverpleging geeft instelling II de hoogste score ('best practice') en voor de gezinsverzorging is het instelling V. Daarentegen geldt dat instelling I de best practice geeft voor het aantal zorguren per F 10.000,- budget.

Kijken we naar de personele samenstelling dan blijkt opnieuw: de ene instelling scoort beter op een variabele, en een andere instelling op een andere. Instelling I scoort het beste op de factor 'percentage uitvoerend personeel' terwijl instelling V het beste uit de bus komt op de factor 'leiding per 100 fte verzorgend personeel'.

Tabel: Kenmerken van het productieproces bij zeven onderzochte instellingen voor thuiszorg vlg. Elias e.a. (1997)

instellingen	I	II	III	IV	V	VI	VII
zorguren wijkverpleging per fte uitvoering wijkverpleging	966	1075	960	nb	726	865	954
zorguren gezinsverzorging per fte uitvoering gezinsverzorging	1421	1401	1434	nb	1490	1449	1477
zorguren per F 10.000 budget	261	212	202	225	225	213	237
aandeel uitvoerend personeel (%)	93	88	92	91	89	91	91
waarvan productief (%)	80	75	71	81	76	77	78
leiding per 100 fte verzorgend personeel	0.92	1.06	0.67	0.85	0.39	0.91	0.56
p&o-personeel per 100 fte verzorgend personeel	0.66	1.48	2.19	1.79	1.26	1.35	1.57
fin. econ. personeel per 100 fte verz. personeel	1.47	2.66	2.67	2.37	2.50	2.92	2.91
automatiseringspersoneel per 100 fte verz. personeel	0.76	0.97	0.84	0.52	0.76	0.65	0.71
facilitair personeel per 100 fte verzorgend personeel	1.34	3.61	3.52	2.67	2.07	3.19	1.99
ziekteverzuim (%)	11	7	11	12	11	9	11
aandeel variabel personeel (%)	15	2	3	19	15	4	8
parttime factor (%)	46	57	48	44	58	42	61

vet: 'best practice'; nb= niet beschikbaar;
bron: Elias e.a., 1997.

Wat weten we nu?

Ten eerste kan men de vraag stellen of de variabelen goed gekozen zijn. Is de doelmatigheid zeer gediend met een gering percentage niet-uitvoerend personeel, bij wijze van spreken met een kleine directie? En is het inderdaad uit een oogpunt van doelmatigheid zonder meer gunstig dat de verhouding tussen facilitair en uitvoerend personeel zo minimaal mogelijk is. Is het meest minimaal, zoals bij instelling I, niet te minimaal? Als er weinig faciliteiten zijn, bestaat de mogelijkheid dat uitvoerend personeel te veel zelf moet doen en niet aan zijn eigenlijke taak

toekomst, aldus Blank (1998). Het is nu wel duidelijk: benchmarking impliceert normatieve keuzen van kengetallen.

De les is: *het is steeds nodig bij benchmarking om de keuze van variabelen te verantwoorden!*

Grenzen van benchmarking

Tot de grenzen van benchmarking behoren:

1. Een eerste grens, zo men wil dreigende beperking, van benchmarking is *het achterwege blijven van de verantwoording* van de variabelenkeuze. Dat kan tot gevolg hebben dat een instelling denkt positief te scoren op een kengetal en daarmee een 'best practice' 'in handen te hebben' terwijl dat in werkelijkheid 'fake' is omdat het kengetal niet wijst op wat men wil meten. De variabele is niet geldig.
2. Een tweede grens houdt in dat de variabele waarop scores vergeleken worden wel *relevant* is. In het veld wordt hoog scoren op zo'n variabele als goed beschouwd, maar waarom eigenlijk? Een voorbeeld. Het wordt als 'goed' beschouwd om veel variabel personeel in dienst te hebben in de thuiszorg en weinig parttimers. Waarom is dat zo goed? Variabele personeel beschikt wellicht over minder 'verzorgende kwaliteiten' dan de parttimers. Onderzoek toont dat parttimers juist de doelmatigheid, althans in ziekenhuizen, bevorderen (Blank, 1998: 433).
3. Een derde grens van benchmarking is dat benchmarking niets zegt over de *oorzaak of achtergronden van verschillen* tussen de scores van organisaties op een kengetal. Of schaafeffecten een rol spelen blijft duister. De grootte van instellingen kan een relevante factor zijn omdat bij grote organisaties de overhead in verhouding tot kleine organisaties waarschijnlijk procentueel geringer is. De *missie van de organisatie* en de *werkfilosofie* wordt ook niet achter de cijfers zichtbaar.
4. Een vierde grens impliceert dat de benchmarking niet, bij een gegeven productie, op zoek gaat naar welke organisaties 'van vlees en bloed' *verschillen in de manier waarop ze die productie realiseren*. Zou dat wel gebeuren dan komen er aanwijzingen naar voren of de ene organisatie dat resultaat doelmatiger bereikt dan het andere. Aan dit bezwaar wordt door het SCP getracht tegemoet te komen.

6 Progressie in benchmarking: verpleeghuizen als domein

Ik ga nu in op verbetering van de benchmark-aanpak en kies als invalshoek doelmatigheid. Wat is qua doelmatigheid (throughput) het beste verpleeghuis?

De aandacht voor doelmatigheid van dienstverlening in de jaren negentig toegenomen is. Een grotere doelmatigheid kan voordelen hebben, zoals kortere wachtlijsten, meer blauw op straat, enz. Het meten en spiegelen van instellingen aan elkaar die eenzelfde product of dienst leveren, is daarbij een hulpmiddel gebleken. Dat vergelijken heet benchmarken. Te onderscheiden is in zinloos en zinvol vergelijken. Van zinloos vergelijken hebben we geen voorbeelden gegeven. Wel gaf ik aan dat benchmarking door *het opsporen van een virtuele, doelmatige organisatie niet genoeg is*. Immers, die *bestaat niet*. De kunst is:

om juist bij eenzelfde hoge productie de verschillen te verklaren. Dat levert gegevens op over soorten ondoelmatigheid en aanzetten (nog geen actieprogramma) over hoe de doelmatigheid te vergroten.

Progressie in benchmarking?

Het Sociaal en Cultureel Planbureau (SCP) denkt progressie geboekt te hebben in benchmarking. Waarom?

Het SCP scoort niet slechts de *best practice* per kengetal voor een virtuele organisatie. Het SCP poogt daadwerkelijk een best practice van een *bestaande* instelling te vinden, een gunstig patroon van productie dus, en deze te verklaren. De onderzoekers spreken van de *data-omhullingsmethode* (Blank, 1998; Blank en Eggink, 1998). Hierbij worden instellingen met eenzelfde productie met elkaar vergeleken. De afhankelijke variabele wordt dus gelijk gehouden.

Drie vormen van ondoelmatigheid

Onderzoek met deze aanpak bij verpleeghuizen laat zien dat dezelfde productie met 30 procent minder middelen gerealiseerd kan worden. Er is blijkbaar inefficiëntie. Daarvan zijn drie vormen bekend: allocatieve, technische en schaal (in) efficiëntie.

1. Die ondoelmatigheid kan deels worden toegeschreven aan een verkeerde 'samenstelling van de ingezette middelen'; *allocatieve inefficiëntie* heet dat (Blank, 1998: 435).
2. Die inefficiëntie kan deels óók worden toegeschreven aan *technische inefficiëntie*: 'het niet maximaal benutten van de ingezette middelen' (Blank, 1998: 435). Door een betere technische efficiëntie zou 12 procent efficiëntievoordeel te behalen zijn.
3. Met schaalverandering en daarmee meer *schaalefficiëntie* is 6 procent kostenbesparing te realiseren.

Volgens de SCP-benadering kan met een regressie-analyse voor elk proceskenmerk worden nagegaan wat de bijdrage is aan de efficiëntie van de instelling. Wat beïnvloedt de technische efficiëntie van een verpleeghuis nadelig? Het SCP-onderzoek 'Zuinig op zorg' toont dat een hoog aandeel van leerling-

verpleegkundigen, een hoog ziekteverzuim, veel in- en uitbested werk en grote afdelingen de technische inefficiëntie op een negatieve manier beïnvloeden. Een positief effect op de doelmatigheid gaat uit van hooggekwalificeerd personeel en een hoge bezettingsgraad (Blank en Eggink, 1996).

Tabel: Efficiënties van verpleeghuizen

totale efficiëntie	0.70
waarvan: allocatieve efficiëntie	0.85
waarvan: schaafefficiëntie	0.94
waarvan: technische efficiëntie	0.88
bijdrage kenmerken productieproces aan technische inefficiëntie	
aandeel leerlingverpleegkundigen	-0.12
kwaliteitsprijs gediplomeerde verpleegkundigen	0.22
kwaliteitsprijs overig personeel	0.20
eigen apotheek	0.01
bezettingsgraad	0.30
ziektervuim	-0.05
uitbested werk	-0.05
inbested werk	-0.03
afdelingsgrootte	-0.04
bron: SCP, Zuinig op zorg, 1996; Blank, 1998	

Daarmee kom ik bij een volgende conclusie:

als het nodig is om bij benchmarking zo te werk te gaan, wordt benchmarking werk van superspecialisten op het gebied van methoden en technieken van sociaal-wetenschappelijk onderzoek.

7 Benchmarking: Riagg's

Ik kom nu bij een ander voorbeeld, dat ook weer de weerbarstige praktijk van benchmarking illustreert. Een levensecht voorbeeld waarmee ik zelf geconfronteerd werd als lid van de Raad van Toezicht van de Riagg OZL (een instelling die intussen gefuseerd is).

De *Commissie Doelmatigheid Riagg-activiteiten* heeft in opdracht van de minister van VWS tussen september 1997 en mei 1998 onderzoek gedaan naar de doelmatigheid van de (organisatie van de) uitvoering van Riagg-activiteiten. Als de commissie is ingesteld door de minister zal er sprake zijn van een *instellingsbeschikking*. Die vinden we in de Staatscourant van 1 oktober 1997.

De taakopdracht van de *commissie-Brouwer* luidde als volgt:

1. onderzoek doen naar de doelmatigheid van de verschillende organisatievormen waarbinnen Riagg-activiteiten worden uitgevoerd;
2. inzicht verkrijgen in de factoren die de doelmatigheid beïnvloeden alsmede inzicht bieden in de mogelijkheden om de doelmatigheid te verbeteren.

Hierover moest gerapporteerd worden aan de minister.

Achtergrondinformatie

Instelling Onderzoekscommissie Doelmatigheid Riagg-activiteiten

21 sept. 1997 nr. GVM/GGZ 974948

De minister van Volksgezondheid, Welzijn en Sport
besluit

§1. *Definities*

Artikel 1

In dit besluit wordt verstaan onder

- a de minister: de minister van VWS;
- b. de commissie: de commissie genoemd in artikel 2;
- c. Riagg-activiteiten: activiteiten zoals uitgevoerd door regionale instellingen voor ambulante geestelijke gezondheidszorg of samenwerkingsverbanden waarvan de Riagg onderdeel uitmaakt.

§ 2. *Instelling commissie*

Artikel 2

1 Er is een onderzoekscommissie Riagg-activiteiten.

2 De commissie heeft tot taak:

- a. onderzoek te doen naar de doelmatigheid van de verschillende organisatievormen van de uitvoering van Riagg-activiteiten;
- b. inzicht te verkrijgen in de factoren die de doelmatigheid beïnvloeden alsmede inzicht te bieden in mogelijkheden die doelmatigheid te verbeteren;
- c. van deze inzichten mededeling te doen aan de minister in een eindrapport.

Artikel 3

1 Tot lid van de commissie worden benoemd:

E. Brouwer RA, tevens voorzitter;

.....

2 De benoeming der leden geschiedt op grond van hun deskundigheid op voor de taak van de commissie relevante terreinen.

3 Als adviserend deskundige, bevoegd tot het bijwonen van de vergaderingen van de commissie, wordt aangewezen:, psychiater, Inspectie Gezondheidszorg.

4 Aan de commissie wordt als secretaris toegevoegd:

§3. *Werkwijze*

Artikel 4

1 De commissie stelt haar eigen werkwijze vast in de vorm van een activiteitenprogramma.

2 De commissie kan zich doen bijstaan door andere personen voor zover dat voor de vervulling van haar taak noodzakelijk is.

3 De commissie is bevoegd om zich voor het verkrijgen van informatie te wenden tot overheidsdiensten en instellingen en groeperingen buiten de overheid.

4 De commissie brengt uiterlijk 1 novembr 1997 een eerste rapportage uit, inhoudende het activiteitenprogramma van de commissie. In dit programma staat tevens het verdere tijdpad van de voorgestelde activiteiten van de commissie aangegeven.

5 De commissie brengt uiterlijk 1 mei 1998 haar eindrapport uit aan de minister.

§4. *Overige bepalingen*

Artikel 5

Het beheer van de bescheiden betreffende de werkzaamheden van de commissie geschiedt op overeenkomstige wijze als bij het ministerie van VWS. De bescheiden worden na beëindiging van de werkzaamheden van de commissie opgeborgen in het archief van het ministerie.

Artikel 6

1 Dit besluit treedt in werking met ingang van de tweede dag na de dagtekening in de Staatscourant waarin mededeling van dit besluit wordt gedaan en werkt terug tot en met 8 september 1997.

2 Dit besluit vervalt met ingang van 1 juli 1998.

Artikel 7

Dit besluit wordt aangehaald als: *Besluit Onderzoekscommissie doelmatigheid Riagg-activiteiten*.

Van dit besluit wordt mededeling gedaan in de Staatscourant. Afschrift van dit besluit zal worden gezonden aan de Algemene Rekenkamer.

De minister van Volksgezondheid, Welzijn en Sport
E. Borst-Eilers

Toelichting

Algemeen

In de notitie 'Wachlijsten in de curatieve zorg' die op 17 december 1996 aan de Tweede Kamer is aangeboden, wordt een tweesporenbeleid aangekondigd inzake een incidentele en structurele aanpak van de wachtlijsten in de curatieve zorg, waaronder de ambulante geestelijke gezondheidszorg (aggz). Het gaat hier om de incidentele aanpak van acute knelpunten in de wachtlijsten van de Riagg. Ook wordt in dat verband onderzocht welke voorwaarden noodzakelijk zijn voor een structurele aanpak van wachtlijsten in de gehele aggz. De structurele aanpak wordt de komende jaren verder uitgewerkt. Hiertoe is noodzakelijk op een aantal punten nadere inzichten te verkrijgen. Wachlijsten worden voor een deel veroorzaakt door een toenemende vraag. Voor een ander deel kunnen wachtlijsten echter ook worden veroorzaakt door ondoelmatigheid van de organisatie van de hulpverlening. Voor doelmatig werkende Riagg-en die voldoen aan een aantal criteria inzake productie, wachtlijstregistratie en -beleid, samenwerking en normoverschrijdende, meer dan gemiddelde wachtlijsten is incidenteel geld beschikbaar gesteld voor de aanpak van acute knelpunten in de wachtlijsten. In andere regio's waar sprake is van wachtlijsten is echter in ieder geval nader onderzoek noodzakelijk naar de achtergronden van het ontstaan van de wachtlijsten.

Onderzoekscommissie doelmatigheid Riagg-activiteiten

Om inzicht te krijgen in de doelmatigheid van de uitvoering van de Riagg-activiteiten door zowel zelfstandige Riagg-en als regionale samenwerkingsverbanden waarvan de Riagg deel uitmaakt, wordt de Onderzoekscommissie Doelmatigheid Riagg-activiteiten ingesteld. Deze commissie heeft tot taak te onderzoeken welke succes- en faalfactoren de doelmatigheid van de uitvoering van Riagg-activiteiten beïnvloeden. Inzicht in de factoren die de doelmatigheid van de Riagg-activiteiten beïnvloeden dragen bij aan de evaluatie van het bestaande beleid ten aanzien van de ambulante geestelijke gezondheidszorg en onderbouwen het te ontwikkelen beleid inzake een structurele aanpak van de aggz en de knelpunten daarin die bijdragen aan het ontstaan van wachtlijsten.

De taakomschrijving, zoals in artikel 2, tweede lid, omschreven zal in ieder geval omvatten:

- a) Onderzoek naar doelmatigheid van de organisatie van Riagg-activiteiten.
- b) Vergelijking van succes- en faalfactoren met betrekking tot doelmatigheid van de verschillende organisatievormen waarin Riagg-activiteiten plaatsvinden.

- c) Antwoord op de vraag of de wijze van samenwerking van de Riagg met andere (ggz-) aanbieders in de regio invloed heeft op de doelmatigheid van de Riagg-activiteiten.
- d) Onderzoek naar de mogelijkheden om de doelmatigheid van de Riagg-activiteiten en van de omgeving waarin deze plaatsvinden te vergroten.
- e) Antwoord op de vraag of de productie-afspraken een goede indicator is voor het meten van de feitelijke doelmatigheid van de Riagg-activiteiten.

De commissie is zodanig samengesteld dat de verschillende deskundigheden die noodzakelijk zijn voor de uitvoering van deze complexe taakopdracht er in zijn vertegenwoordigd.

De vergoeding van de leden, de secretaris en de deskundigen geschiedt op grond van het Vacatiegeldenbesluit 1988 en het Reisbesluit binnenland.

De commissie wordt voor de uitvoering van haar werkzaamheden eenmalig een budget toegekend.

De Minister van Volksgezondheid, Welzijn en Sport,
E. Borst-Eilers

Ter uitvoering van de taak heeft de commissie gesproken met mensen die werkzaam zijn bij een Riagg of in een samenwerkingsverband waar de Riagg deel van uitmaakt, participeren. Tevens is gesproken met enkele zorgverzekeraars, met vertegenwoordigers van de Centraal Orgaan Tarieven Gezondheidszorg (COTG) en de Ziekenfondsraad (ZFR). In de eindfase is een hoorzitting georganiseerd over de concept-aanbevelingen. Ook deze commissie heeft Bakkenist Management consultants ingeschakeld voor onderzoek van doelmatigheid van de zorg, van de bedrijfsvoering in zeven instellingen en van de klanttevredenheid bij cliënten van deze instellingen.

De commissie heeft de volgende omschrijving van doelmatigheid gehanteerd: *'de balans tussen kosten en baten (effect/kwaliteit) van een activiteit. Bij het zoeken naar de meest doelmatige zorg gaat het om de optimale balans tussen budget, en de hoeveelheid verleende zorg c.q. toegankelijkheid van de zorg, met behoud van de noodzakelijke kwaliteit. Met de (immers beperkte) middelen dienen zoveel mogelijk hulpvragen op verantwoorde wijze te worden beantwoord'.*

De commissie heeft geconstateerd dat de Riagg's een brede taakopdracht hebben. De instellingen zijn over het algemeen *kleinschalig* bezig. Het gevolg zou ondoelmatigheid zijn. De commissie geeft de minister ter overweging de taakopdracht van Riagg's, binnen het geheel van de ggz, te herzien. 'In het licht van de toenemende *fusiebewegingen* dient overwogen te worden de taken van de ggz te herschrijven in termen van *zorgprogramma's*, gericht op duidelijk omschreven doelgroepen, zeker als er bij de uitvoering van de zorg sprake is van betrokkenheid van meerder instellingen' (uit: Samenvatting en aanbevelingen').

Met betrekking tot de interne processen van de instelling rond de zorgverlening valt zeker nog *doelmatigheidswinst te behalen*, meent de commissie. 'Waar mogelijk moet gebruik gemaakt worden van landelijk vastgestelde referentiekaders'. Instellingen dienen jaarlijks beleidsplannen op te stellen. Ook acht de commissie van belang: meer uniformering, protocollering en standaardisering van de interne processen, en van hulp aan specifieke doelgroepen.

Het *transparant maken van het zorgproces* door heldere beslismomenten en het goed vastleggen ervan in dossiers acht de commissie ook van belang. Uit een oogpunt van doelmatigheid hecht de commissie eraan dat goede registratie van cliëntgegevens en van de geleverde zorg wordt opgebouwd.

Benchmarking als landelijk traject

Het werken met spiegelinformatie acht de Commissie Doelmatigheid Riagg-activiteiten ook van belang. Dat dient te gebeuren op basis van *kengetallen*. Daarvoor kan worden aangesloten bij het landelijk traject rond *benchmarking* van de sector ggz.

Het onderzoek van Bakkenist Management Consultants heeft een aantal *benchmarkgegevens* opgeleverd die mogelijk een bijdrage aan verbetering van de doelmatigheid kunnen leveren.

De belangrijkste kengetallen die in het onderzoek gebruikt zijn:

1. aantal behandelde cliënten per f 100.000 budget
2. aantal zittingen per f 100.000 budget;
3. 'span of support overhead';
4. beschikbaarheid;
5. aantal hulpverleners binnen de drie afdelingen jeugd, volwassenen en ouderen;
6. aantal behandelde cliënten per fte hulpverlener in de drie afdelingen;
7. aantal cliënten per 100 zittingen in de drie afdelingen;
8. klanttevredenheid.

Daarnaast zijn kengetallen ontwikkeld die betrekking hebben op het aantal inwoners in de regio, het overige ggz-aanbod in de regio, de totale kosten in verhouding tot het aantal inwoners, de omvang van het personeel en het ziekteverzuim van het personeel. Daar bleef het niet bij. Er zijn ook kengetallen beschikbaar gekomen over de inzet van personeel voor hulpverlening in relatie tot andere activiteiten binnen de instelling, de productiviteit per hulpverlener gezien in relatie tot 'face-to-face' uren, aantal zittingen en aantal behandelde cliënten. Ten slotte is nagegaan hoe het personeel verdeeld is over de drie afdelingen.

Zorgkantoren

De commissie acht de vorming van zorgkantoren in het licht van de modernisering van de AWBZ een gunstige ontwikkeling in het bevorderen van een actieve,

sturende rol voor de verzekeraar in de regio. Voor een optimale taakuitvoering door de zorgverzekeraar als zorgkantoor dienen enkele condities vervuld te zijn.

1. Er is een *landelijk referentiekader* nodig voor de kwalitatieve taakuitvoering.
2. Een landelijke *toezichthouder* is gewenst, die de zorgkantoren toetst aan het landelijke referentiekader.
3. De commissie bepleit een 'goed *benchmarkingsmodel* waarmee de instellingen en regio's onderling kunnen worden vergeleken op kwaliteit en doelmatigheid van de organisatie van de zorgverlening'.

Incentives voor doelmatiger werken

De commissie meent dat de doelmatige financiering van doelmatige zorg een degelijk systeem 'met voldoende countervailing power' vergt. 'In dit systeem dienen voor alle partijen *incentives* te worden ingebouwd die het voor partijen interessant maken om zo doelmatig mogelijk te werken. De instellingen moeten zo transparant mogelijk werken, de zorgkantoren dienen een actieve, controlerende functie te vervullen, en de controle op de zorgkantoren moet worden uitgevoerd door de Ziekenfondsraad. Registratie en informatie zijn noodzakelijk voor een goed functionerend systeem.

Regionaliseren budgetsystematiek

De actuele financieringssystematiek is volgens de Commissie Doelmatigheid Riagg-activiteiten te verbeteren. De commissie wil de budgetsystematiek regionaliseren. De aan afzonderlijke instellingen toegekende budgetten kunnen regionaal worden gewijzigd als er sprake is van ondoelmatig functioneren of bijvoorbeeld andere prioriteitsstelling. Wat een instelling aan afgesproken productie niet haalt, zou weggehaald moeten kunnen worden voor intensivering van andere zorg in de regio, is de redenering. De huidige systematiek geeft onvoldoende prikkels om zo te handelen, omdat wat over is, terug vloeit naar de 'Centrale Kas', zoals de commissie het uitdrukt. Wil dit overhevelen van budgetten in een geregionaliseerde financieringsstructuur mogelijk gemaakt worden, dan zijn meer basisgegevens nodig (zie 'Samenvatting en aanbevelingen', 1998).

- 'De commissie stelt zich op het standpunt dat de regionalisatie van de budgetsystematiek kan bijdragen tot een meer effectieve sturing. Als strikte voorwaarde dienen dan wel transparantie en verantwoordingsplicht van zorgaanbieders en zorgkantoren versterkt te worden'.
- 'Voor een dergelijk financieringssysteem zijn bepaalde minimum-registratiegegevens vanuit de instelling noodzakelijk. Deze moeten in de financieringsvoorwaarden worden opgenomen'.
- 'De commissie bepleit dat het in het voorgestelde geregionaliseerde budgetsysteem niet langer mogelijk moet zijn dat de zorgverzekeraars toelaten dat Riagg'en gedurende meerdere jaren meer dan 100% productie realiseren zonder dat aanpassing van de budgetafspraken volgt. De als gevolg van aanpassing beschikbaar

komende middelen kunnen voor andere zorg in de regio worden aangewend'. Einde citaat.

De commissie meent dat productomschrijvingen niet aan bepaalde instellingen gekoppeld moeten worden, maar zo moeten worden geformuleerd dat de uitvoering door diverse instellingen kan plaatsvinden. Logisch, want anders valt er niet goed over te hevelen.

Op termijn zou voor een uniforme financiering van de ggz 'gedacht kunnen worden aan integrale planning van de ggz', stelt de Commissie Doelmatigheid Riagg-activiteiten.

Reflectie op getallen

Voor fijnproevers heeft de commissie nog wat in petto. 'De commissie vraagt zich af of binnen de financieringssytematiek het huidige systeem van "gewogen inwoners", wel doelmatig is. Uit het onderzoek van Bakkenist, en uit de bevindingen van de vertegenwoordigers van de Riagg'en de zorgverzekeraars komen signalen dat de wegingsfactor voor de "gewogen inwoner" in het financieringssysteem voor de Riagg mogelijk te zwaar is. Argumenten hiervoor worden gevonden in mogelijk twee punten. Zo blijkt dat instellingen met een budget dat gebaseerd is op veel gewogen inwoners een relatief lage productie hebben. Bovendien is bij deze instellingen (die in stedelijke gebieden liggen) sprak van een ruim regionaal aanbod aan ander ggz-voorzieningen'. Einde citaat. De commissie meent daarom dat dit systeem van "gewogen inwoners" binnen de voorgestelde regionalisatie van de budgetssystematiek herzien moet worden. De *recent ontwikkelde budgetparameters* worden door de commissie in het licht van de doelmatigheid in de financieringssytematiek wel gewaardeerd omdat ze 'dicht bij het zorgproces aansluiten', aldus de commissie.

8 Wat levert de Algemene Rekenkamer aan inzicht over benchmarking

Het SCP publiceert benchmark-studies maar doet de Algemene Rekenkamer aan benchmarking en licht ze zichzelf ook door? Het antwoord is niet bemoedigend voor de benchmark-adepten.

De Algemene Rekenkamer doet niet aan benchmarking. De Algemene Rekenkamer onderzoekt wel de rechtmatigheid en doelmatigheid van rijksuitgaven en ze beklemtoont ook sinds jaar en dag de wenselijkheid van *kengetallen*. Kengetallen zouden de Tweede Kamer informatie moeten verschaffen over kosten en prestaties van publieke voorzieningen. De oogst blijkt echter nog mager. Circa de helft van honderd onderzochte, en volgens het ministerie van Financiën ook zinvol toe te lichten, artikelen uit de begroting 1997 ging vergezeld van volgens de eisen opgestelde ramingskengetallen (De Groot, 1998: 241). Ramingskengetallen zijn kengetallen waarmee het begrotingsbedrag wordt onderbouwd. Slechts een kwart

was toegelicht met doelmatigheidskengetallen, zoals kosten per prestatie en nog veel minder artikelen met doeltreffendheids- en/of kwaliteitsinformatie. Het is dus te hoog gegrepen om van dit college van staat te verwachten dat die institutioneel, over de volle breedte van rijksbeleid komt met benchmarking. Dat is ook onmogelijk.

Plaatst de Algemene Rekenkamer zichzelf in een benchmark? Een overheidsorganisatie als geheel is te bezien door te kijken naar de toelichting op de begroting in termen van onderbouwing met kengetallen over kosten en prestaties. Dat is bij nader inzien niet zo eenvoudig. De Algemene Rekenkamer (AR) zelf rapporteerde tot 1996 *over de eigen organisatie* (!! in elk geval nog nauwelijks in die zin. Dat is dan ook gelaakt (Dolmans, 1995). Dat de AR in haar jaarverslagen als sinds lang ook informatie opnam over de inhoud van haar rapporten kan niet als een excuus gelden.

Intussen is geprobeerd iets aan de ontbrekende rapportage van kengetallen te doen. Het gaat om stappen op weg naar verdere benchmarking. In het *Jaarverslag 1997* zijn kengetallen opgenomen die informatie geven over kosten en prestaties. Welke onderdelen bevat die analyse? De Groot (1998) geeft hiervan een overzicht, dat we samenvatten.

a *De productmeting*: productie van rapporten aan de Staten-Generaal. Dat is een kwalitatieve analyse. Een onderdeel hiervan zijn de rapporten van de AR bij de financiële verantwoordingen van de ministeries op basis van een wettelijke verplichting, vastgelegd in de Comptabiliteitswet. Daarnaast is sprake van een tweede categorie: onderzoek naar financieel beheer. Een derde categorie zijn de rapportages in het kader van het doelmatigheidsonderzoek. Die trekken doorgaans de meeste aandacht in de media en samenleving.

b *De tijdsbestedingsanalyse*: analyse van de inzet van personeel voor de productie van rapporten. Het gaat hier voornamelijk om onderzoeksdagen waarover informatie beschikbaar is in de vorm van een tijdschrijfsysteem. De helft van de totale tijd ging in 1997 op aan 'directe onderzoekstijd', de rest aan 'ontwikkeling en programmering' (22%), en aan beheer (28%).

c *De middelenanalyse*: de analyse van uitgaven voor productontwikkeling. Hierbij gaat het om de groei of daling van de personele uitgaven in verschillende categorieën. Het absolute niveau van de personele uitgaven blijkt vergelijkbaar met die voor andere onderzoeks- en adviesorganen. De materiële uitgaven liggen evenwel opvallend laag vergeleken met instellingen als het Centraal Planbureau of de Raad van State.

d De bepaling van de *integrale kostprijs* per onderzoeksdag. Een stijging hiervan wordt vooral bepaald door loonstijging.

In het artikel van Hans de Groot zijn tabellen opgenomen. De Algemene Rekenkamer probeert de komende jaren een vergelijking te maken met de bedrijfsvoering van andere rekenkamers. Door ijkpunten op te stellen ('benchmarks') kan de doelmatigheid van de Rekenkamer verder extern worden getoetst.

9 Bedrijfsvergelijking bij andere organisaties

Pogingen tot benchmarking komt intussen op alle overheidsniveaus voor. Neem gemeenten. De gemeenten Barneveld, Heusden, Noordoostpolder en Soest hebben samen hun eerste ervaringen met benchmarking opgedaan.

Benchmarking kent volgens Helder, die de ervaringen als sector controller van Soest met benchmarking beschrijft, vier basiselementen.

Het trachten te leren van ervaringen buiten de eigen omgeving is het eerste element.

Ten tweede, benchmarking is niet alleen financieel of instrumenteel gericht.

Ten derde, benchmarking is een vergelijking gericht op best practice.

Ten vierde, benchmarking is op continue verbetering gericht.

Deze aanduidingen zijn nodig omdat de indruk kan bestaan dat benchmarking alleen gericht is op wat meetbaar is. Immers, wat meetbaar is, is makkelijk te vergelijken. Daardoor zou benchmarking een zeer eenzijdig middel worden en zou het gebruik van organisatievergelijking toch benchmarking in de instrumentele hoek drijven. Wie aan benchmarking doet, zal rekening moeten houden met waarderingen door burgers en beleidsdoelstellingen. Benchmarking is zinvol als valkuilen vermeden worden.

Een item dat aan benchmarking is onderworpen betreft onderhoud van wegen. Het verschil tussen gemeenten bleek op dit vlak groot te zijn.

Literatuur

- Bogt, H. ter, Prestatiegegevens: nuttig voor sturing van gemeentelijke taken, in: Public Controlling, juni 2004, pp. 52-58.
- Groot, H. de, (red.), Benchmarks in de publieke sector, Sdu, Den Haag, 2004 (serie Public controlling, nr. 4).
- Hoogwout, M., Het ontwerpen van overheidsbenchmarks – Een krachtig management- en verantwoordingsinstrument, in: Public Controlling, juni 2004, pp. 32-42.
- Smit, B., T. Tadema-Van Dam, en D. van der Zwaag, Van beheren naar beheersen – Sturen op prestaties, financiën en bezetting bij de gemeente Leiden, in: Public Controlling, juni 2004, pp. 16-28.
- Strikwerda, J., Shared service centers en control in de publieke sector – Hoe verdere decentralisatie leidt tot betere control, in: Public Controlling, juni 2004, pp. 22-27.
- Strikwerda, J., Shared service centers: van kostenbesparing naar waardecreatie, Van Gorcum, Assen, 2003.
- Waal, A. de, en L. Kerklaan, De resultaatgerichte overheid – Op weg naar de prestatiegedreven overheidsorganisatie, Sdu, Den Haag, 2004.

10 Sturing op prestaties en perverse effecten neutraliseren

We komen nu bij een specifieke problematiek.

Eind april presenteerde de lijsttrekker van de PvdA het plan *'100 Dagen, 10 Punten, 1 Land: Contract met Nederland'*. Daarin stonden de belangrijkste tien punten die de PvdA wil aanpakken als deze partij deel gaat uitmaken van de regering. Het was de basis voor de inhoudelijke campagnelijn tot 15 mei 2002. Hoewel de PvdA bij de verkiezingen fors verloor, zit de partij op *een trend naar sturen op prestaties*. De PvdA zegt toe dat het volgend kabinet een plan presenteert *'waarmee het probleem van de wachtlijsten uiterlijk in 2006 is weggewerkt'*. De PvdA wil afspraken over prestaties. Feitelijk komen prestatieafspraken al voor.

Het beleid van het kabinet-Kok I en II laat zien dat diverse pogingen zijn gedaan overheidssturing van bovenaf te vervangen door een soort markt gecombineerd met output- of resultaatmeting. Op een echte markt bestaat geen stelsel van prestatieafspraken en -meting. Is het product wat u kocht niet goed, dan ruilt u het of gaat u de volgende keer naar een andere zaak. Maar om een voorbeeld te geven, een onderwijsinstelling krijgt wel een prestatieverplichting, bijvoorbeeld een gemiddelde score op Cito-toetsen. Dat kan leiden tot perverse effecten, bijvoorbeeld het niet mee laten doen van zwakke leerlingen met toetsen of het organiseren van trainingen in Cito-toetsen in plaats van leerlingen andere dingen te leren (Hilhorst, 2001: 77). Prestatiemeting is daarmee een intrigerend verschijnsel. Prestatiemeting lijkt gewenst maar is niet probleemloos. Bezinning is nodig want in de coalitiebesprekingen zal een verdere druk om te komen tot afspraken over te leveren prestaties ontstaan. De grote winnaar van de Tweekamerverkiezingen de Lijst Pim Fortuyn (LPF) is voorstander van prestatieafspraken. De 6 mei 2002 overleden lijsttrekker Pim Fortuyn van LPF was voorstander van meer prestatieafspraken en tegelijk geen voorstander van meer bureaucratie, bijvoorbeeld door uitgebreide verslagen over prestatiemeting.

Nemen we enige afstand tot de politieke actualiteit dan is sturing op prestaties nog steeds een trend. Respondenten uit een onderzoek van Berenschot gaven een hoge score aan *'sturing op prestaties'* als wenselijkheid, die past in een bestuursbeeld van de *'transactiesamenleving'*. De publieke zaak is volgens de deelnemende professionals gediend met een meer bedrijfsmatige overheid maar signaleren ook dat er nog weinig resultaat met deze aanpak behaald wordt (Van Twist e.a., 2002: 85, 95).

De verdere beschouwing gaat over prestatiemeting in het openbaar bestuur. Daarbij komen de volgende vragen aan bod:

- Welke gevolgen kunnen optreden: positieve en perverse;
- Welke verklaring is te geven voor perverse effecten?;
- Welke strategieën zijn mogelijk om perverse effecten van prestatiemeting te neutraliseren of minimaliseren?

- Welke uitkomsten blijken? We richten ons op de agentschappen.

Begrippen

We kennen meerdere vormen van overheidssturing. Te noemen zijn aanbod- en vraagsturing. Aanbodsturing betekent dat het aanbod regulerend is voor de vraag. Neem een ziekenhuis waarvan niet meer bedden bezetten kunnen worden door patiënten dan er bedden zijn. Dat leidt tot de problematiek van wachtlijsten. Ook is het mogelijk om burgers een zogenaamd persoonsgebonden budget te geven, waarmee deze burger zorg kan inkopen. Dat is een voorbeeld van vraagsturing. We gebruiken hier een ander onderscheid, namelijk in:

- *inputsturing*,
- *throughputsturing* en
- *outputsturing*.

Hier hebben we het over outputsturing of *sturing op resultaat*. Sturing op resultaat staat hier gelijk aan *sturing op prestaties*.

Voor outputsturing bestaan zeker mogelijkheden. Beleid voeren betekent immers vrijwel altijd het vaststellen van de wenselijkheid van een prestatie en de organisatie daarvan. In het openbaar bestuur komen we dan ook veel prestaties tegen zoals het geven van een bekeuring of boete, het in bewaring stellen van een moordenaar, het afgeven van een paspoort of van een uittreksel uit een register of van een vergunning, het verstrekken van een uitkering of subsidie, het sluiten van huwelijken of toewijzen van een woning, het kopen van grond, het ophalen van huisvuil, het oprichten van een museum of schouwburg, het draaiende houden van een schouwburg. In het openbaar bestuur kennen we niet alleen de levering van prestaties zelf, zoals een woningtoewijzing of huisvuilophaal, maar ook daadwerkelijke prestatiemeting. Dan gaat het om de vraag: zijn de gewenste prestaties ook geleverd? Hoeveel bezoekers bezocht toneel-, ballet- en muziekvoorstellingen in de gemeentelijke schouwburg? En, kwamen er genoeg bezoekers naar de programma's die de schouwburgdirectie bood om het bedrijf rendabel te houden?

De gedachte achter prestatiemeting is eenvoudig: een overheidsorganisatie beoogt prestaties te leveren dus ligt het voor de hand om te kijken of die ook *daadwerkelijk* vallen aan te treffen, hoeveel en of het de goede zijn, en of het bijvoorbeeld foutloos (als dat van toepassing is), of op tijd of klantvriendelijk gebeurt. Daarvoor moet men *prestatieindicatoren* gebruiken. Denk aan het aantal afgegeven vergunningen in een jaar in verhouding tot het aantal medewerkers op een afdeling.

Als een *prestatienorm* ontbreekt, kan men die construeren. Een manier hiervoor is om vergelijkbare organisaties erbij te betrekken. *Benchmarking* dient zich dan aan. Benchmarking betreft de vergelijking van organisaties op het punt van werkwijzen en/of prestaties en/of effecten. Denk hierbij aan benchmarking van sociale diensten.

Het gaat in werkelijkheid natuurlijk niet altijd alleen maar om een prestatie (output) maar om een *effect* (= *outcome*). Die outcomes zijn vaak moeilijk meetbaar. Het kan namelijk een tijd duren voor de consequenties van een prestatie zichtbaar worden. Denk aan maatregelen om de leefbaarheid of veiligheid in een wijk te bevorderen. Nu kan wel onderscheid maken in *directe* en *indirecte effecten*. Een direct effect van een verkeersveiligheidsmaatregel kan zijn dat het aantal verkeersslachtoffers in een wijkje daalt, maar doordat verkeer ook andere routes gaat volgen kunnen er elders meer slachtoffers vallen. Wanneer wordt dat zichtbaar? Prestaties als drempelplaatsing worden wel betiteld als directe effecten (De Bruijn, 2002: 141).

Functies van prestatiemeting

Prestaties worden gevraagd van politiemensen (bijvoorbeeld verkeerscontroles, bekeuringen, opheldering van misdrijven, enz.), scholen (rendementen), schouwburgen (aantallen bezoekers, aantallen voorstellingen), musea (aantallen collecties, aantallen bezoekers), sociale diensten (aantallen verstrekte uitkeringen).

Prestatiemeting kan een aantal doelen of functies vervullen. De Bruijn noemt de volgende:

1. *Transparantiebevordering*. Bijvoorbeeld van de vraag: wat was de gewenste productie en welke kosten zijn daarvoor gemaakt? Deze vraagstelling komt in de buurt van een kosten-batenanalyse.
2. *Leren*. Na het meten van prestaties kan ene organisatie zich afvragen of een normverandering nodig is op het vlak van kwantiteit of kwaliteit van de prestaties respectievelijk een verlaging van kosten wenselijk is. Waar zijn verbeteringen mogelijk?
3. *Oordelen*. Prestatiemeting biedt de mogelijkheid om een oordeel uit te spreken over het functioneren van een organisatie. Dat kan alleen onder voorwaarden. Men zal een prestatiemeting in den brede moeten plegen.
4. *Afrekenen*. Oordeelsvorming kan een sanctie voor goed of slecht presteren tot gevolg hebben.

Deze functies hebben een opklimmende mate van ambitie en dwang.

De daadwerkelijke positieve effecten van prestatiemeting

Denk eens aan de prestaties van scholen, musea, schouwburgen, sociale diensten, politiecorpsen. Welke effecten kunnen van prestatiemeting uitgaan. We onderscheiden positieve en perverse effecten. Prestatiemeting kan de volgende positieve gevolgen hebben, aldus De Bruijn (2002) in een artikel in Bestuurswetenschappen, dat we hier samenvatten. .

1) Prestatiemeting leidt tot transparantie.

Een prestatiemeting kan een organisatie inzicht geven in de omvang van het aantal verstrekte of geleverde producten, de kosten per product of hoeveel een

organisatieonderdeel bijdraagt aan de productie van een organisatie. Zeg: de gemeente x verstrekt per jaar zoveel paspoorten, zoveel vergunningen en haalt bij zoveel eenheden zus en zoveel huisvuil op en het ophalen ervan vertoont een stijgende lijn. In principe kan transparantie verder bevordert worden en uiteindelijk bijdragen tot rationalisatie van de bedrijfsvoering (bijvoorbeeld: zelfde productie met minder mensen?) of verandering in de productieomvang (het proberen de productie te verhogen of verlagen). Het is mogelijk om bijvoorbeeld de betekenis van nieuwe productiewijzen op de productie te meten.

2) Prestatiemeting is een prikkel voor productie.

Prestatiemeting is niet gericht op de inspanning (input) en evenmin op de activiteiten en dis de inspanning (throughput) maar op de prestaties en dus de resultaten (output). Het hanteren van een outputmodel in een organisatie kan ertoe leiden dat de productie ook omhoog gaat. Dat was bijvoorbeeld bij universiteiten het geval. Als een vakgroep beloond wordt voor het aantal publicaties, zullen leidinggevenden het publiceren stimuleren en zal het aantal artikelen en boeken ook toenemen. Ook in andere organisaties wordt wel een verband aangetroffen tussen prestatiemeting en toenemende productie.

3) Prestatiemeting is een vorm van verantwoording.

Organisaties krijgen in een complexe maatschappij veel vrijheid om professionele expertise op eigen wijze in te zetten. Denk aan politiecorpsen. Om toch zicht te houden op de productie worden prestatieafspraken gemaakt. Prestatiemeting is dan een vorm van terugkoppeling. Het laten zien dat er waar voor het geld is geleverd. Een organisatie kan zich zo verantwoorden en inzicht geven in de geleverde prestaties. Dat lukt meer naarmate de informatie communiceerbaar is, de taakuitvoering is geobjectiveerd en de informatie met zekere regelmaat kan worden aangeleverd.

Perverse effecten van prestatiemeting

Prestatiemeting kan ook (enigszins) negatieve gevolgen hebben. We spreken van perverse effecten omdat deze niet bedoeld zijn en vaak ongunstig. Ze kunnen te maken hebben met de opstelling van degenen die betrokken zijn bij de meting van prestaties of de gevolgen ervan ondervinden. De Delftse hoogleraar De Bruijn noemt de volgende.

1) Prestatiemeting leidt tot strategisch gedrag.

Prestatieafspraken en meting kan een incentive zijn voor strategisch gedrag. Dat betekent bijvoorbeeld dat een organisatie alle kaarten zet op de gevraagde prestaties terwijl prestaties die niet gevraagd worden maar wellicht uit professioneel of klantperspectief belangrijk(er) zijn worden verwaarloosd. Een voorbeeld zijn de universiteiten. Als een minister van onderwijs met universiteiten afsprekt om ze te financieren op basis van aantal bachelor- en masterdiploma's en deze

financieringsgrondslag naar faculteiten doorvertaling vindt, zal de neiging bestaan om studenten makkelijker te laten slagen. Immers, hoe meer aantal afgestudeerden hoe meer geld. Gaat de minister betalen op basis van aantal publicaties van medewerkers en niet op basis van een onderwijscriterium dan zal de nadruk op aantal publicaties vallen en onderwijs intern in minder aanzien gaan staan, op den duur. Wat je erin stopt, komt eruit. Prestatieafspraken en meting kunnen zogezegd verloederend werken. De actoren uit een universiteit kunnen zich dus gaan gedragen zoals de incentivestructuur wil dat ze zich gedragen.

2) *Prestatiemeting versterkt de interne bureaucratie.*

Onderzoek laat zien dat die organisaties goed scoren in een stelsel van prestatiemeting die goed scoren in procedureel-organisatorisch opzicht. Wie als universiteit veel geld wil vangen voor rendementen in het onderwijs, moet dus de aantal afgestudeerden goed tellen en zo mogelijk handigheidjes bedenken waardoor de rendementen omhoog gaan. Dat is ook gebleken. Een bepaalde hogeschool zorgde ervoor dat (min of meer fictieve) studenten uit Vlaanderen werden ingeschreven: kunstmatige opkrikking van de cijfers. Een universiteit kan doorstroomstudenten van andere opleidingen een doorstroomprogramma aanbieden of eerst een bachelor laten halen. In het laatste geval wordt het aantal bachelors opgekrikt.

3) *Prestatiemeting verhindert innovaties.*

Een prestatie leverende organisatie richt zich meestal ook op optimalisering van het productieproces dat tot prestaties moet leiden. Daarbij zal vooral ingestoken worden op die processen die leiden tot prestaties waarvoor de meeste beloningen zijn te halen. Als een universiteit wordt betaald op basis van aantallen afgestudeerden hebben studierichtingen met veel studenten het makkelijk maar lopen andere studierichting het gevaar van opheffing. Er bestaat ook een gevaar dat innovatie niet telt want die levert weinig geld op omdat die niet in een prestatie-indicator is gevangen. Wie innoveert is bezig met het onbekende en dat wordt meestal niet in prestatie-indicatoren gevangen. Extreem gesteld, een research & development-afdeling zal buiten de prestatiemeting gehouden worden. Prestatiemeting 'beloont het voortdurend reproduceren van het bestaande', zegt De Bruijn (2002: 143).

4) *Prestatiemeting blokkeert ambities.*

Organisaties die met prestatieafspraken te maken hebben en met prestatiemeting neigen ertoe om hun prestaties te vergroten door de output te verbeteren met zo min mogelijk throughput. Moeilijke gevallen moeten vervallen want die kosten geld. Denk aan onderwijs voor bijzondere leerlingen. De Bruijn (2002: 144) ziet het gevaar van inleveren op ambities.

5) *Prestatiemeting verdrijft de professionele habitus*

Een systeem van prestatiemeting beloont doorgaans niet het riskante in de productie maar de herhaling, het kwantitatieve, het meetbare. Daarom komt prestatiemeting

vaak voor bij organisaties met 'bulk'-productie, dus met grote aantallen producties of diensten en niet bij bijvoorbeeld een geheime dienst. Een museum dat werkt met prestatie-indicatoren zal zeker een indicator opnemen voor bezoekersaantallen. Voor zaken als de integriteit van de collectie of de betekenis van een collectie voor verder onderzoek zal minder aandacht bestaan. Prestatiemeting kan door de sterke focus op goed te omschrijven taken en uitkomsten de professionele habitus in musea en ook schouwburgen naar de achtergrond verdrijven

6) Prestatiemeting verdrijft bestelverantwoordelijkheid.

Het effect van de terugtrek van de professionele habitus als gevolg van prestatiemeting kan zich tussen organisaties voordoen en zo doorwerken in een bestel. Een schouwburgdirectie die alleen let op bezoekersaantallen zal geen contracten afsluiten met gezelschappen die modern, abstract toneel brengen maar de voorkeur geven aan 'hapklare brokken' die veel bezoekers trekken. Dat heeft een verschralend effect op het geheel van podiumkunsten. 'Prestatiemeting kan (..) bestelverantwoordelijkheid verdrijven' zegt De Bruijn. Verzwakking van bestelverantwoordelijkheid kan zich ook voordoen in de sector onderwijs. Prestatiemeting van scholen kan leiden tot meting van aantallen leerlingen en rendementen. Worden scholen elkaars concurrent dan zullen ze minder bereid zijn om inzicht te geven in de eigen 'best practices'.

7) Prestatiemeting bestraft goed presteren.

Als organisaties meewerken aan prestatiemeting kan dat in geval van transparantie leiden tot inzicht in welke organisatie beter presteert en welke minder. Het gevolg hiervan kan zijn dat een externe financier tot het inzicht komt dat dezelfde output met minder input te bereiken is, waarna het budget verlaagd wordt. Deelnemers aan prestatiemeting nemen dus het risico dat ze in eigen vlees snijden.

Conclusies

Uit beide overzichten van effecten vallen duidelijke conclusies te trekken. Vanuit het *new public management* - denken bestaat er aandacht voor prestatiemeting en productiviteit. Het is inderdaad mogelijk om veel beoogde prestaties (gewenste directe effecten) van veel overheidsorganisaties in kaart te brengen maar dat kunnen zowel positieve als perverse effecten zijn. Vallen die perverse effecten van prestatiemeting te verklaren en wat is er tegen te doen? Daarover gaat het vervolg.

Verklaring voor het optreden van perverse effecten van prestatiemeting

Er zijn verschillende verklaringen mogelijk.

1 De eerste verklaring voor perverse effecten bestaat hierin dat *professionals het systeem daadwerkelijk kunnen perverteren en zich ook hiertoe gelegitimeerd achten* (De Bruijn, 2002: 145).

Toelichting: Professionals kunnen leren zich af te schermen tegen systemen van prestatiemeting, zoals bij de eerste twee genoemde perverse effecten aan de orde is. Dat kunnen ze doen door het meetsysteem van 'kunstmatige' informatie te voorzien die als voor hen gunstig 'in zijn gevolgen' wordt gezien. Het voorbeeld laat zien dat *prestatiemeting een ethisch vraagstuk* kan zijn.

Een voorbeeld: De therapeuten van een Riagg moesten aantal contacten opgeven als indicator voor productie, waarbij jaarlijks een bepaalde productie gehaald moest worden. De directeur zette een meldingssysteem op. De therapeuten kunnen ertoe neigen om een afspraak met een cliënt in meerdere contacten om te zetten, zodat de zelfde hoeveelheid tijd en dezelfde therapeutische handeling nu tot meer productie 'op papier' leidt. Het gevolg is dat het systeem bij de neus genomen wordt. Er ontstaat een transparante papieren werkelijkheid die niet meer spoort met de werkelijkheid. Maar beschikt een directeur eenmaal over deze informatie dan werken prikkels krachtig op deze in om hiermee te volstaan. Welke directeur wil immers dat de productie verlaagd wordt, waarna de organisatie straf krijgt omdat ze niet aan een norm voldoet.

Bij de opstelling van professionals ten aanzien van prestatiemeting komen we in de praktijk volgende overwegingen tegen.

1. Ten eerste wordt prestatiemeting wel *als armoedig gezien*. Een aantal therapeuten uit ons voorbeeld denkt er zo over. Ze vinden dat therapie zoveel 'waarden' vertegenwoordigt dat het tellen van contacten daarmee niet eens schril afsteekt.
2. Een tweede overweging uit de praktijk behelst de stelling dat prestatiemeting *unfair* is, omdat prestaties het resultaat zijn van samenwerking (coproductie). De therapeuten uit ons voorbeeld zullen als argument tegen prestatiemeting in de vorm van het tellen van aantal contacten aanvoeren dat therapie iets is van therapeut en cliënt en dat het werkelijke effect niks zegt over aantallen contacten.
3. Een derde overweging tegen prestatiemeting is dat die *statisch* is terwijl prestaties vaak dynamisch zijn. Een schouwburgdirecteur kan geneigd zijn tot een contract met een groep voor experimenteel toneel maar het uitvallen hiervan kan aanleiding geven tot een contract met toneelgroep die een veel publiek trekkend stuk op de planken brengt. Wat telt nu daadwerkelijk?
4. Een vierde overweging is dat bij prestatiemeting de *context* niet wordt meegenomen. We geven een fictief voorbeeld. Neem een museumdirectie die een tentoonstelling over kunst van moslim fundamentalisten uit Afghanistan houdt. Stel, deze tentoonstelling opent twee weken na de ramp met de vliegtuigkapers in de Twin Towers. Dat kan van de tentoonstelling een sterk de aandacht trekkende gebeurtenis maken of iets dat 'niet loopt'. Prestatiemeetsystemen houden met dit soort omstandigheden geen rekening mee.

2 Een tweede verklaring voor het optreden van perverse effecten van prestatiemeting luidt volgens De Bruijn: *'hoe meer bestuurders willen sturen door prestatiemeting, des te minder effectief zal prestatiemeting zijn'*.

Toelichting: Prestatiemeting kan vanuit meerdere doelen gebeuren. Een ervan was – zoals geformuleerd – transparantiebevordering, een ander 'leren', een derde 'oordelen' en een vierde 'afrekenen'. De neiging is groot om prestatie-meetsystemen te gebruiken voor een *command and control* - achtige aanpak. Een bestuurder kan in prestatie-afspraken een prestatienorm neerleggen om vervolgens als een organisatie na meting blijkt dit doel niet te halen te gaan afrekenen in termen van positieve of negatieve sancties. Prestatieafspraken kunnen als doelen makkelijk gehaald worden, gebruikt worden om de prestatienorm op te schroeven, dus om ambitieuze doelen te realiseren. Wie prestatieafspraken maakt vanuit een command and control-optiek houdt evenwel geen rekening met de *paradox van de toenemende perverse effecten*. De paradox van de toenemende perverse effecten houdt in: 'naarmate de bestuurder door prestatiemeting meer invloed wil uitoefenen op het primaire proces, neemt het aantal perverse effecten toe'. Dat komt omdat professionals hun pogingen versterken om zich van de prestatiemeting af te schermen. Professionals zullen zich ook wapenen tegen afrekenen door te benadrukken dat hun productie co-productie met anderen is, dat omstandigheden van invloed en een dynamiek kent.

Dit doet zich bijvoorbeeld voor als een ministerie prestatieafspraken maakt met de Open Universiteit Nederland over aantal propedeuseverklaringen en aantal afstudeerders. Het aantal studenten in verschillende studierichtingen heeft een universiteit niet in de hand. De universiteit kan moeilijk de maatschappelijke aantrekkelijkheid voor studies als bedrijfskunde of rechten en daarmee de toestroom beïnvloeden. Medewerkers zullen dus geen bezwaar maken tegen prestatiemeting gericht op transparantie of leren maar wel tegen oordelen zonder deze factoren mee te wegen of tegen afrekenen.

Prestatiemeting is niet effect als professionals zich niet conformeren aan een systeem van prestatiemeting. Dan treden de perverse effecten 1 en 2 op. Als professionals zich wel aan een systeem van prestatiemeting conformeren, treden de perverse effecten 3 tot 7 op, stelt De Bruijn (2002).

11 Strategieën ter beperking van perverse effecten van prestatiemeting

Welke strategieën hanteren organisaties om om te gaan met een tweeslachtig beeld van prestatiemeting: enerzijds gewenste positieve effecten en anderzijds perverse effecten? Hoe kan prestatiemeting zo worden vormgegeven dat perverse effecten uitblijven of minimaal blijven? De Bruijn noemt er een aantal. Strategie 2 komt voor eigen rekening.

Figuur: *Strategieën ter beperking van perverse effecten van prestatie-meetsystemen* vlg. o.a. De Bruijn (2002):

Strategie 1: tolerantie voor een variëteit aan (concurrerende) productdefinities.
Strategie 2: verdisconteer omstandigheden zoals coproductie en dynamiek.
Strategie 3: een verbod op een monopolie op de betekenisgeving.
Strategie 4: beperking van functies en fora.
Strategie 5: strategische selectie van producten.
Strategie 6: management van de concurrerende benaderingen 'product' en 'proces'.

Strategie 1: tolerantie voor een variëteit aan (concurrerende) productdefinities.

Iedere productdefinitie voor een meervoudig product is discutabel. Bij organisaties kan de neiging ontstaan om productie te reduceren tot een specifieke productdefinitie. Als dat gebeurt, zoals in geval dat een universiteit alleen aantal afstudeerders telt en een gevangenisdirectie alleen het aantal vrijdelde uitbraakpogingen, dan liggen perverse effecten op de loer omdat werknemers dergelijke meting als unfair ervaren. Het antwoord zou moeten zijn om tolerant te zijn in termen van het toelaten van ene variëteit aan productdefinities, meent De Bruijn. Hij heeft gelijk.

Een voorbeeld hiervan is het inrichten van een prestatie-meetsysteem voor onderzoek van universiteiten. Tellen Nederlandstalige artikelen in een wetenschappelijk en vaktijdschrift mee en Engelstalige, maar wordt dan een weegfactor ingevoerd voor die artikelen die verschijnen in 'goede' wetenschappelijke A-tijdschriften? Hoe zwaar tellen artikelen in tijdschriften voor ambtenaren of andere doelgroepen? Tellen boeken in het Engels zwaarder dan die welke verschijnen bij een Nederlandse uitgever in het Nederlands? Wordt rekening gehouden met het feit dat een artikel of een boek een coproductie is van meerdere auteurs? Wordt dan dubbeltelling voorkomen? Telt ook de begeleiding door een promotor van een proefschrift mee in de telling? Kortom, de lezer voelt het al: een prestatie-meetsysteem van zoiets als onderzoek kent een aantal valkuilen en arbitraire keuzen. Ook onbillijkheid is mogelijk en verschil tussen studierichtingen. Juristen richten zich sterk op het Nederlands recht. Hun betekenis zou afnemen als ze uitsluiten in Engelstalige tijdschriften publiceren.

Als verschillende productdefinities naast elkaar bestaan, dan heeft dat voor- en nadelen.

Tot de voordelen behoort dat ze conflictreducerend zijn.

1. Het hanteren van variëteit in productdefinities kan ook doelmatig zijn omdat een organisatie zich het werk bespaart om alle producten te omschrijven op een eenduidige manier.
2. Een tweede voordeel van variëteit in prestatie-meting is dat een rijker beeld van de prestaties van ene organisatie wordt verkregen.
3. Een derde voordeel is de matiging van prikkels om te komen tot pervers gedrag van professionals.

4. Een vierde voordeel is de aantrekkelijkheid voor bestuurders resp. directies. Het gezag van de uitkomsten is groter bij het werken met variëteit in prestatiesystemen. Als door toelating van variëteit toch zichtbaar wordt dat een prestatie niet gehaald wordt, dan kan het voor professionals die de prestatie voortbrengen moeilijker worden om niet te proberen de prestatie wel te verbeteren.

Strategie 2: verdisconteer omstandigheden zoals coproductie en dynamiek.

Het is mogelijk bij een prestatie-meetsysteem van een overheidsorganisatie niet alleen de variëteit in prestaties voor het voetlicht te brengen om zo perverse effecten van een prestatie-meting te voorkomen of beperken maar het is ook mogelijk zichtbaar te maken of omstandigheden of turbulenties inwerken op het wel of niet bereiken van bepaalde prestatienormen. Dat voorkomt dat professionals het gevoel krijgen dat de verkeerde 'partijen' het (niet) halen van prestaties aangerekend wordt.

Strategie 3: een verbod op een monopolie op de betekenisgeving.

Bij prestatie-meting is een belangrijke vraag wie tot betekenisverlening en oordelen ('meaning making rights') bevoegd is. Alles zal afhangen van de vraag of sprake is van een vrijwillige prestatie-meting of een interbestuurlijke verplichte? En van de vraag of prestatie-meting een verhullende werking heeft. Een prestatie-meting kan op grond van twee mechanismen een *verhullende werking* hebben.

- 1) Eerste mechanisme dat bijdraagt aan een verhullend effect van prestatie-meting: Als de afstand tussen de cijferproducent en de cijferontvanger groot is, is het voor de ontvanger moeilijker om deze cijfers van betekenis te voorzien. De ontvanger staat verder af van het primair proces waardoor het juist zicht op wat de cijfers wel of niet betekenen verloren gaat. Dit verschijnsel namen we zelf waar binnen de Open Universiteit toen daar cijfers over studentvorderingen werden geproduceerd en de controller de analyse verrichtte.
- 2) Tweede mechanisme dat bijdraagt aan een verhullend effect van prestatie-meting: naarmate de distantie tussen degen die de cijfers produceert en degene die de productiecijfers ontvangt groter is, neemt 'de vermeende hardheid van de cijfers toe', aldus De Bruijn (2002: 150). De betekenis van het cijfers is relatief, maar wie verder van het primair proces afstand wordt minder gehinderd door de relativering. Bovendien nodigen cijfers ertoe uit om ze als onbetwistbaar te beschouwen. Neem de vergelijking van geweldplegingen in gevangenissen tegen personeel. Gevangenis A kent 10 geweldplegingen per half jaar per 100 werknemers, gevangenis B 25 en gevangenis C 3. Gevangenis B doet het dus slechter dan A en C. wie niet meer weet, komt tot die conclusie. Maar als men in beschouwing neemt dat B de enige gevangenis is met lang gestraften en zwaar gestraften en dit aantal een

aandeel heeft van de helft van de totale populatie, wordt het anders. Want het zijn nadrukkelijk de lang gestraften die in een uitzichtloze situatie zitten omdat ze vermoedelijk nooit vrij komen, die tot geweldplegingen overgaan. Gevangenis C is een doorgangshuis naar de gevangenis waar men langer moet verblijven. Zo gezien kan het cijfer voor B gerelativeerd worden.

Een monopolie op de keuze van cijfers en op de beoordeling ervan is dus gevaarlijk. Het geeft een risico op vertekende conclusies. Advies: vermijd dit. Zorg voor afspraken over spelregels die betrekking hebben op het gebruik van prestatiecijfers in een organisatie. Het is mogelijk om de cijfers te bespreken in een managementteam zodat de betekenisverlening- perspectieven van een controller en hoofden duidelijk wordt.

Strategie 4: beperking van functies en fora.

Het is mogelijk om cijfers over prestaties te verzamelen en intern te bezien voor alle organisatiedelen van een en dezelfde organisatie (forum 1), of prestaties van afdelingen van een gemeente a en gemeente b te verzamelen en vergelijken en deze te bezien in gemeenteraden (forum 2), of prestaties van eenzelfde afdeling 1 te vergelijken voor een groot aantal gemeenten (vergelijking van burgerzaken in tien gemeenten) en deze te laten bekijken door een hoofd van een afdeling (forum 3). Enzovoorts. Daarmee wordt duidelijk dat het domein van de prestatiemeting kan verschillen en het forum dat de uitkomsten duidt en bespreekt. De wet van De Bruijn luidt:

'hoe meer functies een productiecijfer moet vervullen en hoe meer verschillende fora een productiecijfer zullen gebruiken, des te groter de kans dat de paradox van de toenemende perverse effecten zich manifesteert' (De Bruijn, 2002: 152).

De oplossing om dit te vermijden is dat bestuurders zich hiervan expliciet rekenschap geven en afspraken maken over functies en fora. Een eenzijdige wijziging van functies en fora moet worden vermeden, houdt De Bruijn ons terecht voor. Voor professionals is het aantrekkelijk vooraf zekerheid te hebben over waartoe prestatiemeting dient en voor welk forum die meting bestemd is. Prestatiemeting moet gebaseerd zijn op *vertrouwen*, dat noodzakelijk is voor leren.

Strategie 5: strategische selectie van producten.

Prestatiemeting bij een scala van producten heeft allerlei nadelen zoals de grote hoeveelheid informatie (door de bomen het bos niet meer zien). Daarom kan het zin hebben dat een organisatie allereerst eens een onderscheid maakt in soorten producten: hoofd- en bijproducten, of in primaire en secundaire producten, of in meervoudige en enkelvoudige producten. In werkelijkheid is vervolgens een onderscheid te maken tussen soorten producten waarop prestatiemeting losgelaten wordt en te variëren naar functie van de prestatiemeting. Voor de primaire producten vindt bijvoorbeeld meting plaats van de prestaties met de doelen

transparantie, leren, oordelen en afrekenen, voor de secundaire producten is de doelstelling beperkter.

Denk hierbij bijvoorbeeld aan onderwijs voor kansarmen door de Open Universiteit die is opgericht om hoger onderwijs voor kansarme studenten mogelijk te maken. Het ministerie zou deze prestatiemeting kunnen eisen en daarbij ook de meting van prestaties van anderen kunnen betrekken. Aangezien de OU ook onderzoek doet maar dit geen primaire taak is zou prestatiemeting hier kunnen geschieden vanuit een lagere doelstelling, bijvoorbeeld transparantie bevorderen en leren.

De Bruijn (2002) houdt ons voor dat een dergelijk onderscheid in soorten producten waarbij prestatiemeting geschiedt, de 'overload' aan informatie beperkt. Bovendien kan een externe financier van een organisatie, zeg een ministerie, door een intelligente selectie van producten toch invloed houden op een organisatie, zoals een universiteit. In het algemeen is het dus niet zonder meer aan te raden om zware prestatiemeting bij alles te plegen. Daarbij is ook te overwegen of prestatiemeting van meervoudige of juist enkelvoudige producten positieve leereffecten heeft op onderdelen van een organisatie. Het is mogelijk dat een selectie van voor prestatiemeting bestemde producties goed valt bij professionals.

Strategie 6: management van de concurrerende benaderingen 'product' en 'proces'.

In het voorgaande is volop gebleken dat er kritiek mogelijk is op prestatiemeting. Prestatiemeting kan een heel beperkt zicht geven op prestaties als het productieproces in een zwarte doos blijft en de kostentoe rekening onduidelijk is. Welke activiteiten heeft een werknemer verricht om het resultaat te bereiken? Hoeveel tijd is besteed in geval van co-productie aan samenwerking of overgave van een dossier? Welke omstandigheden hadden een positieve of negatieve invloed? Wil een gebruiker cijfers kunnen interpreteren dan zal veelal overleg nodig zijn met professionals over de totstandkoming en de omstandigheden. Vindt dit overleg plaats dan zal het begrip voor de omgang met prestatiemeting toenemen. Prestatiecijfers kunnen binnen een organisatie tussen centrale leiding en decentrale hoofden een voortuig zijn voor gesprek over werk en prestaties terwijl de dialoog ook weer de risico's van prestatiemeting kan beperken. Zo wordt een procesbenadering in een systeem van prestatiemeting geïncorporeerd (De Bruijn, 2002: 156).

Aanvulling: strategie 7?

Zou er nog een andere strategie, strategie 7, mogelijk zijn? Ja, het is mogelijk om de functies van prestatiemeting te beperken. Denk hierbij aan het niet hanteren van prestatiemeting voor oordelen of afrekenen. Het is mogelijk dat prestatiemeting die teruggekoppeld wordt naar professionals juist voor hen tot leerervaringen kan opleveren.

12 Typen prestatie­meting

Er valt ook enige kritiek te formuleren op de beschouwing van Hans de Bruijn die wij hier deels volgden. Steeds blijkt de vraag om de hoek te komen: om welk type prestatie­meting gaat het nu? Gaat het om een studieuze prestatie­meting door een Sociaal en Cultureel Planbureau gebaseerd op bestaande cijfers over zeg verpleeghuizen, of gaat het om prestatie­meting van vijf sociale diensten die onderling afspreken te komen tot een prestatievergelijking of is sprake van een systeem van prestatieafspraken tussen een hogere overheid en een zelfstandig bestuursorgaan als een universiteit?

Bij prestatie­meting en ook de effecten moet er sterk rekening mee gehouden worden of sprake is van een bepaald type prestatie­meting. We onderscheiden:

- Prestatie­meting van een organisatie of van meerdere organisaties tegelijk. Een vergelijking van organisaties gericht op verbetering en good or best practice heet benchmarking..
- Horizontale prestatie­meting (vergelijking tussen zelfde organisaties of delen daarvan) of verticale prestatie­meting (prestatie­meting in een interbestuurlijke stelsel).
- Prestatie­meting gebaseerd op vrijwillige of verplichte deelname.
- Prestatie­meting die top-down gestuurd wordt of bottom-up.
- Prestatie­meting met het doel transparantiebevordering, leren, oordelen of afrekenen.

In werkelijkheid komen combinaties voor.

Neem de visitatie van studierichtingen in het hoger onderwijs. Hier is sprake van een verplichting. Geen studierichting kon er zich in de jaren negentig van de vorige eeuw aan onttrekken. Elke universiteit stelde een zelfstudie op op basis van een VSNU-protocol, waarna een visitatiecommissie de zelfstudie beoordeelde, een bezoek ter plaatse aflegde en gesprekken voerde, voorlopige bevindingen opstelde waarna een definitief rapport werd geproduceerd. Hier is sprake van verplichting, van een combi van top-down en bottom-up en van een verticaal geheel want er volgt een rapport dat weer door de inspectie hoger onderwijs wordt beoordeeld. Er is ook sprake van transparantiebevordering, 'leren' en oordelen. Afrekenen blijft uit tenzij een studierichting geheel beneden de maat is.

Schema: Typen prestatie­meting

Typen prestatie­meting	Deelname	Transparantie-functie	Functie: Leren	Functie: Verantwoording/ Oordelen	Functie: Verantwoording en afrekenen/ sanctietoepassing
Horizontale prestatie­meting	Vrijwillige deelname	Bijv. vergelijking van vijf sociale diensten	Bijv. vergelijking van vijf sociale diensten		
Horizontale	Verplichte				

prestatie­meting	deelname				
Verticale (interbestuurlijke) prestatie­meting	Vrijwillige deelname				
Verticale (interbestuurlijke) prestatie­meting	Verplichte deelname				

Uitkomsten

Tot welke uitkomsten leidt prestatie­meting nu? We geven een voorbeeld. Veger (2002) deed onderzoek naar resultaatgerichte bedrijfsvoering. Het kenmerk daarvan is een combinatie van integraal management, centrale sturing op hoofdlijnen, afspraken over te behalen prestaties, verantwoording over behaalde prestaties en een systematische evaluatie. Leden van managementteams van tien agentschappen, in totaal 37 personen, deden mee aan een (klein) onderzoek waarin onder meer is gekeken naar de betekenis van de overgang naar resultaatgerichte bedrijfsvoering voor de organisatie van de rijksoverheid. Vanuit het gegeven schema is deze prestatie­meting te typeren als enigszins horizontaal (een agentschap blijft deel uitmaken van een ministerie maar komt toch op afstand te staan), gedwongen, gericht op transparantie, leren, verantwoording en wellicht ook afrekenen. Deze prestatie­meting sluit aan op de VBTB-operatie, zoals de nota 'Van beleidsbegroting naar beleidsverantwoording' laat zien. De regering-Kok II wil meer inzicht in de relatie tussen geleverde prestaties en de daarmee gemoeide kosten. Aangezien agentschappen als een gerichtheid op prestaties kenden, dat was veelal zelfs een argument voor de instelling, is het interessant om te zien of deze agentschappen de gewenste prestaties ook daadwerkelijk leveren.

De scores liepen op een vijf­puntsschaal uiteen van 'nee, beslist niet' (1), naar 'nee, eigenlijk niet' (2), 'soms wel, soms niet' (3), tot 'ja, eigenlijk wel' (4) en 'ja, beslist' (5). De scores staan in de tabel. Er bleek sprake van forse afwijkingen van het gemiddelde. Dat betekent dat er sprake is van forse spreiding. Tal van agentschappen slaagden er in 2000 nog niet in om het merendeel van hun afspraken na te komen. De respondenten geven aan dat men verbeteringen in de bedrijfsvoering over het algemeen wel 'een beter inzicht in de behaalde prestaties heeft gekregen' (Veger, 2002: 8).

Tabel: Gemiddelde scores voor de prestaties van tien onderzochte agentschappen op een vijf­puntsschaal vlg. Veger

vraagstelling	Gemiddelde oordeel
Prestaties op het gebied van personeelstevredenheid	3.08
Prestaties op het gebied van tevredenheid klanten-opdrachtgevers	3.41
Prestaties op het gebied van financieel-economische doelstellingen	3.57
Prestaties op het gebied van productie	3.41
Prestaties op het gebied van organisatieontwikkeling	3.57

De tien onderzochte agentschappen kunnen over het algemeen nog niet goed inhoud geven aan resultaatgericht management. De oorzaken daarvoor liggen meer in de interne en externe invloeden dan op de wijze waarop afspraken in overleg met het naast hogere niveau plaatsvinden.

Literatuur

- Behn, R. & P.A. Kant, Strategies for avoiding the pitfalls of performance contracting, in: Public Productivity and Management Review, 1999, nr. 4.
- Bordewijk, P. en H. Klaassen, Wij laten ons niet kennen, VNG, Den Haag, 2000.
- Bruijn, H. de, Prestatiemeting in de publieke sector, Lemma, Utrecht, 2001.
- Bruijn, J. de, Prestatiemeting in de publieke sector – Strategieën om perverse effecten te neutraliseren, in: Bestuurswetenschappen, 2002, nr. 2, pp. 139-160.
- Eggink, E. & J. Blank, Verstandig verzorgd, SCP, Den Haag, 2001.
- Haselbekke, A.G. e.a., Prestaties tellen, VNG Uitgeverij, Den Haag, 1990.
- Hilhorst, P., De wraak van de publieke zaak, De Balie, Amsterdam, 2001.
- Huppes, W., Kwaliteit resultaatanalyse systeem, Achtergrondstudie bij het Advies Volksgezondheid en zorg, Zoetermeer, 2001.
- Leeuw, F.L., Performance auditing, new public management and performance improvement: questions and answers, in: Accounting, Auditing and Accountability Journal, 1996, nr. 2.
- Oosterom, R.P. van, Resultaatgericht management bij uitvoeringsorganisaties van de rijksoverheid, in: Management in overheidsorganisaties, april 2002, C2150.
- Osborne, D. & T. Gaebler, Reinventing government, Reading, 1992.
- Thiel, S. van, en F.L. Leeuw, The performance paradox in the public sector, in: Public Productivity and Management Review, march 2002, pp. 267-281.
- Twist, M.J.W. van, e.a., Beelden van bestuur – Berenschot trendstudie, Utrecht, 2002.
- Veger, R., Agentschappen en resultaatgerichte bedrijfsvoering, in: Openbaar Bestuur, juni-juli 2002, nr. 6/7, pp. 7-10.
- Vierbergen, R. & A. Brouwer, Bestuurlijk relevante prestatiegegevens in de publieke sector, in: Overheidsmanagement, 2002, nr. 1, pp. 4-7.
- Vogelzang, P., Efficiency politie is moeilijk in cijfers te vangen, in: Volkskrant, 6 april 2002, p. 9.

12 Afsluiting

De kwaliteit van een private – op winst gerichte - organisatie kan worden afgemeten aan de vraag naar producten. Geen vraag, geen privaat bedrijf. Een manier om de kwaliteit van een overheidsorganisatie te beoordelen is moeilijker want burgers hebben doorgaans geen andere keuzemogelijkheid dan verhuizen. Marktwerking ontbreekt. Een mogelijkheid die wel bestaat, is bijvoorbeeld beleidsevaluatie van een beleidsprogram. Binnen het geheel van mogelijkheden tot beleidsevaluatie is een manier om de doelbereiking na te gaan van een specifiek programma en te onderzoeken of het een beleidsprogram was dat bijdroeg aan de doelbereiking. Dit type staat bekend als een *beleidseffectiviteitsstudie*. Wie zicht wil krijgen op de kwaliteit van een gehele organisatie zou dus alle beleidsprogramma's moeten evalueren. Wie een zelfde organisaties wil beoordelen zou dus al die beleidsprogramma's moeten evalueren. Als het niet gebeurt op basis van een beleidseffectiviteitsmodel kan het nog volgens een *goal free approach* of anderszins.

Hoe dan ook, dit is bijna een onmogelijke opgave voor scholen, musea, gevangenissen, sociale diensten, waterschappen en dergelijke. Dus wordt gezocht naar next best-aanpakken. Zo lichtten medewerkers van het departement van SZW sociale diensten door. En een inspectie van BZK ziet toe op bepaalde gemeenten met een zwakke financiële basis, zoals de artikel 12-gemeenten. En er bestaan *visitaties*.

Benchmarking is slechts één benadering van kwaliteitszorg, en wel een die aansluit bij het (bedrijfsmatig gerichte) denken in termen van *new public management*. Het is een aanpak die veel methodische kennis vereist om de valkuilen te omzeilen. Hierbij komen gebruikers niet aan bod. Er is geen sprake van interactiviteit. De aanpak is er binnen grenzen. Wie belevingsinformatie over en emotie of ander gevoel bij dienstverlening van groot gewicht acht, heeft aan benchmarking doorgaans niets omdat de beleving niet meegenomen wordt in de vergelijking.

Benchmarking is te zien als een verschijnsel met drie kenmerken: een vergelijking van organisaties qua throughput, en/of output en/of outcomes, gericht op continue verbetering vanuit het perspectief van best practices. De toepassing van deze benadering is niet zonder problemen. Onder voorwaarden kan benchmarking zinvol zijn. Een van de genoemde voorwaarden is dat de aanpak alleen kan worden toegepast door onderzoekers met benul van methoden en technieken. Is er niet een softere vorm van organisatievergelijking mogelijk? Ja, bijvoorbeeld een vergelijking van sociale diensten kan nuttig zijn om te zien of de snelheid van behandeling van aanvragen resp. de duur van de verstrekking relatief lang is in vergelijking met andere diensten (verg. een studie van Van Westerlaak). Dan ligt de ambitie laag.

Bij benchmarking is sprake van een aanpak met verschillende functies en een aantal valkuilen. Aangegeven is welke. Er zijn strategieën genoemd om aan perverse effecten te ontkomen.

Wat in onze beschouwing ontbrak en in nieuw onderzoek aan bod kan komen is: wat is het gebruik en de doorwerking van benchmarkresultaten? (zie over benutting publicaties van Van de Vall, Korsten, Bekkers, Hupe, Weiss, Landry). Nog onbewezen is of benchmarking echt tot verbetermanagement leidt. Niet uit te sluiten is dat veel benchmarksstudies leiden tot de uitroep: 'bevestigd wordt wat we al dachten'.

Literatuur over benchmarking

- Aardema, H. en A.F.A. Korsten, De staat van de gemeente. Op weg naar een handzame landelijke gemeentemonitor, 2005.
- Aardema, H., Stille waarden, OU/BMC, den Haag, 2005.

- Admiraal, R.J. en van Helden, G.J., De balanced scorecard binnen een benchmarkcontext: bedrijfsvergelijking zuiveringsbeheer bij Nederlandse waterschappen, in: MAB, jrg. 77, 2003, nr. 6, pp 279-286.
- Ammons, D. e.d., Proper Mentality for Benchmarking, in: Public Administration Review, vol. 59, 1999, nr. 2, pp. 105-119.
- Bakker, W. en K. Yesilkagit (red.), Publieke verantwoording, Boom, Amsterdam, 2005.
- Behn, R.D., Rethinking democratic accountability, The Brookings Institution, Washington, 2001.
- Behn, R.D. & P. Kant, Strategies for avoiding the pitfalls of performance contracting, In: Public Productivity and Management Review, 1999, nr. 4, pp. 470-489.
- Bekkers, V. e.a., Doorwerking van strategische beleidsadviesing, Rotterdam/Tilburg, 2004.
- Bentlage, F., e.a., De excellente overheidsorganisatie – De invoering van Total Quality Management in de publieke sector, Kluwer, Deventer, 1998.
- Berg, C. e.a. (red), Moderne bedrijfsvoering bij de overheid, Sdu, Den Haag, 2003.
- Blank, J. en P. Vrancken, Produktiviteit in de publieke sector, in: Jaarboek overheidsuitgaven 1992, Academic Service, Schoonhoven, 1991, pp. 133-149.
- Blank, J., Benchmarking of de kunst van het vergelijken, in: ESB, 29 mei 1998, pp. 432-435.
- Blank, J.L.T. en E. Eggink, Doelmatigheid van verpleeghuizen, in: Openbare Uitgaven, 1998, nr. 1, pp. 14-22.
- Blank, J.L.T. en E. Eggink, Zuinig op zorg- Een empirisch onderzoek naar de produktiviteitsstructuur van verpleeghuizen, SCP/Vuga, Rijswijk, 1996.
- Blank, J.L.T., Benchmerken of de kunst van het vergelijken, in: ESB, 29 mei 1998, p.p. 432-435
- Blank, J.L.T., Produktiviteitsonderzoek naar de publieke sector, in: Groot, H. de, en C.G.M. van Oosteren (red.), De toekomst van de publieke sector, Vuga, Den Haag, 1994, pp. 123-137.
- Bogt, H.J. ter & Helden, G.J. van, Kwaliteit van prestatiemeting bij gemeentelijke overheden, in: Bestuurskunde, jrg. 3, 1994, nr.1, pp. 2-12.
- Bogt, H.J. ter, Managementvernieuwingen bij de overheid, Groningen, 2005.
- Bordewijk, P. en H. Klaassen, Wij laten ons niet tellen, VNG Uitgeverij, Den Haag, 2000.
- Bordewijk, P.&Klaassen, H.L., Afrekenen met de productbegroting, in: B&G, januari 2001, pp. 16-21
- Bouckaert, G., en T. Auwers, Prestaties meten in de overheid, Brugge, 1999.
- Bowerman, M., Francis, G., Ball, A. & Fry, J., The evaluation of benchmarking in UK local authorities, in: Benchmarking – An international Journal, vol. 9, nr. 5, pp. 429-449.
- Bruijn, J.A. de, Prestatiemeting in de publieke sector, Lemma, Utrecht, 2001.
- Bruijn, J.A. de, Prestatiemeting in de publieke sector: strategieën om perverse effecten te neutraliseren, in: Bestuurswetenschappen, 2002, nr. 2, pp. 139-160.
- Bruijn, J.A. van, Wendel de Joode, R., Voort, H.G. van der, Potentials and risks of benchmarking, in: Journal of Environmental Assessment Policy & Management, vol. 6, 2004, nr. 3, pp. 289-309.
- Bullivant, J.R.N., Benchmarking for continuous improvement in the public sector, Longman, Essex, 1994.
- Camp, R.C., Benchmarking – Het zoeken naar de beste werkmethode die leiden tot superieure prestaties, Kluwer, Deventer, 1992.
- Commissaris, E.J. en Hedde, M. ter, Gluren bij de burens – Benchmarking kritisch beschouwd, Webside Berenschot, 2006
- Commissie Doelmatigheid Riagg-activiteiten, eindrapport, 1998.
- Commissie Toekomst Lokaal Bestuur, Wil tot verschil, VNG, gemeenten in 2015, Den Haag, 2006.
- Davies, P., The burgeoning of benchmarking in British local government: the value of “learning by looking” in the public services, in: Benchmarking for Quality Management and Technology, vol. 5, 1998, nr. 4, pp. 260-270.
- Dolmans, L.J.F., Algemene Rekenkamer, in: Groot, H. de, en C.A. de Kam (red.), Jaarboek overheidsuitgaven 1996, Academic Service, Schoonhoven, 1995.
- Dorsch, J.J., Yasin, M.N., A framework for benchmarking in the public sector, in: International Journal of Public Sector Management, vol. 11, 1998, nr. 2-3, pp. 91-116.

- Drew, S., From knowledge to action, the impact of benchmarking on organizational performance, in: Long Range planning, vol. 30, 1997, nr. 3, pp. 427-441.
- Driel, M. van, S. van Wersch e.a., Beoordeling door cliënten - Handreiking voor kwaliteitsbeoordeling door cliënten in zorginstellingen, Nederlandse Vereniging voor Kwaliteit en Zorg, Uitgeverij SWP, Utrecht, 1998.
- Eggink, E. & J. Blank, Verstandig verzorgd, SCP, Den Haag, 2001.
- Elias, J., J.Th. van Ieperen e.a., Doelmatigheid in de thuiszorg, Bakkenist Management Consultants, Amsterdam, 1997.
- Farrell, M.J., The measurement of productivity efficiency, in: Journal of the Royal Statistical Society, 1957, pp. 253-281.
- Ferlie, E. e.a. (eds.), The Oxford Handbook of Public Management, Oxford University Press, Oxford, 2005.
- Groot, H. de e.a., Benchmarking in de publieke sector, Sdu, Den Haag, 2004.
- Groot, H. de, Algemene Rekenkamer in bedrijf, in: Openbare Uitgaven, 1998, nr. 5, pp. 241-245.
- Groot, H. de, Goudriaan, R., Hoogwout, M., Jong, A. de, Roerstamper, R., Benchmarking in de publieke sector: vergelijken van prestaties als managementinstrument, Public Controlling Reeks nr. 6, Sdu, Den Haag, 2004.
- Hakvoort, J. & H. Klaassen, Benchmarking in non profit organisaties, in: Beleidsanalyse, 1999, nr. 3, pp. 11-20.
- Harrington, H. & J. Harrington, High performance benchmarking, New York, 1996.
- Harten, W.H. van, Bouwen aan een kwaliteitssysteem in de revalidatiezorg, Rotterdam, 1997 (diss.).
- Haselbekke, A.G. e.a., Prestaties tellen, VNG Uitgeverij, Den Haag, 1990.
- Helden, G.J. van, Brouwer, N., Benchmarking en prestatieverbetering in de publieke sector: ervaringen met de bedrijfsvergelijking zuiveringsbeheer, in: MAB, jrg. 79, nov. 2005, nr. 11, pp. 574-583.
- Helden, G.J., Tillema, S., In search of a benchmarking theory for the public sector, in: Financial Accountability & Management, vol. 21, nr. 3, pp. 337-361.
- Helder, S., Op weg naar structurele inzet benchmark-instrumenten, in: Bank&Gemeente, maart 2001, pp. 12-14.
- Heuvelhof, E.F. ten, Nauta, C.M., Milieu-effectrapportage als leerproces, Beleidswetenschap, jrg. 11, 1997, nr. 12, pp. 25-40.
- Hilhorst, P., De wraak van de publieke zaak, De Balie, Amsterdam, 2001.
- Hood, C.A., Public management for all seasons= Public administration 69[1], 3/19, 1991
- Hoogwout, M., Het ontwerpen van overheidsbenchmarks, in: Public Controlling, vol. 2, 2004, nr. 2, pp. 32-36.
- Hughes, O., Public management & administration - An Introduction, St. Martin's Press, Basingstoke, 1994.
- Hupe, P., Meer, F.B. van der, Cruciale koppelingen; over doorwerking van emancipatie-effectrapportages in beleidsprocessen, in: Beleidswetenschap, jrg. 20, 2006, nr. 2, pp. 3-21.
- Huppel, W., Kwaliteits Resultaat analyse Systeem, Zoetermeer, 2001 (Achtergrondstudie bij het Advies Volksgezondheid en zorg)
- Keehley, P., e.a., Benchmarking for best practices in the public sector, Jossey-Bass Publishers, San Francisco, 1996.
- Keen, J., On the nature of audit judgements: the case of value for money studies, in: Public Administration, vol. 77, 1999, nr. 3, pp. 509-527.
- Kerr, C. A. e.a., Best-practice measures of resource utilization for hospitals: a useful complement in performance assessment, in: Public Administration, vol. 77, 1999, nr. 3, pp. 639-651.
- Korsten, A.F.A., Wat is goed genoeg?, Kobra, Amsterdam, 1983.
- Kouzmin, A., Loffler, E., Klages, H., Korac-Kakabadse, N., Benchmarking and performance measurement in public sectors, in: International Journal of Public Sector Management, vol. 12, 1999, nr. 2, pp. 121-145.

- Kuhry, B. & A. van der Torre, Maten voor gemeenten - Een verkennende studie naar prestatieindicatoren in het kader van het 'Plan van aanpak Transparantie in de financiële verhouding', SCP, juni 2001.
- Kuhry, B., Veldheer, V., Maten voor gemeenten 2005, Een analyse van de prestaties van de lokale overheid, SCP, en Haag, 2005.
- Landry, R., Lamari, M., Amara, N., The Extent and Determinants of the Utilization of University Research in Government Agencies, *Public Administration Review*, vol. 63, 2003, nr. 2, pp. 195-205.
- Laughlin, K., S. Osborne & E. Ferlie (eds.), *New public management*, Routledge, Londen, 2002.
- Leeuw, F.L., Onbedoelde neveneffecten van outputsturing, controle en toezicht?, in: Raad voor Maatschappelijke Ontwikkeling, Aansprekend burgerschap, Den Haag, 2000, p. 149-171.
- Leeuw, F.L., Performance auditing, new public management and performance improvement: questions and answers, in: *Accounting, Auditing and Accountability Journal*, 1996, nr. 2.
- Lovell, C.A.K., *Productivity frontiers and productivity efficiency*, in: *The measurement of productive efficiency: techniques and applications*, Oxford University Press, 1993.
- Maas, M. e.a., *Benchmarking bij overheid en non-profitorganisaties*, Elsevier, Den Haag, 1998.
- Maas, M., H. Witjes & I. Zaat, *Benchmarking bij overheid en non-profitorganisaties*, Elsevier, Den Haag, 1998.
- McLaughlin, K. e.a., *New Public Management current trends and future prospects*, London: Routledge, Londen, 2002.
- Meer, F.B. van der, Vries, G.J.D. de, Vissers, G.A.N., Evaluatie en leerprocessen bij de overheid; de rol van institutionele condities. *Beleidswetenschappen*, jrg.14, 2000, nr. 3, pp. 253-277.
- Meurs, P. en T. van der Grinten, *Gemengd besturen – Besturingsvragen en trends in de gezondheidszorg*, Academic Service, /Sdu, Den Haag, 2005.
- Ministerie van BZK en VNG, *De Gids Prestatievergelijkingen en Benchmarks binnen de openbare sector*, zie www.andereoverheid.nl.
- Morgan, C. e.a., *Total Quality Management in the Public Sector*, Open University Press, 1994.
- Mulder, H.P. et al., gebruik van beleidsevaluatieonderzoek bij de rijksoverheid, in: *Beleidswetenschap*, jrg. 5, 1991, pp. 203-228.
- Nelissen, N.J.M. & P. de Goede, Benchmarking: vergelijken en verbeteren, in: *Bestuurskunde*, 1999, nr. 2, pp. 54-71.
- Nutt, P.C., Decision-making success in public, private and third sector organizations, finding sector dependent best practice, in: *Journal of Management Studies*, vol.. 37, 2000, nr. 1, pp. 77-108.
- Oosterom, R.P. van, Resultaatgericht management bij uitvoeringsorganisaties van de rijksoverheid, in: *Management in overheidsorganisaties*, april 2002, C2150.
- Osborne, D. & T. Gaebler, *Reinventing government*, Reading, 1992.
- Patton, M.Q., *Utilization-focused evaluation*, Sage, Thousand Oaks, 1997.
- Peursen, K.A. van, M. Pratt & S. Lawrence, Health management performance - A review of measures and Indicators, in: *Accounting, Auditing and Accountability Journal*, 1995, nr. 5, pp. 34-70.
- Pollit, C. & G.B. Bouckaert, *Quality improvement in European public services: concepts, cases and commentary*, Sage, Londen, 1995.
- Pollitt, C., M. Cave & R. Joss, International benchmarking as a tool to improve public sector performance – A critical overview, in: *OECD, Performance measurement in government – Issues and illustrations*, Paris, 1994.
- PricewaterhouseCoopers en Berenschot, *Benchmarkonderzoek thuiszorg biedt aanknopingspunten voor instellingen en overheid*, Utrecht, maart 1999.
- PricewaterhouseCoopers, *Integratie van milieubeleid: degraderen of excelleren - Lessen uit een benchmark van externe integratie bij vijf grote gemeenten*, Utrecht, 2000.
- Prins, S.J., *Zoek, rapporteer, vergelijk en verbeter! – Een verkennend onderzoek naar de zin en mogelijkheden van het gebruik van benchmarks binnen het prestatie-meetsysteem*, Moret Fonds, Rotterdam, 1997.

- Raad voor het openbaar bestuur, Presteren door leren, Benchmarken in het binnenlands bestuur, Den Haag, 2002.
- Rainey, H.G., Backoff, R.W., & C.Levine, Comparing public and private organizations, in: Public Administration Review, vol. 36, 1976, nr. 2, pp. 233-244.
- Serban, A. & J. Burke, Meeting the performance challenge - A nine-state comparative analysis, in: Public Productivity and Management Review, 1998, nr. 2, p. 157-176.
- Severijnen, P., Benchmarking voor gemeenten volgens de vergelijkingsmethode, in: De Overheidsmanager, jg. 7, oktober 2000, nr. 10.
- SGBO, Verkennend onderzoek kwaliteitshandvesten gemeenten, Den Haag, 2005.
- Spendolini, M., The Benchmarking Book, AMACOM, New York, 1992.
- Swanborn P.G., Evalueren, Boom, Amsterdam, 1999.
- Swanborn, P.G., Case-study's: Wat, wanneer en hoe, Boom, Meppel, 1996.
- Thiel, S. van, en F.L.Leeuw, The performance paradox in the public sector, in: Public Productivity and Management Review, march 2002, pp. 267-281.
- Tillema, S., Gebruik van bechmarkinginformatie in de publieke sector, in: MAB – maandblad voor Accountancy en Bedrijfseconomie, jrg. 80, juni 2006, nr. 6, pp. 300-309.
- Titulaer, R., K. Lammers e.a., Kwaliteit bij burgerzaken: stap voor stap op weg naar rekenschap - Rekenschapsonderzoek naar resultaten en verklaringen van de kwaliteit van de dienstverlening van afdelingen burgerzaken, mei 2001.
- Twist, M.J.W. van, e.a., Beelden van bestuur – Berenschot trendstudie, Utrecht, 2002.
- Vall, M. van de, Sociaal Beleidsonderzoek, Samson, Alphen, 1980.
- Veldkamp, C.M.H., Beelden van kwaliteit - Normanalyse als bruikbaar hulpmiddel bij kwaliteitszorg in verpleeghuizen, Universiteit Twente, Enschede, 1998 (diss.).
- Verheyen, P.A. (red.), Non-profit in bedrijf, Kluwer, Deventer, 1994.
- Verheyen, P.A., De economie van non-profitinstellingen, Kluwer, Deventer, 1993.
- VNG, Gemeenten benchmarken. VNG Uitgeverij, den Haag, 2005.
- Vries, J. de, en J. van der Togt, Benchmarking in 9 stappen, Kluwer, Deventer, 1995.
- Waalewijn, P. e.a., Benchmarking in Nederland op zoek naar het ideale proces, in: Holland Management Review, 1996, nr. 51, pp. 57-69.
- Watson, G.H., Strategisch benchmarken, Scriptum, Schiedam, 1998.
- Weiss, C.H., Bucuvalas, M.J., Sociale science research and decision-making, New York Guildford, Columbia University Press, 1980.