

# Mythen over benchmarking

*Feiten en ficties rond het vergelijken van prestaties tussen organisaties, met bijzondere aandacht voor gemeenten*

door

Arno F.A. Korsten, Klaas Abma en Anne-Douwe van der Meer<sup>1</sup>, 11 juni 2013<sup>2</sup>

## Inhoud

### 1 Inleiding

### 2 Benchmarking: achtergronden en eerste typering

### 3 Kenmerken van benchmarking

### 4 Ervaringen met verschillende benchmarkingtrajecten: verdiepende impressies

### 5 Veronderstellingen achter benchmarking

### 6 Uitkomsten van benchmarking: enkele casus

### 7 Bespreking van benchmarkrapportages: gebruiken, leren en verbeteren?

### 8 Valkuilen bij benchmarking

### 9 Enkele conclusies

---

<sup>1</sup> Prof.dr. A.F.A. (Arno) Korsten is honorair hoogleraar Bestuurskunde van de lagere overheden aan de Faculteit rechtsgeleerdheid van de Universiteit Maastricht en emeritus hoogleraar Bedrijfs- en bestuurswetenschappen, i.h.b. bestuurskunde aan de Faculteit managementwetenschappen van de Open Universiteit. Zie [www.arnokorsten.nl](http://www.arnokorsten.nl). Dr. K. (Klaas) Abma is als adjunct-gemeentesecretaris werkzaam in de gemeente Littenseradiel en als kartrekker 'decentralisaties' bij de gemeente Sud-West Fryslan. Drs. A.D. (Anne-Douwe) van der Meer is als controller werkzaam bij Bureau Jeugdzorg, Agglomeratie Amsterdam.

<sup>2</sup> Deze tekst van drie auteurs is opgesteld in het kader van het promotietraject van Anne-Douwe van der Meer.

# Mythen over benchmarking

*Feiten en ficties rond het vergelijken van prestaties tussen organisaties, met bijzondere aandacht voor gemeenten*

door

Arno F.A. Korsten, Klaas Abma en Anne-Douwe van der Meer, 9 juni 2013

## 1 Inleiding

Dit betoog gaat over benchmarking, een onderzoeksmethode op het vlak van resultaatsturing of 'performance management'. Benchmarking staat voor Spiegelen (vergelijken), Spieken (bij andere organisaties kijken) en Sprankelen (verbeteren). Bij benchmarking gaat het - anders gesteld - om meten, vergelijken en verbeteren.

Het is mogelijk om vanuit een wetenschappelijke doelstelling een benchmarkonderzoek op te zetten, maar hier gaat het over benchmarking als toegepast of praktijkgericht onderzoek ten behoeve van opdrachtgevers, zoals bij voorbeeld besturen van gemeenten of waterschappen. Dit benchmarkonderzoek is gericht op prestatievergelijking van een aantal organisaties, het rangordnen van prestaties op basis van scores, het zoeken naar goede of excellente praktijken en uiteindelijk om verbeteren. Het is dus een methode voor prestatie management met vier elementen, dus meer dan de genoemde 3 S'en.

Achter dergelijke praktijkgerichte benchmarking gaan hardnekkige ambities schuil die steeds maar weer gedeeld en herhaald worden. We noemen ze mythes. Zoals de mythe dat benchmarkonderzoekers ook in werkelijkheid steeds gericht zouden zijn op 'best practices' en die ook steeds weten te vinden en aan te wijzen. En de mythe dat benchmarking op basis daarvan daadwerkelijk kan en zal leiden tot prestatieverbeteringen bij overheidsorganisaties. Is daar op basis van onze ervaringen en tal van verrichte studies grond voor? Worden goede praktijken daadwerkelijk getraceerd en aangewezen, en zijn ze referentiepunt voor verbetering of is dat niet zo? Deze beschouwing gaat vooral maar niet uitsluitend over mythes. Blijken zal dat het nodig is om goed op de kwaliteit van benchmarking te letten en die zo nodig te verbeteren, zodat de ambities achter benchmarking in stand kunnen blijven.

We gaan in deze beschouwing op zoek naar een aantal kenmerken van benchmarking, veronderstellingen achter benchmarking en veel voorkomende valkuilen, die mogelijk een verklaring zijn voor het achterblijven van benchmarking bij verwachtingen zoals die in de literatuur te vinden zijn en vaak ook in praktijkgerichte documenten van overheidsbesturen.

We verwijzen naar een voorbeeld: het rapport *'Rotterdam vergelijkenderwijs – Benchmarks voor de verbetering van de bedrijfsvoering'* uit 2002 (Doorduyn e.a., 2002). Het gemeentebestuur geeft daarin aan dat prestaties van het gemeentebestuur te meten en te vergelijken zijn, en dat het nuttig is om ook van benchmarking gebruik te maken. 'Bij het gebruik bepaalt een organisatie een aantal ijkpunten waaraan ze zich wil meten, haar positie wil bepalen ten opzichte van een of meer andere vergelijkbare organisaties' (Doorduyn e.a., 2002: 1).

Met deze aanpak hopen we aanzetten te geven voor verbetering van benchmarkpraktijken.

## Doel

Het doel van deze beschouwing is om door vermelding van de *echte werkelijkheid* van benchmarking teleurstellingen over doelstellingen en verwachtingen rond benchmarking te voorkomen of beperken. Zijn de risico's bij benchmarking te ondervangen? Sommige wel, andere minder of niet. Blijken zal dat de verwachtingen over *'analysis by numbers'* (vrij naar Hood, 2007), zoals benchmarking ook wel eens wordt aangeduid, op basis van praktijkervaringen niet zelden te hooggespannen zijn. Kortom, wie succes wil hebben met benchmarking zal uit de goede vaatjes moeten blijven tappen.

### **Bronnen**

Wat we hier weergeven komt niet uit de lucht vallen. Deze beschouwing baseren we op literatuurstudie van benchmarking, op empirisch onderzoek in binnen- en buitenland (o.a. Korsten, 1983; Aardema en Korsten, 2005; Abma en Korsten, 2009) en eigen ervaringen als deelnemer aan een benchmark, opdrachtgever hiertoe, toezichhouder of 'gebruiker' van uitkomsten van benchmarking. We putten dus ook uit eigen ervaringen, 'inside information' en interacties in praktijkcontexten. Waar hier voorbeelden gegeven worden zullen die vooral betrekking hebben op gemeentelijke organisaties maar niet uitsluitend (Blank, Felsö en De Groot, 2011; Hulst en De Groot, 2011; Bouckaert e.a., 2011; Peeters & Verschraegen, 2013).

Alvorens iets te kunnen zeggen over deze aspecten en mythes is het nodig eerst aan te geven wat benchmarking is en hoe dit type vergelijkend onderzoek in de praktijk van decentrale overheden in Nederland en in andere landen in West-Europa voorkomt.

## **2 Benchmarking: achtergronden en eerste typering**

In de jaren tachtig van de vorige eeuw werd door tal van politieke, bestuurlijke en ambtelijke opinieleiders gedacht dat overheidsorganisaties eindelijk eens meer werk moest maken van bedrijfsmatigheid, dus van meer bedrijfsmatig werken. Ambtenaren moesten zakelijker, effectiever en doelmatiger gaan functioneren en overheidsorganisaties moesten ook meer 'publiek ondernemerschap' gaan tonen. De overheid moest het functioneren over een andere boeg gooien, zo heette het in de internationale bestseller *'Reinventing government'* (Osborne & Gaebler, 1992). Een daarop aansluitend adagium luidde: laat je als overheid inspireren door het succesvolle bedrijfsleven om zo op ideeën te komen over hoe het beter, anders of goedkoper kan (o.a. Korsten, 1994; Korsten en Noordegraaf, 1995; Teunisse, De Kiewit en Boshove, 2005). Privatisering was niet langer een vies woord. En ook het op afstand zetten van overheidsorganisaties in de vorm van zelfstandige bestuursorganen, met meer eigen beheermogelijkheden, paste in dat perspectief (Van Thiel, 2002; Jurriëns en Sylvester, 2005). Sommige universiteiten, als de Universiteit Twente, werden 'omgekat' tot een ondernemende universiteit die meer budget uit de markt kon en wilde halen. Gemeenten als Groningen en Delft maakten in de voetsporen van Tilburg (het Tilburgs model) de stap naar outputsturing en intern contractmanagement; met meer ruimte voor decentrale ambtelijke vrijheid van werken op basis van ambtelijke afspraken en uitbesteding.

### **Bedrijfsmatig denken**

Deze 'aanwijzingen' voor een meer bedrijfsmatige aanpak binnen de publieke sector culmineerden in een stroming of denkrichting onder de vlag van het *New Public Management* of *NPM* (Zifcak, 1994; Hood, 1995; Lane, 2000; Barzelay, 2001; Laughlin e.a., 2002; Van Helden en Jansen, 2002; Blank, Felsö en De Groot, 2011; Bekkers, 2012: 104). Uiteindelijk was deze beweging naar meer bedrijfsmatigheid bij overheidsorganisaties gericht op bevordering van publiek ondernemerschap; op transparantie en verantwoording; op beter leer- en oordeelsvermogen; en op innovatie (Bekkers, 2012: 105).

De belangrijkste omslag die NPM wilde maken, is dat de oude sturing op 'input' (budget) en 'throughput' (werkprocessen) plaats maakt voor sturing op 'output' (prestaties) en zo mogelijk op 'outcomes' (beleidseffecten).

Een voorbeeld betreft verkeersveiligheid. Wie hierbij als gemeentebestuur stuurt op input kan denken aan de hoeveelheid budget voor maatregelen op het gebied van verkeersveiligheid. Wie outputsturing wil, komt met gewenste beleidsresultaten of prestaties voor de dag, zoals de plaatsing van verkeersborden van 30 km max. rijsnelheid en de totstandkoming van verkeersluwe gebieden of de aanleg van een aantal verkeersdrempels in wijken met tot dan veel verkeersslachtoffers op onveilige plekken. Bij sturing op effecten gaat het er om meer verkeersveiligheid te bereiken door onder meer vermindering van het aantal verkeersslachtoffers in aangewezen gebieden.

NPM schoot in Nederland onder meer wortel in de vorm van de komst van productbegrotingen en later van programmabegrotingen, en van meer marktwerking.

Deze beweging naar meer bedrijfsmatigheid zou nog heel lang doorwerken. Ondanks dat Paul Bordewijk en Henk Klaassen (2000) in het kader van een analyse van kengetallen en productbegrotingen in *'Wij laten ons niet kennen'* met passie en kracht van argumenten stellen dat een gemeente geen bedrijf is. Ondanks een kritisch proefschrift als dat van Kor Grit (2000) over het op bedrijfsmatige leest geschoeide Tilburgs (gemeente)model en het concept van de ondernemende universiteit. Grit vroeg aandacht voor doorgeschooten bedrijfsmatigheid (ook Karsten en Van Veen, 1998).

### ***Vershil tussen overheidsorganisaties en private organisaties***

Dat een gemeente of provincie of waterschap geen bedrijf is (of niet primair zo gezien moet worden), ook al zijn organisatiedelen die tamelijk bedrijfsmatig werken aanwijsbaar, heeft redenen. We noemen er hier een aantal.

- 1) In de private sector is sprake van *marktwerking* en in de publieke sector *niet*. Burgers kunnen *op de markt kiezen* tussen aanbieders terwijl burgers in de publieke sector *niet* kunnen kiezen, tenzij ze verhuizen. In de publieke sector gelden andere principes dan marktprincipes, zoals sociale rechtvaardigheid, toegankelijkheid, gelijke behandeling, vrijheid, openbaarheid, redelijkheid en proportionaliteit, naleving van regels (uitvoeriger in Korsten, 2010a). Evenals in de private sector komen in de publieke sector overigens wel onderhandelingen voor.
- 2) Private bedrijven produceren goederen en diensten in een bepaalde kwantiteit tegen een *bepaalde prijs*. Wie de prijs niet betaalt, wordt *uitgesloten* van een goed of dienst. In het kader van overheidsbeleid vindt er vaak discussie plaats over de vraag of een goed of dienst inkomensafhankelijk kan worden verkregen of dat kwijtschelding of vrijstelling plaats heeft of een andere oplossing opportuun wordt geacht (Borst, Lako en De Vries, 2012: 16).
- 3) In de publieke sector komt overheidsbeleid tot stand dat vaak handelt over *collectieve keuzen*. Burgers worden als categorie of groep aangesproken, bij voorbeeld als belastingbetaler of huizenbezitter of energieverbruiker, en niet zozeer als consumenten. In de private sector worden burgers als *individuele* consumenten gezien (Borst, De Vries en Lako, 2012: 16). Pogingen om door middel van kartelvorming iets af te spreken voor consumenten in een bepaalde sector, bij voorbeeld een minimumbenzineprijs, zijn hier en daar verboden.
- 4) Een gemeentelijke organisatie produceert naast de verstrekking van identiteitsbewijzen en paspoorten - voorbeelden van concrete producten -, ook *slecht identificeerbare* en *niet-homogene producten uitbrengt*, waarvoor geen bedrijfseconomische criteria gelden of

waarbij deze criteria secundair zijn. Voor overheidsbesturen gelden tal van *andere criteria* dan in de private sfeer, mede omdat de response op uiteenlopende beleidsproblemen deels een andere is dan in de private sfeer. Denk aan weerbarstige probleem die alleen in netwerken van van elkaar afhankelijke 'actoren' zijn aan te pakken, zoals voetbalvandalisme (Baakman e.a., 2001; Ankersmit en Klinkers, 2008; Korsten, 2010a; Borst e.a., 2012).

- 5) Een regulier overheidsorganisatie als een gemeente is geen bedrijf omdat een gemeentebestuur niet uitsluitend producten voortbrengt of voorzieningen 'opricht' of 'verwerft' (o.a. terreinen, gebouwen) maar ook stuurt door middel van *processen*, zoals onderhandelingen of door op cohesie gericht overleg (communicatieve sturing naast of met juridische en/of economische sturing; denk aan pr, propaganda en convenanten). Soms zal een overheidsorganisatie ook symbolisch beleid formuleren terwijl een private organisatie dit minder gauw zal doen.
- 6) Een gemeentebestuur voert ook wel *bijzonder beleid* dat uit zuiver financieel-economisch perspectief en dus rendementsoverwegingen zou moeten worden geschrapt. Denk aan een beleid in het kader van de sociale werkvoorziening of het in stand houden van een zwembad. Soms legt een overheid geld bij omdat iets maatschappelijk van belang wordt geacht.
- 7) Wie gemeentebegrotingen bekeek en bekijkt, ziet dat weliswaar sprake is van pogingen om de drie w's in te vullen (wat willen we, wat doen we daarvoor en wat mag het kosten) en dat gepoogd wordt om op output te sturen of op outcome (gewenste effecten) maar ook dat dit maar heel beperkt lukt. Van uitsluitend *sturen op output* blijkt in gemeenten *geen* sprake en van sturen op outcome evenmin (Haselbekke, 1998; Bordewijk en Klaassen, 2000; Wassenaar en Verhagen, 2006; Bouckaert e.a., 2011).

Maar daarmee is de NPM-wind nog niet helemaal gaan liggen. NPM kan zeker een impuls hebben gehad die positief is geweest of nog is.

### ***Benchmarking past in NPM***

Benchmarking past(e) naadloos in deze meer bedrijfsmatige aanpak binnen de publieke sector. Maar benchmarking zou ook een geheel eigenstandige nuttige functie kunnen vervullen in het openbaar bestuur, los van NPM en met een relativering van bedrijfsmatigheid, mits rekening wordt gehouden met eisen van deugdelijk onderzoek, met de grenzen van benchmarking, met eigen kenmerken van overheidsprocessen en producten en met beoordelingscriteria voor overheidshandelen of niet-handelen (Callahan, 2007; Blank, Felsö en De Groot, 2011: 4). De bestuurskundigen Jan Hakvoort en Henk Klaassen (2004) nemen benchmarking dan ook op in hun meer tijdloze overzicht van '*bedrijfsvoeringstechnieken voor overheidsorganisaties*'. Anderen scharen benchmarking onder *prestatie management van overheden* (Van Dooren, Bouckaert & Halligan, 2010; Bouckaert e.a., 2011; Noordegraaf e.a., 2011). Tegen beide keuzen bestaat geen bezwaar mits onderkend wordt dat de 'techniek' van benchmarking flexibel moet worden toegepast en benchmarking een instrument kan zijn voor sturing door managers en bestuurders (Bouckaert e.a., 2011).

Benchmarking is, na de opkomst in het begin van de jaren negentig, jaren lang een populair managementinstrument bij gemeenten en in de rijksdienst gebleven, ook al is hier benadrukt dat er verschillen zijn tussen marktgerichte organisaties (lees: private bedrijven) en overheidsorganisaties. Immers, ook de overheid moet bij tal van taken en ambities letten op kwaliteit van product- en dienstverlening, op zorgvuldigheid van processen en andere kwaliteitsaspecten, en op doelmatigheid en prijs (Ammons, 1996; Bentlage e.a., 1998; Nelissen en De Goede, 1999; De Groot, 2004; Blank, Felsö en De Groot, 2011).

### ***Het gras is elders groener en hoe lukt ze dat?***

Benchmarking is dus een vanzelfsprekendheid geworden en geaccepteerd geraakt (Francis & Holloway, 2007: 171) maar waar hebben we het eigenlijk over bij benchmarking? Het wordt tijd om preciezer te worden dan tot nu toe in dit betoog. Benchmarking wordt wel eens frivool omschreven als een antwoord op de vraag *'is het gras bij de burens groener? En zo ja, hoe lukt ze dat?'* Waar het bij benchmarking om gaat, is ook anders te formuleren. Benchmarking is hier opgevat als een vorm van toegepast onderzoek waarbij organisaties, onderdelen of substantiële, samenhangende productstromen kwantitatief worden vergeleken om zo te zien wat de beste, meest lucratieve, effectieve of doelmatige organisatie is, en wat tot de beste praktijken behoort ('best practice'). Een dominante gedachte was en is dat van prestatievergelijking valt te leren en ook van vergelijking tussen achterliggende processen en werkwijzen. De zwakkere organisatie kan zich aan de sterkere praktijken optrekken (De Vriend en Timmerman, 1995; Maas e.a., 1998; Nelissen en De Goede, 1998; Martens e.a., 2002: 38; Van der Zee, 2005).

Benchmarking was dus oorspronkelijk gericht op verbeteren (ook wel *'benchlearning'* genoemd) en niet op louter transparant maken en zich verantwoorden naar de samenleving of naar specifieke groepen of 'stakeholders'. Werkt die benchmarking in de publieke sector? In elk geval is op tal van terreinen sprake geweest van toepassing, zoals bij voorbeeld in de zorg (verpleeghuizen, ziekenhuizen, regionale centra voor geestelijke gezondheidszorg, jeugdgezondheidszorg), bij pensioeninstellingen, bij rechtbanken, bij regionale opleidingscentra, bij universiteiten, bij politiekorpsen, bij water- en zuiveringsschappen en bij diverse gemeentelijke diensten (o.a. Van Helden en Brouwer, 2005; Cleuver en Blank, 2006; Brans e.a., 2008; Abma en Korsten, 2009; Van Tits, 2009). Waarom en waar? Daar waar productiestromen zich lieten en laten kwantificeren, valt vaak ook benchmarking waar te nemen. In benchmarks uit die sectoren komen we dan ook vooral kwantitatieve indicatoren van kwaliteit, effectiviteit of doelmatigheid tegen, zoals een vergelijking tussen ziekenhuizen qua aantal ligdagen van patiënten in ziekenhuisbedden, van aantallen behandelingen door medewerkers van verschillende riagg's, omvang en kwaliteit van afvalzuivering van water, trajectduurverschillen en overeenkomsten bij behandeling van bijstandsaanvragen, de behandelduur van vergunningaanvragen, de stoelbezetting in schouwburgen of productie- en werkdrukanalyses bij personeel.

### **Waar benchmarking gedijt**

Benchmarking gedijt van oudsher in *'blue collar service organizations'*, zoals die van het transport, en daar waar sprake is van laboratoriumachtige processen maar drong later ook door in andere servicegerichte organisaties (Francis & Holloway, 2007: 173). Benchmarking gedijt tegenwoordig vooral in wat Mintzberg (1980, 2006) ooit *machinebureaucratieën* noemde; het gaat daarbij om een bepaalde organisatieconfiguratie, waarin veel eenheden 'verwerkt' worden, zoals vroegere sociale diensten (centra voor werk en inkomen), scholen ('de schoollijst van Jaap Dronkers') en bepaalde onderdelen van politiekorpsen, verzorgings-, verpleeg- of ziekenhuizen (Cleuver en Blank, 2006; Blank, Dumaij e.a., 2011). Delen van de productie laten zich hier relatief makkelijk kwantitatief 'vangen'.

Benchmarking wordt daarentegen (vrijwel) nooit toegepast bij het vergelijken van bepaalde andere organisatietypen, zoals

- geheime diensten (producten worden meestal niet gespecificeerd),
- 'high performance'-organisaties, waar 'zero fault' de leidraad is en geheimhouding ook een rol speelt (bij voorbeeld kerncentrales),
- of bij sterk op creativiteit gerichte professionele organisaties, zoals architectenbureaus (zie Baakman e.a., 2001).

### **Waar wel of geen benchmarking?**

Welke organisaties van eenzelfde soort passen meer of minder vaak benchmarkonderzoek toe? Welke gemeenten doen juist wel of niet mee aan benchmarkonderzoek? De sterke indruk bestaat

dat benchmarking meer voorkomt in *op innovatie gerichte* gemeenten waar het grote belang van 'een lerende overheid' door het bestuur een en andermaal uitgesproken wordt en minder voorkomt in *bestuurlijke probleemgemeenten*. Probleemgemeenten zijn gemeenten waar in een jaar of twaalf vaker burgemeesters en/of wethouders sneuvelen en coalities uit elkaar vallen. Daarachter gaat vaak schuil dat de bestuurlijke oriëntatie op de langere termijn via strategische beleidsvoering afwezig is of niet sterk ontwikkeld is. Het ad hoc-denken is juist in deze probleemgemeenten sterk ontwikkeld. Er wordt minder een algemene koers of beleidslijn gevolgd of makkelijk daarvan afgeweken. In die lijn past dat veel beleidskwesties als individuele gevallen worden behandeld. Daardoor neemt de kans op cliëntelisme, wispelturigheid en conflicten in de besluitvorming toe. Vaak is ook sprake van een rauwe politieke cultuur. De conflicttemperende krachten zijn daarentegen betrekkelijk zwak.

Een voorbeeld van een gemeente die niet of nauwelijks meedeed met visitaties, bestuurskrachtmetingen en evenmin aan benchmarking is de gemeente Maasdriel. In de literatuur vinden we een lange lijst van andere gemeenten (Schoenmaker, 2011). Het is wel voorgekomen dat probleemgemeenten als Den Helder wel meededen met een benchmark, zoals '*De staat van de gemeente*'; dan scoren ze op tal van indicatoren voor kwaliteit van het gemeentebestuur relatief lager dan tal van andere gemeenten (Aardema en Korsten, 2005; Abma en Korsten, 2009).

Bestuurders van probleemgemeenten kunnen niet goed voldoen aan een eis uit de '*Code voor goed openbaar bestuur*', namelijk beschikken over zelfreinigend vermogen. Ze krijgen dit vermogen door middel van evaluaties en benchmarking niet goed geactiveerd (Korsten en Aardema, 2006; Korsten, 2010; Korsten en Schoenmaker, 2011; Schoenmaker, 2011).

### **Onderwerp**

Dat was een eerste kennismaking. Deze beschouwing handelt verder over het onderzoeksinstrument benchmarking en dan in het bijzonder over de mythen. Mythen worden hier opgevat als '*overgeleverde verhalen*', dat wil zeggen aan benchmarking toegeschreven kwaliteiten en hardnekkige beelden waarvan de vraag is of ze bij nader inzien juist blijken. Een dergelijke mythe is dat benchmarking daadwerkelijk bijdraagt aan bedrijfsmatigheid in het functioneren van organisaties. In dit essay wordt getracht over die mythen iets te zeggen aan de hand van onze ervaringen met benchmarking in het openbaar bestuur, in het bijzonder over gemeenten. Daarbij richten we ons met name op risico's bij benchmarking in de vorm van valkuilen. Risico's worden hier opgevat als onverwachte of ongewenste omstandigheden, gebeurtenissen of dreigende effecten bij de toepassing van benchmarking die maken dat benchmarking niet beantwoordt aan de doelstellingen of achterblijft bij de verwachtingen. Valkuilen in onderzoek of daarbuiten worden tot de risico's gerekend.

### **Mythen**

Welke mythen onderscheiden we hier? We noemen er vier maar er zullen er meer te formuleren zijn. Een eerste mythe luidt: *organisaties zoeken zelf naar partnerorganisaties waarmee ze het eigen presteren willen vergelijken en waarvan ze verwachten dat die ook qua prestaties vergelijkbaar zijn.*

Een tweede mythe luidt: *Benchmarking leidt tot het opsporen van 'best practices' (goede, navolgbare praktijken) en bezinning daarop in benchmarkrapportages.*

Een derde mythe: *Benchmarking draagt bij aan overheidsorganisaties die beleid, in casu prestaties, producten, processen, daadwerkelijk verbeteren en zich daarmee een lerende organisatie tonen.* Uitkomsten van benchmarks moeten worden benut en dat gebeurt ook door veranderprocessen, gericht op verbetering.

En een vierde mythe: *Verbeteren is in elk geval financieel-economisch gericht*. Benchmarking van vergelijkbare organisaties op onder meer doelmatigheid draagt bij aan een meer bedrijfsmatig functionerende overheidsorganisaties. Lees: benutting van benchmarkrapporten leidt in de praktijk tot een meer doelmatige uitvoering van taken, dus met minder middelen en meer voortgang of snelheid (minder personeel, minder budget, betere afstemming van werkzaamheden wat leidt tot meer snelheid, enz.).

Op het einde van deze beschouwing geven we het antwoord maar we gaan dit antwoord eerst opbouwen.

### **3 Kenmerken van benchmarking**

Leidinggevenden van overheidsorganisaties, zoals besturen en topambtenaren, bedienen zich van allerlei bestuursmiddelen. Een van die middelen is het opstellen van een begroting en het werken met kengetallen daarin, een ander is benchmarking. Benchmarking past in een denken over de bedrijfsmatige gemeente of breder 'de lerende organisatie', dus van een organisatie die door het organiseren van kritische feedback op eigen handelen in de vorm van vergelijking en leren van andere organisaties zichzelf verder helpt. Anders gesteld, benchmarking past in de *Code voor goed openbaar bestuur* zoals die door een van de laatste kabinetten-Balkenende is vastgesteld. Leren en zelfreiniging passen daarin (Addink, 2010; Korsten, 2011; Abma, 2012: 353 e.v.).

#### ***Geen beleidsevaluatie of monitoring***

Benchmarking is een vorm van vergelijkend cijfermatig beoordelen van wat organisaties deden maar het is *geen* beleidsevaluatie *vanuit* overheidsdoelstellingen. Beleidsevaluatie impliceert doorgaans het beoordelen van een beleidsprogramma, zeg voor verkeersveiligheidsbeleid, in termen van officiële beleidsdoelstellingen (gewenste prestaties en effecten) en het nagaan of het wel of niet bereiken van wat men wil, veroorzaakt is door, of anders gesteld toe te schrijven is aan het beleid zelf of aan omstandigheden daarbuiten.

Beleidsevaluatie kan op het verleden gericht zijn of meer prospectief. Benchmarking impliceert daarentegen *zelden of nooit* het beoordelen van een bestuur op beleidsdoelstellingen. In een benchmarkrapport valt nooit te lezen dat welke doelstellingen partnerorganisaties erop nahielden of nahouden en evenmin of de beleidsdoelstellingen wel of niet bereikt zijn. Benchmarking is altijd *ex post*. Benchmarking kan zeker toekomstgericht zijn maar is dat zelden.

Overigens wordt in begrotingen vanuit outputsturing wel nagestreefd om te laten zien of een gewenste – smart geformuleerde - prestatie bereikt is. De begroting van de gemeente Amsterdam van de afgelopen jaren toont op een groot aantal terreinen gekwantificeerde doelstellingen. Maar dat als evaluatie betitelen, gaat ons te ver.

Benchmarking is ook *geen* vorm van interbestuurlijke monitoring De Kool (2007) want dat impliceert doorgaans dat in een bestuurlijk bestel van overheidslagen een hogere overheid informatie vraagt en ontvangt over de beleidsinput van een andere overheid bij voorbeeld een gemeente, of over de throughput of de output. Benchmarking vervult dat niet of nauwelijks.

De rijksoverheid is overigens uit een oogpunt van verantwoording naar het parlement vaak wel geïnteresseerd in wat een gemeentebestuur aan beleid voert omdat het gemeentebestuur het rijksbeleid geheel of grotendeels (in medebewind) uitvoert. Daarvoor worden (interbestuurlijk gerichte) monitoring gearrangeerd (verder De Kool, 2007). Denk ook aan het Sociaal en Cultureel Planbureau dat de uitvoering van de Wet maatschappelijke ondersteuning (Wmo) al jaren volgt (SCP, 2010).

Omdat benchmarking geen vorm van beleidsevaluatie vanuit doelstellingen is noch van interbestuurlijk gerichte monitoring, kan men de vraag stellen: zouden lokale rekenkamer(commis)sie(s) die weliswaar onafhankelijk zijn maar toch nogal eens een verzoek krijgen om een onderzoek naar een politiek relevant 'dossier' of thema te verrichten gebruik maken van



benchmarking? Dat is sinds de instelling van lokale rekenkamer(commissie)s na 2002 over het algemeen niet het geval. Van de meer dan duizend onderzoeken zijn er, uitgaande van de titels en korte beschrijvingen van de inhoud, vrijwel geen te typeren als benchmark-onderzoek. Wie ook benchmarkings verricht, lokale rekenkamer(commissie)s niet.

We kennen ook nog een andere vorm van monitoring. In het boekje 'Nederland aan de monitor' (Engbersen e.a., 1997: 12) is sprake van monitoring als:

1. De informatieverzameling periodiek en systematisch uitgevoerd wordt;
2. De verzamelde informatie actueel is;
3. De gebruikte indicatoren zijn steeds per item hetzelfde;
4. De verzamelde informatie is beleidsrelevant; en
5. De monitor kent een ijkpunt.

Monitoring is dan iets wat een gemeente periodiek uitvoert. Bij benchmarking heet dit de variant van 'benchmarking in de tijd'. Gemeentebesturen kunnen dan door de jaren heen de ontwikkelingen volgen. Bij 'De Staat van de Gemeente' deden gemeenten die vaker meededen dat ook door de uitkomsten van de meest actuele meting te vergelijken met die uit voorgaande jaren (Abma, 2012: 249-260; Abma en Korsten, 2009: 48-56).

### **Vergelijken**

Benchmarken impliceert, we herhalen het nog eens, a) het vergelijken van organisaties of delen of aspecten daarvan in kwantitatief opzicht door middel van scores, b) het (rang)ordenen van uitkomsten op vergelijkingspunten en het begrijpen van verschillen, c) het zoeken naar 'best practices', en d) het trachten te verbeteren van de organisatie in de richting van de betere organisaties of betere praktijk (cf Van Tits, 2009: 329). Denk daarbij aan het vergelijken van vijftien gemeenten op aspecten van gemeentelijke dienstverlening, zoals bij voorbeeld de tijdigheid van een ambtelijke of bestuurlijke reactie op brieven van burgers. Dit is dus de *brede* opvatting van benchmarking. Nelissen en De Goede (1999) maakten onderscheid tussen vergelijken, wat zij 'performance measurement' noemden en verbeteren ('continuous improvement'). Pas wanneer vergelijken en verbeteren methodisch worden gekoppeld is er volgens hun sprake van 'echte' benchmarking. Dit in tegenstelling tot de *smalle* opvatting waarbij het alleen gaat over vergelijken en wat dus eigenlijk het instrumentele deel betreft.

Vergelijken zonder dat bekend is wat een goede, voldoende of (on)aanvaardbare score is, is lastig. Vaak hebben gemeenten evenwel een *kwaliteitshandvest* opgesteld en zijn daarin normen opgenomen voor kwaliteit (bij wijze van voorbeeld: 'elke beller moet binnen tien minuten goed doorverbonden kunnen worden'). Het nadeel is dat kwaliteitshandvesten een beperkt bereik hebben. Ze gaan slechts over dienstverlening terwijl het geheel van gemeentelijke taken en ambities veel en veel breder is. Denk maar eens aan het bezit en onderhoud van sportvelden, het bieden van voorzieningen voor openbaar onderwijs, maatschappelijke ondersteuning, reïntegratie naar een baan, opvang van daklozen. Over het omgaan met daklozen gaat een kwaliteitshandvest nooit (o.a. Hiemstra, 2003).

Gezien het voorgaande zal het niet verrassen: een benchmarking-aanpak moet speciaal *ontworpen* worden (Hoogwout, 2004; Bouckaert e.a., 2011). Zijn daarin ook normen of andere streefcijfers nodig?

### **Normen**

Vergelijken met of vanuit normen of standaarden? Kathe Callahan (2007: 64) stelt in haar boek '*Elements of Effective Governance*':

'The usefulness of a performance measurement system is enhanced when it includes benchmarking. In this process, performance indicators are compared to other performance indicators or to established standards of performance. Baseline data provide a standard or a

norm against which the performance of a program or service can be compared. Programs can be benchmarked over time’.

Haar boodschap is dat in een vergelijkende analyse, een benchmark, bij voorkeur prestaties met normen of standaarden vergeleken moeten en kunnen worden.

Vaak is in de praktijk van gemeentelijke en andere benchmarking overigens *geen* door een beroepsgroep opgestelde of wettelijk verankerd streefcijfer, richtlijn of andere norm voor een gewenste of vereiste minimale kwaliteit voorhanden noch een normering door gemeentebesturen zelf. Bij benchmarkrapporten in de medische of milieusfeer kan dit anders liggen want op dit vlak bestaan wel normen voor dreiging, veiligheid of (gezondheids)risico's.

Het gaat dus om vergelijken, om het zoeken naar de plaats van de organisatie te midden van een veld van vergelijkbare organisaties en het opsporen van de 'best practices' voor verbetering, en daarmee om wat zich aandient als een navolgbare goede praktijk. Onderzochte organisaties spiegelen zich dus aan elkaar of worden gespiegeld aan een streefcijfer.

#### **Soorten performance-indicatoren**

In de literatuur over performance management en benchmarking zijn diverse begrippen in omloop. We noemen er enkele.

Input: het geheel aan middelen om resultaten te boeken, zoals budget, personeel, gebouwen, grond.

Throughput: het geheel aan activiteiten om met behulp van input te komen tot prestaties.

Output: het geheel aan prestaties of resultaten die voortvloeien uit activiteiten van een organisatie.

Outcomes: gewenste en ongewenste effecten van prestaties van een overheidsorganisatie, of van andere invloeden.

Productiviteit: output gedeeld door een specifieke input. Bijv. het percentage opgeloste misdrijven per rechercheur.

Efficiency: de verhouding van output ten opzichte (een index van) van alle inputs, zoals alle gemaakte kosten.

Effectiviteit: outcome of gewenst effect gedeeld door output. Bijv. De verhouding tussen aantallen verkeerslachtoffers (gewenst effect) ten opzichte van de plaatsing van verkeersdrempels (prestatie) in een stad.

(Bron: Bouckaert en Van Dooren, 2009: 155; Haselbekke, 1998).

Over het opstellen van streefcijfers en het onderscheid naar typen streefcijfer en de problemen die zich daarbij voordoen, is in de literatuur veel te doen (Haselbekke, 1998; Bordewijk en Klaassen, 2000; Behn, 2003; De Groot, 2004; De Vries en Lako, 2009; Blank, 2010; Bouckaert e.a., 2011; Bongers en Verlet, 2013).

Ten eerste blijkt het niet zo eenvoudig om al het overheidsbeleid in duidelijke producten te specificeren en te komen tot streefcijfers. Onderwijs is vanuit gemeentelijk perspectief niet in hapklare brokken 'output' of prestaties uit te drukken. Hetzelfde geldt voor leefbaarheid of welzijn, stellen wij in de voetsporen van Bordewijk en Klaassen (2000).

Ten tweede kunnen perverse effecten optreden (De Bruijn, 2001, 2003, 2007). Als door een gemeentebestuur aan een schouwburgsubsidie een forse stoelbezetting per gemiddelde uitvoering gekoppeld wordt, kan de schouwburgdirectie ertoe overgaan slechts publiekstrekking te programmeren en/of evidente succesnummers in een a- en b-gedeelte te splitsen en zo twee uitvoeringen te gaan tellen.

Ten derde, een analyse van toepassingen van kengetallen in Tilburgse begrotingen wees uit dat sprake was van een rommelig geheel. Gemeente-ambtenaren bleken niet in staat om steeds kengetallen te produceren en vergelijken. Lang niet iedereen wist aan welke eis een kengetal moest voldoen (Haselbekke, 1998).

#### **Scoringseisen**

Benchmarking meet prestatiekwaliteit en streeft die na, maar moet op zich ook 'high performance' bieden (cf Van Dooren, Bouckaert en Halligan, 2010). Daarom moet benchmarking zelf aan eisen voldoen. Welke? We kunnen denken aan eisen te stellen aan de opdracht, het meetsysteem, de rapportage en de manier van werken in het kader van de bezinning op uitkomsten. We belichten hier de prestatie-maten.

Een goede set van prestatie-maten moet aan eisen voldoen. Van Dooren, Bouckaert & Halligan (2010: 71) wijzen op drie kwaliteitssets:

- functionaliteit van het meetsysteem,
- geldigheid van de meting zelf en betrouwbaarheid,
- en legitimiteit van het meetsysteem.

Callahan (2007: 68-69) accentueert een cluster van de drie: de kwaliteit van de meting. Maar zij heeft ook oog voor relevantie en dat raakt aan functionaliteit en voor de communicatie rond benchmarkonderzoek en daarmee raakt zij aan legitimiteit. Zij formuleert de volgende kwaliteitseisen: omvattendheid (comprehensive), begrijpelijkheid (understandable), validiteit van de maten (meten ze wat ze moeten meten?), actualiteit van de cijfers, betrouwbaarheid (reliability; blijkt bij herhaling hetzelfde?), resistentie tegen perverse effecten, afwezigheid van redundantie (niet kiezen voor scores die zelfde prestaties meerdere malen meten), lage verzamel-prijs van data. Het moet bij meting gaan om relevante en controleerbare aspecten van prestaties. Elk prestatie-metingssysteem of meetaanpak moet volgens haar in essentie accuraat zijn en waardevolle informatie genereren. Om dat te bereiken zal het aan de dag leggen van flexibiliteit nodig zijn.

De meest succesvolle pogingen tot benchmarking zijn die waarbij de unieke kenmerken van meting en meetobject erkend worden, dus van wat in onderzoek is, en rekening gehouden wordt met waaraan opdrachtgevende organisaties behoefte hebben, aldus Callahan. Daarmee geeft zij aan dat de entreebesprekingen tussen opdrachtgever en onderzoeker voorafgaand aan de daadwerkelijke uitvoering van onderzoek van belang zijn.

Komen de eisen die aan indicatoren gesteld worden ook bij andere auteurs terug? Bouckaert e.a. (2011: 122) besteden hier nogal aandacht aan. Ze noemen: validiteit, accuraatheid, begrijpelijkheid, betrouwbaarheid, specificiteit en voegen draagvlak toe.

Alle criteria zijn van belang. Maar niet vergeten moet worden dat de benchmarkopzet verantwoord moet worden en dat de toepassing tot een leesbaar en dus begrijpelijk rapport moet leiden. Dat laatste is lang niet altijd het geval, zoals nog zal blijken.

### ***Het kwantitatieve is dominant***

Doorgaans zal een benchmark uitgevoerd worden door een opdrachtnemer in opdracht van een opdrachtgever; meestal is de uitvoerder een onderzoeks- of adviesbureau. Daarbij is vrijwel altijd sprake van een kwantitatieve, cijfermatige vergelijking. Een benchmarkrapport staat dus bol van de scores en soms ook van zgn. spinnenwebben. Het is dan ook te begrijpen dat aan de oorsprong van benchmarking instituten stonden op het gebied van statistiek.

Benchmarking tracht iets over de kwaliteit van kerntaken van organisaties (b.v. een scholen, ziekenhuizen) of organisatiedelen (dienstverlening van gemeenten vergeleken; vergelijking van centra voor werk en inkomen; of van rioleringszorg) te zeggen door middel van kwantitatieve uitspraken: scores. Soms kan ook sprake zijn van vergelijking vanuit een focus; bij voorbeeld een belichting van gemeentelijke dienstverlening gezien vanuit het perspectief van klantgerichtheid. Maar bijna altijd is sprake van een zekere omvattendheid in de beoordeling: van een organisatie als geheel of een substantieel deel, of een substantieel geheel van producten of processen.

Als er al van kwalitatieve uitspraken sprake is dan zal het gaan om het begrijpen van datgene wat kwantitatief gevat is.

Er komen uitzonderingen voor op het dominant kwantitatieve. Zo waren eerste visitaties van (universitaire) wo-opleidingen in de vorige eeuw vooral kwalitatief, aangevuld met elementen van kwantitatieve vergelijking. Een zekere benchmarking trad op omdat de visitatiecommissies ‘best practices’ opspoorde en vermeldde in een overal rapport van de beoordeelde zelf opleidingen (bijvoorbeeld alle wo-opleidingen bestuurskunde). In bestuurskrachtonderzoeken werd sporadisch gebruik gemaakt van benchmarks (Abma, 2012: 54) maar die speelden in de uiteindelijke beoordeling een marginale rol van betekenis,

Benchmarkingrapporten bevatten over het algemeen overigens geen concrete gespecificeerde aanbevelingen voor elke gemeente of elk waterschap dat deelneemt aan een benchmarking naar aanleiding van de uitkomsten.

### **Doelen van benchmarking**

Benchmarking is bijzonder. Benchmarkonderzoek appelleert aan de nieuwsgierigheid. Velen willen weten: waar staat mijn organisatie te midden van andere? En scores die goed uitkomen, kunnen een motivatie-effect hebben op organisatieleden. Ze kunnen het elan bevorderen. Scores en rangordeningen kunnen ook gebruikt worden als marketinginstrument, bij voorbeeld voor universiteiten om studenten te werven of om de aantrekkelijkheid voor toetreding van medewerkers te vergroten (Abma en Korsten, 2009: 142). Toch zijn dat geen voldoende primaire redenen om aan benchmarking te doen. Doel van benchmarking is volgens degenen die benchmarking bepleiten ‘spiegelen en verbeteren’ (Van Tits, 2009: 239; Van Dooren, Bouckaert en Halligan, 2010)). Benchmarking staat ook in functie van transparantie en verantwoording naar controlerende organen en de samenleving. Veel benchmarking in de publieke sector is daarom openbaar.

### **Vormen van benchmarking**

Vermeldenswaard is dat benchmarking geen eenheidsworst is. Er bestaan verschillende vormen. In de literatuur wordt daar melding van gemaakt en wordt gesteld dat het waarnemen van deze variëteit van belang (Francis & Holloway, 2007: 174-176). Elke benchmarkstudie is te typeren op basis van een aantal onderscheidingen (tabel 1). Aan de literatuur ontleen we de volgende onderscheidingen: een onderscheid van benchmarkvormen is dat in een benchmark vanuit een leerstrategie van de organisatie of een benchmark waarbij die strategie ontbreekt; een partiële of integrale benchmark; een externe en interne benchmarking; benchmarking van resultaten of processen of beide; benchmarking met vermelding van ‘best practices’ of zonder vermelding van ‘best practices’; benchmarking die wel of niet vergezeld gaat van een verschillenanalyse en dus een verklaring van verschil tussen scores; benchmarking met één overal rapport voor alle deelnemende partnerorganisaties of een overal rapport plus een rapport voor elke deelnemende organisatie; een rapport dat wel of niet openbaar is.

Tabel 1: Vormen van benchmarking (bm)

1 Bm is gevolg van een leerstrategie van organisaties		Bm is geen gevolg van een leerstrategie van organisaties
2 Vrijwillige deelname aan bm		Verplichte deelname aan bm
3 Leiding van bm door visitatiecommissie		Leiding door onderzoeksbureau
4 Onafhankelijke bm t.o.v. deelnemende partnerorganisaties		Coöperatieve deelname van partnerorganisaties
5 Externe vergelijking tussen organisaties		Interne vergelijking binnen een organisatie
6 Partiële bm /aspectmatig		Integrale bm
7 Resultaat- en procesanalyse		Resultaat- of procesanalyse
8 Cijfers over meerdere jaren		Cijfers over een jaar
9 Vermelding van ‘best practices’ in rapport		Geen vermelding van ‘best practices’
10 Verklaring van verschillen tussen scores		Geen verklaring van verschillen
11 Overalrapport en rapport per partnerorganisatie		Alleen een overall-rapport

Er is natuurlijk sprake van een favoriet benchmarking-onderzoek, namelijk: passend binnen een leerstrategie; onafhankelijk en geleid door een commissie; met een vergelijking tussen externe partners; resultaatgericht; met aanduiding van best practices; met een verschillenanalyse; met een overallrapport en een rapport voor elke afzonderlijke deelnemer; met openbaarheid van rapportage.

Een integrale benchmarking van organisaties, zoals een ziekenhuis, is minstens zo uitdagend als een partiële vergelijking omdat deze complexer is (Abma en Korsten, 2009).

#### **4 Ervaringen met verschillende benchmarkingtrajecten: verdiepende impressies**

Benchmarking is een veel toegepast instrument.

‘Since the 1980s, in the US, the use of the benchmarking in the public services has grown spectacularly’ ....’(Triantafillou, 2007: 829).

Daarbij wordt vaak gezegd dat de groei vooral zit in de toepassing in bepaalde sectoren. Maar hoe dan ook: ‘Public sector benchmarking is of increasing importance’ (Van Helden & Tillema, 2005: 337). Heeft benchmarking dan alleen voordelen?

##### ***Veronderstelde voordelen***

Diverse voordelen van benchmarking zijn genoemd, zoals de aanzet tot verbetering in wat een organisatie doet en voortbrengt (ook Van Dooren, Bouckaert en Halligan, 2010).

Een van de toe te voegen voordelen is dat over het algemeen geen enkele politieke partij zich er tegen verzet. Benchmarking past in het denken van politiek links tot politiek rechts, zo lezen we in de internationale literatuur (Triantafillou, 2007:830).

Andere aan benchmarking toegeschreven voordelen zijn: deze methode verschaft nieuwe inzichten; de methode verschaft vroeger waarschuwingen over waar een organisatie voorop loopt of achterblijft; de methode werkt motiverend voor betrokken overheidspersoneel (Francis & Holloway, 2007: 181).

Benchmarking lijkt daarom een hoerabegrip; je kunt moeilijk op benchmarking tegen zijn. Benchmarking is misschien mede daarom ook wel ‘vermaatschappelijk’. Het verlangen naar rangordes blijkt groot. Benchmarking prikkelt immers de nieuwsgierigheid bij gebrek aan een marktaandeel-indicator zoals die voor private bedrijven geldt. Gepredikt wordt dat burgers in de relevante rol (als woningzoeker, patiënt, cliënt, ouder, etc.) zelf aan de hand van vergelijkende overzichten meer en beter zouden kunnen en moeten kiezen. De redactie van het weekblad Elsevier publiceert met zekere regelmaat rangordes van de beste woongemeente, de veiligste gemeente, de innovatiefste regio, de beste ziekenhuizen, de beste middelbare scholen en universitaire opleidingen.

Maar alleen de presentatie van hitlijsten, geordend naar prestatieomvang of kwaliteit of een ander aspect, volstaat niet om te kunnen spreken van benchmarking (cf Blank, Felsö en De Groot, 2011: 5). Er is altijd inzicht nodig in *het waarom*, dus in wat verschillen tussen organisaties verklaart, en waarom bepaalde gemeenten bovenaan staan en andere onderaan. Want alleen als je daarover veel achterhaalt, zijn in principe verbeteracties mogelijk die zijn te betitelen als ‘*benchlearning*’ (cf Tebn Tije e.a., 2010; Blank e.a., 2011) en het daadwerkelijk verbeteren als ‘*benchaction*’ (Ten Tije e.a., 2010).

##### ***Veronderstelde nadelen***

Er bleken ook wat andere problematische kanten. Een zo'n kritiek is dat benchmarking niet ambities oproept maar verdrijft (De Bruijn, 2001, 2003, 2007; Cie-Bovens, 2006). We noemen nog enkele andere nadelen:

- benchmarking leidt tot eindeloze rangordes van de rangorde-industrie; voor elk wat wils?;
- het reductie-argument: alleen het kwantitatieve komt in beeld en het kwalitatieve niet?;
- het outcome-argument: prestatie management moet niet gericht zijn op prestaties maar op gewenste effecten;
- het normalisatie-argument: benchmarking leidt tot uitbanning van innovatie, variëteit en pluriformiteit en drang naar het gemiddelde;
- het monocentrisme-argument: het netwerk van partnerorganisaties en samenleving komt bij benchmarkonderzoek niet in beeld terwijl de rol van de ene overheidsorganisatie overschat wordt?;
- het schuttersputje-argument;
- het argument van schijnsucces door strategisch gedrag rond normen;
- de groei van de bureaucratie door administratieve lasten rond deelname aan benchmarking;
- permanente veranderingen door juist veel benutting van uitkomsten;
- schadelijkheid van prestatie meting voor professionaliteit?

We lichten deze nadelen toe.

### **1 Het argument van de eindeloze rangordes**

De gemeente Tilburg is de best presterende gemeente in de Benchmark Wet Werk en Inkomen 2012, een jaarlijkse ranglijst van SGBO en Stimulansz. Eerder won Rijswijk, en daarvoor Tilburg en Hengelo. Aan het onderzoek deden 49 gemeenten mee. Als de samenwerkingsverbanden worden meegeteld, stijgt het aantal naar 67. De ranglijst was gebaseerd op tien, elk even zwaar wegende indicatoren, waaronder het gemiddelde bedrag van de verstrekte uitkering; de incassoquote; de doorlooptijd van aanvragen WWB levensonderhoud; het aandeel gegronde bezwaarschriften; de kosten per klant; het bijstandsvolume gerelateerd aan december 2011. Tilburg had korte doorlooptijden, weinig gegronde bezwaarschriften en lage kosten per klant. Amsterdam bleek de grote stijger. Het gevolg was dat de Amsterdamse wethouder Andrée van Es ook direct een persbericht liet uitgaan, waarbij ze de medewerkers van de dienst feliciteerde. Ook de Tilburgse wethouder reageerde enthousiast en benadrukte wat er allemaal goed was. Bij deze rangorde bleef het in 2013 niet. Er ontstond het afgelopen decennium nationaal en internationaal een ware *rangorde-industrie* (Hood, Dixon & Beeston, 2008). De VNG telde rond 2006 al ruim dertig benchmarks. En er kwamen lijstjes van bij voorbeeld de beste woongemeente, de beste universiteit of school, het meest doelmatige afvalbeheer en de veiligste gemeente. Het hangt soms maar van de indicatoren af wie als hoogste (of beste, doelmatigste of effectiefste) uit een rangordening komt. Daarmee wordt direct al een complexiteit genoemd: welke set criteria is relevant, geschikt én voldoende om iets te zeggen over de beste universiteit of het meest deugdelijk gemeentebestuur? Wie lang genoeg zoekt, zal zien dat een organisatie altijd wel ergens de beste in is, waardoor andere, lagere scores wat gerelativeerd worden (verg. Hood e.a., 2008; Abma en Korsten, 2009). Benchmarking kan *pervers gedrag* oproepen. Wie niet de beste in jaar x, kan de prestatie meting voortzetten door over de jaren u,v,w en x te meten. De 'rangorde-industrie' heeft meerdere manieren voortgebracht om slechte scores onder het tapijt te vegen. Dergelijk gedrag kan ontstaan als kerntakendiscussies opkomen of een bezuiniging dreigt.

### **2 Het reductie-argument**

Benchmarking werkt verhullend in zoverre de werkelijkheid teruggebracht wordt tot het meetbare, tot het kwantitatief meetbare in op toepassing gericht onderzoek. Niet alles is te beperken tot het kwantitatieve. De statistisch geaggregeerde werkelijkheid, die gebaseerd is op prestatie-indicatoren, doet geen recht aan de multidimensionaliteit en kwalitatieve kant van bepaalde maatschappelijke vraagstukken. Sommige waarden laten zich bovendien niet of niet makkelijk kwantificeren, zoals

mededogen. Denk aan de beoordeling van prestaties van de sociale werkvoorziening, die werk biedt aan personen met een lichamelijke of geestelijke beperking. Een dergelijke organisatie kan niet optimaal concurreren met andere organisaties (Bekkers, 2012: 106).

### **3 Het outcome-argument**

Bij overheidsbestuur gaat het veel niet alleen om prestaties (outputs) maar juist om gewenste effecten van beleid (outcomes). Benchmarking meet vaak wel prestaties maar juist geen gewenste effecten en leidt niet tot vergelijking daarvan (Van Dooren, Bouckaert & Halligan, 2010: 176). Het gaat niet om het aantal op straat zichtbare politiemensen maar om het voorkomen van misdrijven, het registreren van misdrijven en het oplossen daarvan en om het bevorderen van een daarmee sporend veiligheidsgevoel bij burgers. 'Performance management should primarily focus on outcomes', aldus Van Dooren c.s. (2010: 176).

### **4 Het normalisatie-argument**

De vergelijking tussen biersoorten kan uitwijzen dat alle bieren van bierproducent Hanekos de beste zijn, waarna andere bierproducenten allemaal gaan werken richting de standaard die Hanekos heeft gezet als 'best practice'-organisatie. Het gevolg kan zijn dat producenten gedrag gaan vertonen dat ertoe leidt dat alle bieren op elkaar gaan lijken ('isomorfisme'; Francis & Holloway, 2007: 183). De commissie-Bovens (2006: 28) spreekt in dit verband van nivellering als gevolg van benchmarkonderzoek en meer van een oriëntatie op het gemiddelde. Normalisatie of nivellering, variëteit in producten wordt hoe dan ook niet vergroot, eerder teruggedrongen; diversiteit neemt af. Dat kan ook gebeuren bij overheidsorganisaties. Dit proces is normalisatie te noemen (ook Triantafillou, 2007). De consequentie kan zijn dat innovatiepogingen en durf om af te wijken worden teruggedrongen (Cie-Bovens, 2006: 28).

### **5 Het monocentrisme-argument**

Benchmarking brengt over het algemeen niet netwerkproductiviteit in beeld terwijl gemeenten voor tal van taken, opgaven en middelen afhankelijk zijn van mede-overheden en andere actoren. Daarmee moeten gemeenten samen een strategie volgen. Ze moeten komen tot gemeenschappelijke beeldvorming en hun doelen en middelen verknopen (Cörvers, Glasbergen en Korsten, 2003; Kaashoek e.a., 2010; De Rynck, Voets en Van Dooren, 2010; Van Dooren, Bouckaert en Halligan, 2012). Denk maar aan gemeenten die woningbouw plannen, en corporaties en projectontwikkelaars die bouwen. Deze ontkenning van 'governance' is als een nadeel van benchmarking te beschouwen. Benchmarkonderzoekers wekken ten onrechte de indruk dat een te benchmarken groep organisaties zelfstandig aan alle knoppen zit.

### **6 Het argument van schijnsucces**

Outputsturing, die benchmarking als hulpmiddel gebruikt, heeft als potentieel nadeel dat deze *vernieuwing in praktijken* kan *belemmeren* en kan bijdragen aan *verstarring* (Commissie-Bovens, 2006: 28). Dat kan doordat organisaties doelen en prestaties gaan selecteren die makkelijk (maximaal) te realiseren zijn. Dan ontstaat *schijnsucces*. Door bovendien eenvoudige maatstaven in benchmarks te selecteren, kan er een discrepantie ontstaan tussen de papieren werkelijkheid van uit benchmarking blijkende goede prestaties en de prestaties die in werkelijkheid tegenvallen. Daarmee betreden we de wereld van de prestatieparadox, perverse effecten en het bedrog (Wittebrood, 2002; Van Thiel en De Leeuw, 2003; Triantafillou, 2007; Bekkers, 2012: 105). Deze aanpak kan verrassen zodra maatschappelijk tumult ontstaat, iets wat bij voorbeeld dreigt bij het gebruik van veiligheidsbenchmarks. Dan kan een benchmark uitwijzen dat in de gemeente x het percentage opgeloste inbraken iets is gestegen maar het veiligheidsgevoel juist is gedaald.

### **7 Het bureaucratie-argument**

Benchmarking is niet gratis; denk aan het opstellen van een offerte en de selectie van een onderzoeksteam, bezinning op de opdracht, indicatorensselectie, gegevensverzameling, analyse,

rapportage. Indien gegevens over meerdere jaren in de analyse betrokken worden, moet er ook sprake zijn van het bijhouden van bestanden. Er kan sprake zijn van aanzienlijke administratieve lasten (De Bruijn, 2001, 2003; Frey & Osterloh, 2009; Noordegraaf en Sterrenburg, 2009; Bekkers, 2012: 106). Dat wordt ook gezien als een nadeel. Een gemeentelijke informant zegt het zo: 'Het aanleveren van gegevens kost veel tijd. Wij zijn een kleine gemeente en hebben geen tijd om aan verschillende onderzoeken mee te doen' (Ten Tije e.a., 2010: 40).

### **8 Permanent veranderen?**

Een aanverwant kritisch argument houdt in dat benchmarking een organisatie in een nagenoeg permanente staat van verandering kan brengen (Triantafillou, 2007). Het produceren moet immers al maar beter. Goed is niet goed genoeg en zwak moet beter als gevolg van benchmarking. Hier is tegenin te brengen dat het aantal benchmarks in de publieke sector niet moet worden overschat. En het nog maar de vraag is of sprake is van 'research utilization' en 'double loop learning', waarbij bestaand beleid volledig ter discussie komt en sprake is van het ter discussie stellen van doelstellingen en waarden (Argyris, 1977; Francis & Holloway, 2007: 180). Er bestaan echter reserves over juist dat lerend effect. Immers, zo zegt de commissie Toekomst Lokaal Bestuur: benchmarks werken nivellerend en drijven ambitie uit (Commissie-Bovens, 2006: 28). Dat duidt niet direct op double loop learning.

### **9 Het schuttersputje-argument**

Als benchmarking op grote schaal plaatsvindt, werkt dat *verkokering en fragmentatie* in de hand. De redenering is als volgt. Als sterk op benchmarking geleund wordt, zullen vooral die prestatiegebieden in beeld komen die overzichtelijk zijn en gerelateerd zijn aan taakgebieden, afdelingen en bepaalde leidinggevenden. Maar er bestaan ook *transversaal beleid* dat aan te duiden is als *programmamanagement*: het gaat daarbij om beleid dat dwars door bestaande afdelingsgrenzen heen loopt, zoals nogal eens met beleid gericht op de bevordering van duurzaamheid het geval is (Korsten, De Jong en Breed, 2010). Dit transversaal beleid raakt aan onder meer wonen, milieu en energie. Het overstijgt de grenzen van afzonderlijke taakgebieden. Dergelijk beleid kent echter minder hechte verankering bij 'eigenaren'. Het gevolg is dat organisaties minder bereid zijn om zich in te spannen voor benchmarking op dat terrein. Benchmarking kan bevorderen dat iedereen zich terugtrekt in het eigen 'schuttersputje', meent Victor Bekkers (2012: 106).

### **10 Benadeling van professionaliteit**

Er is nog een andere kritiek op benchmarking. Die houdt in dat prestatiemeting schadelijk is voor de professionaliteit van professionals. Het zijn professionals in wie een overheidsbestuur *vertrouwen* moet hebben. Onderzoek naar prestaties is zo gezien niet nodig; bestuurders kunnen het presteren aan de medewerkers overlaten. Maar toch vindt prestatiemeting plaats, juist om de prestatie te vergroten. Daardoor komt vertrouwen op de tocht te staan (Van Dooren, Bouckaert & Halligan, 2010: 176). Het gevaar bestaat bovendien, misschien wel daardoor, dat professionals niet meer doen wat primair door hun eigen vakgebied is ingegeven en wenselijk is maar 'dingen doen' waarop te scoren valt. De Bruijn (2003) geeft het voorbeeld van de muskusrattenvanger die niet de muskusrat direct vangt maar afwacht tot het muskusrattepaar een nest jongen heeft om daarna pas te gaan vangen. Tel uit je winst? Hij benadrukt een en andermaal dat prestatiemeting strategisch gedrag uitlokt in de vorm van productie-op-papier, die nergens op slaat (De Bruijn, 2001, 2007). Hoe je streefcijfers op een wat *perverse* wijze kunt halen? Een voorbeeld is het veiligheidsbeleid; we richten ons op de politiechef (zie ook Plasterk, 2002). De politiechef heeft te maken met delicten. Tip 1: Niet veel tijd besteden aan moeilijke gevallen of zaken. Een zelfstandig geval telt immers niet extra mee in een prestatieafspraken. Voor een meting zijn al gauw alle gevallen gelijk. Een weegfactor zal er niet zijn. Tip 2: Burgers ontmoedigen om aangifte te doen, zeker van die zaken die moeilijk oplosbaar zijn. Immers, als er veel input is (meldingen zijn), moet de output (oplossingen) ook groeien. En dat kan lastig zijn. De output-inputratio wordt gunstiger als de input afneemt bij gelijktijdige output. Tip 3: Kies een gunstige definitie van een zaak. Als een inbreker betrokken was bij tien inbraken, zorg dan



dat de tien inbraken als tien gevallen (outputs) tellen, die zijn opgelost. Tip 4: Stel dat een veroordeling van een verdachte telt in een prestatiemeting en een vrijspraak niet. Dan zal alles gezet kunnen worden op veroordelingen. Laat ontlastende informatie weg. Immers, met ontlastende informatie haal je de prestatiedoelen niet. Tip 5: Als de prestatiedoelen niet gehaald worden, is een maand voor de meetdatum is verstreken nog wel wat 'laag hangend fruit te plukken'. Even een paar avonden politiecontrole op snelheid via radar of anderszins en het doel is alsnog bereikt.

### ***Academische of reële bezwaren?***

Is dit beeld van benchmarkonderzoek dat vertrouwen in ambtenaren ondergraaft en pervers gedrag uitlokt concreet voorstelbaar? En, zijn deze bezwaren ook waargenomen? Een van ons (Korsten) heeft in het bestuurlijk deel van zijn leven, gedurende tientallen jaren, steeds wel met benchmarking te maken gehad. Er volgen vijf illustraties: een uit de wereld van de ggz (riagg), een uit de wereld van het middelbaar beroepsonderwijs (mbo, roc), een uit de universitaire wereld, een betreft gemeenten en een laatste het binnenlands bestuur. Ze worden hier weergegeven tegen de achtergrond van bevindingen uit de literatuurstudie. De benchmarks uit de hoek van riagg's, universitaire opleidingen en gemeenten betreffen min of meer integrale benchmarks. Dat wil zeggen dat een organisatie als geheel vergeleken wordt (riagg's, gemeenten) of een productiecluster (universitaire opleidingen van verschillende universiteiten op een zelfde vakgebied).

Onder meer zal blijken dat het genoemde bureaucratie- en het schuttersputje-argument *niet* zijn waargenomen. Dat benchmarking variëteit vermindert, is zwaar overtrokken want dat kan alleen als benchmarkrapporten ook benut worden en leiden tot 'leren'. Dat benchmarkonderzoek tot pervers gedrag leidt, kan voorkomen als sprake is van een verbinding van prestatiemeting met prestatiecontracten, zoals bij de politie een tijd het geval was. Maar bij vrijwillige deelname aan prestatiemeting die niet gepaard gaat met prestatiecontracten, is de inbreuk op de professionaliteit van werknemers door ons niet waargenomen. De doorwerking van benchmarkstudies ('the uses of ...') wordt door theoretici (De Bruijn, 2001, 2007; Van Dooren, Bouckaert en Halligan, 2010) blijkbaar zwaar overschat.

Deze argumenten zijn blijkbaar vooral bedachte argumenten die in de literatuur te vinden zijn. Daar tegenover staat andere oogst uit het volgende betoog, zoals inzicht in de context van de opdrachtverlening, de argumenten om deel te nemen en wat besturen of raden van toezicht doen met uitkomsten.

### ***Benchmarking bij riagg's***

Als lid van de raad van toezicht van een riagg (Riagg OZL, vestigingsplaats Heerlen) bleek midden jaren negentig van de vorige eeuw dat de ambitieuze directeur-bestuurder op een bepaald moment aan een vergelijking met vier andere riagg's wenste deel te nemen. Hij wilde in het kader van de benchmarking zijn ambitieus tonen en vergelijken en stelde bij de beste vijf riagg's van het land te willen horen. Dat streven en het gratuite 'laten we eens meedoen, het kost niet veel' leidde tot deelname aan een integrale benchmark met nog vier andere riagg's; een integrale benchmark want de benchmark ging in op het algeheel functioneren van riagg's; alle taakvelden, zoals hulpverlening aan ouderen in geestelijke nood kwamen aan bod. Het gevolg was één rapport, opgesteld na analyse door externe adviseurs, voor vijf organisaties. Het rapport is daadwerkelijk gelezen en besproken in het managementteam van de riagg- staf en in het overleg tussen raad van toezicht en directeur-bestuurder. Iedereen van de raad van toezicht had het rapport redelijk tot goed gelezen. 'Onze' riagg kwam er niet slecht uit op de verschillende onderdelen ('de scoring') maar van een daarop volgende omwenteling in het beleid was geenszins sprake. De bespreking van het rapport werd na 20 minuten beëindigd met de conclusie 'dat het aardig was geweest om eens te zien waar je als Riagg staat te midden van enkele andere riagg's en dat het rapport een versterking was van de lijn die bestuur en Raad van toezicht toch al hadden ingezet'. Zo ging het en zo gaat het wel vaker bij benchmarking.

### **Benchmarking bij roc's**

De tweede ervaring werd opgedaan in de raad van toezicht van een regionaal opleidingscentrum (roc) in het eerste decennium na 2000. In het ruim acht jaren durend toezichthouderschap van een onzer, dat gepaard ging met het bijwonen van grofweg veertig - vijftig vergaderingen, kwam een benchmark waarin alle roc's van Nederland deelnamen elk jaar wel minstens een keer langs, meestal in de vorm van een grafiek en enkele tabellen. Van toelichtende of verklarende tekst was nauwelijks sprake; de collegevoorzitter gaf doorgaans uitleg van wat er stond en hoe 'we' de uitkomst moesten begrijpen. Ook hier weer nauwelijks rapportage en analyse. Van een aanduiding van best practices was volstrekt geen sprake en aanbevelingen ontbraken. De leden van de raad van toezicht keken altijd even waar de eigen roc stond en die bleek stevast in diverse opzichten te behoren tot de beste bovenste helft van roc's. Misschien geen excellente benchmarkkwaliteit, maar tevredenheid alom.

Alle toezichthouders wilden de benchmark wel onder ogen krijgen maar de gebruiksurgentie was meestal laag. Meer dan een signaalfunctie vervulde de benchmark niet. Van beleidswijziging als gevolg van die benchmarking was bij deze roc nooit sprake. In alle voorstellen die in die acht jaar door het bestuur werden gedaan kwam de benchmarking wel af en toe voorbij maar over het algemeen zelden als argument voor het een of ander, anders dan 'dit zijn we van plan'. Dat kwam natuurlijk mede door de lage urgentie. 'Onze' roc scoorde immers niet zo dat het behoorde tot de slechtste helft of het slechtste kwart van de roc's. Op grond hiervan valt de veronderstelling te formuleren: bij benchmarking zullen vooral de deelnemers met een score in het laagste kwart of – ruimer – evident in de laagste helft in de benchmarkuitkomsten een aansporing tot verbetering ervaren; tenminste als de uitkomsten relevant, actueel, begrijpelijk en interpreteerbaar zijn. Maar is dan ook sprake van een vertaalslag naar handelen? Weet men wat te doen? Niet zo maar. Verderop keert dit onderwerp terug.

Eind 2012 zijn enkele benchmarks geraadpleegd die uitgevoerd zijn door Pwc en uitgegeven werden door de MboRaad. Een van deze benchmarks betreft studiesucces (Pwc, 2013). Deze rapporten springen positief in het oog doordat een rapportage *per instelling* plaatsvond, een positief punt uit het oogpunt van benuttingskansen. Tabellen en diagrammen geven de positie aan van de betreffende roc te midden van andere roc's. Toelichtingen op tabellen zijn in sobere vorm aanwezig.

Een van de twee bestuursleden is schriftelijk gevraagd wat hij van deze brede benchmarks anno 2013 vond. Hij constateert dat het sectororgaan, de MboRaad, de benchmarking volop steunt, dat alle roc's meedoen, dat alle roc's meebetalen aan de benchmark, dat veel van de bedrijfsvoering in de benchmarking wordt meegenomen, dat de benchmark zuiver kwantitatief is, dat volop sprake is van rangordening per indicator, dat de 'best practices' niet expliciet aangewezen worden en dat de toelichting op tabellen en diagrammen heel summier is. Van een echte verschillenanalyse is geen sprake. De bestuurder voegt toe:

'We weten natuurlijk wel waar de andere roc's ongeveer staan. Op basis daarvan vraag je collegabestuurders in den lande wel eens: volgens mij scoren jullie op item x erg hoog: hoe doen jullie dat?'

Hij acht de benchmarks een handig instrument om eigen prestaties afgezet te zien tegen die van de rest. 'Voor mij heeft dit bestuurlijke waarde'. Ik moet mij regelmatig afvragen:

'Wat ik nu wil (hogere ratio's bij voorbeeld) kan dat eigenlijk wel? Dan is het nuttig om de reeds bestaande score te zien in groter verband'. 'Ik ontvang die benchmarks graag want zo kan ik eigen cijfers meer betekenis geven'. 'We bespreken die benchmarks ook als bestuur en in de conversatie met de raad van toezicht'.

### **Vergelijking van universitaire opleidingen**

Een derde ervaring betreft de visitatie van universitair onderwijs in de vorige eeuw. Een van de eerste visitaties van bestuurskunde-opleidingen was zowel kwalitatief als kwantitatief gericht, een afwijking van het hoofdbeeld bij benchmarking. De visitatie was ingekaderd in afspraken die de

colleges van bestuur met elkaar gemaakt hadden en was uitgewerkt door de VSNU-staf. Dat leidde tot de afspraak over een startmoment, de samenstelling van een visitatiecommissie van onafhankelijke leden, een protocol en een zelfevaluatie op basis van het protocol. Eens per vijf jaar zou elke wo-opleiding in een dakpansgewijze systeem gelijktijdig beoordeeld worden, waarbij elke opleiding eens per vijf jaar aan bod zou komen. Een visitatiecommissie zou dan ineens alle opleidingen van eenzelfde soort bekijken, ook door de zelfevaluatie tegen het licht te houden en kritisch ter plaatse te gaan kijken en vragen te stellen aan staf, studenten en afgestudeerden. Er kwam een visitatierapport per opleiding tot stand en een overall-rapport. In beide gevallen ging het om uitgeschreven tekst met oordelen en onderbouwingen, waarin ook cijfers aan bod kwamen. In die visitatie werden ook rangordes aangebracht en 'best practices' aangewezen. Ergo, hier was sprake van visitatie met zekere benchmark-elementen. Later zijn de benchmarkelementen uit de visitatierapporten verdwenen. De suggestie van 'best practices' werkte weinig door. Wie gaat immers navlooiën als blijkt dat de onderwijsorganisatie bij opleiding x van universiteit 1 goed is (best practice) en slecht bij 2. Dat betekent dat de 'best practice' geen makkelijk transport kent. De visitaties van deze opleiding en andere opleidingen maakte overigens duidelijk dat elke universiteit altijd wel ergens goed in is. Er valt in de 'ranking industry' te scoren: universiteit 1 is de beste van de grote klassieke universiteiten, universiteit 2 is de beste van de middelgrote en universiteit 3 van de universiteiten voor afstandsonderwijs. Zo wint iedereen een prijs. Benchmarking kan zo leiden tot een collectief 'hiep hiep hoera'-effect (Abma en Korsten, 2009).

### **Benchmarking van gemeenten**

Een vierde ervaring met benchmarking betreft de analyse van 'De staat van de gemeente' (Aardema en Korsten, 2005; Abma en Korsten, 2009). In het midden van het eerste decennium na 2000 nam de Vereniging van Gemeentesecretarissen (VGS) het initiatief tot een instrument dat 'de staat van de gemeente' zou gaan heten. In VGS-kringen had men de buik vol van een rijksoverheid die steeds maar weer gemeenten tot rapportages dwong over de uitvoering van rijksbeleid. Als alle monitors nu eens vervangen konden worden door één instrument, aldus de VGS, dan zou dat doelmatiger zijn. Een werkgroep werd opgetuigd om tot een instrument te komen. Deze stap leidde tot een instrument waarbij de oordelen van burgers over gemeenten werden 'opgehaald' en de burgers in zes rollen aan bod kwamen: als kiezer, als klant, als wijkbewoner, als belastingbetaler, als onderdaan en als partner. Een gemeente kon op basis van pakweg zestig indicatoren, verdeeld over de zes rollen, beoordeeld worden. Na enkele pilots is dit instrument bij veel gemeenten toegepast. Abma en Korsten (2009) hebben in 'Gemeenten in rapportcijfers' hiervan verslag gedaan. Het bleek dat gemeenten over het algemeen op de rol 'kiezer' het slechtste functioneerden. Dat impliceert dat burgers relatief kritisch zijn op het coelge van B&W en de gemeenteraad. Het vertrouwen daarin is in veel gemeenten, althans volgens 'De staat', gematigd negatief. Overigens heeft het VGS-initiatief tot benchmarking er niet toe geleid dat het aantal monitors op rijksniveau waaraan gemeenten moesten meedoen afnam. Logisch want de rijksoverheid is een eilandenrijk waarin ministers ministerieel verantwoordelijk zijn voor een 'eiland'. De rijksoverheid spreekt dus niet altijd met één mond. Een effect dat de benchmark 'De staat van de gemeente' tot de politiek-bestuurlijke dood van tal van monitors leidt, trad niet op. Logisch want de VGS had niet de macht om met het kabinet in gesprek te komen en het bestuur van de Vereniging van Nederlandse Gemeenten haalde de hete kolen niet uit het vuur.

### **Een systeem bouwen?**

De vijfde ervaring slaat op de participatie in de Raad voor het openbaar bestuur (Rob). Een van de auteurs (Korsten) was daar in het eerste decennium na 2000 lid van. De Raad bracht in dat verband een advies uit aan het toenmalige kabinet-Balkenende over 'presteren door leren'. Dat advies handelde over de rol van benchmarking in het binnenlands bestuur. Om beslagen ten ijs te komen, werd niet alleen literatuurstudie verricht en werden benchmarks bekeken maar is ook een expertmeeting gehouden met kenners op het gebied van lokaal bestuur, van statistiek, kengetallen

en begrotingen, en van benchmarking zelf. Met de voormalige wethouders Walter Etty (Amsterdam) en Paul Bordewijk (Leiden), met SCP-medewerker Vic Veldheer, enz. Er werd ook kennis genomen van de aanpak die het Sociaal en Cultureel Planbureau volgde om prestaties te vergelijken, zoals die later terug te vinden is in *'Maten voor gemeenten'* (o.a. Kuhry en Jonker, 2009). De voorbereidende raadscommissie was op grond daarvan positief over benchmarking maar dan wel in een kleine setting, dus een vergelijking van bij voorbeeld 15 organisaties. Want de taak van onderzoekers was om in geval van benchmarking te beschikken over relevante en actuele cijfers over de kern van een prestatie, productie of proces (en een dekking van de essentie) en om vooral te begrijpen wat er *achter* de scores en de kwantitatieve vergelijking schuil gaat. Een benchmarkrapportage moet idealiter duidelijk maken waarom een score hoog of laag is. Benchmarking moet immers geen moderne tovenarij worden; geen dans rondom een getallenmix die niet te begrijpen is.

De suggestie om door middel van een uitgebreid interbestuurlijk stelsel van benchmarking - dat mede gebaseerd was op statistische gegevens die toch al beschikbaar waren en vormgeving van een stelsel waarin alle gemeenten participeren -, om zo te komen tot een bijdrage aan de verantwoording door middel van de VBTB-operatie op rijksniveau was volgens de experts, en in de voetsporen daarvan volgens de Raad voor het openbaar bestuur, gedoemd te mislukken (verg. ook Haselbekke, 1998). Het is niet goed mogelijk om voor alle taakgebieden en ambities, zoals die ook in een gemeentebegroting aan bod komen, relevante, dekkende en actuele kernvariabelen te vinden die iets zeggen over de omvang en kwaliteit van de prestaties (voor reïntegratiebeleid, subsidiebeleid, ouderenbeleid, jeugdbeleid, sportbeleid, recreatiebeleid, ruimtelijke ordening, etc.), die te vergelijken en te interpreteren. Derhalve is het ook niet goed mogelijk om via een dergelijk systeem eenvoudig antwoorden te geven op informatieverzoeken vanuit het parlement in de geest van: kunt u eens aangeven hoeveel havengemeenten jaarlijks uitgeven aan budget voor het verminderen van havenslibproblemen?; welke havengemeente biedt in welk opzicht 'best practice'.

### **Enkele conclusies**

Uit deze analyses blijkt dat benchmarking niet steeds (goed) begrijpelijke rapportages biedt en gericht is op een verklaring van verschillen en ook niet expliciet en nadrukkelijk op verbeteren. Een gemeentelijke informant zegt het zo:

'Het college heeft het altijd over cijfers, dus die willen eigenlijk alleen het cijfer over klanttevredenheid horen. Terwijl ik weleens denk: wat zegt dat nou over dienstverlening? En hoe moeten we dit dan verbeteren? Daar schiet benchmarking wel tekort' (Ten Tije e.a., 2010: 36).

Soms heeft een opdracht, zoals bij de riagg, iets van een verkenning ('laten we eens meedoen'). Of het had iets van een rituele dans ('we doen al jaren zo'n benchmark'). De lezer moet beseffen dat verbeteren bestuurlijk en ambtelijk niet altijd makkelijk is; het vergt juist energie; en daar tegenover staat 'behoud het goede'. Dat benchmarking tot verbeteringen leidt is dus verre van vanzelfsprekend. Een mythe lijkt zo ondergraven te worden. Verderop komen we hierop terug en zullen we bezien of sprake is van 'leren'.

Zijn we de in de literatuur genoemde negen bezwaren tegengekomen? Het normalisatie-argument impliceerde dat organisaties door vergelijking en het overnemen van goede praktijken meer gelijk worden. Dit bezwaar tegen benchmarking is niet in de genoemde praktijken vastgesteld. Daarvoor blijken de producten en processen in organisaties toch teveel contextgebonden. Een praktijk dat organisaties allemaal op de beste willen gaan lijken en dit effect ook optreedt, is niet waargenomen. Het beeld van wie de beste is of wat het beste is, is niet opgenomen of is onscherp en teveel gehuld in 'mist'. Blijkbaar wijkt benchmarking in de publieke sector toch af vanwege de aard van gemeten en vergeleken producten, processen en contexten toch teveel af van private organisaties, zoals bedrijven. Dat een benchmarkonderzoek de bestuurlijke ambitie zou remmen is niet gebleken.

Dat benchmarking tot wel heel veel veranderingen in deelnemende organisaties zouden leiden is in de praktijk van de benchmarking zeker niet waargenomen: geen 'double loop learning' als gevolg van benchmarking bij het Riagg, de universiteit of het roc.

Het reductie-argument is wel aangetroffen evenals het monocentrisme-argument. Benchmarking hield in de casus, die hier de revue passeerden geen rekening met het bestaan van de netwerksamenleving en het feit dat organisaties voor het realiseren van tal van doelstellingen mede afhankelijk zijn van andere organisaties. Van het optreden van meer bureaucratie als gevolg van benchmarking was geen sprake. Dat benchmarking de verkokering en fragmentatie binnen organisaties zou versterken, bleek ook niet. Zo gezien moeten we de genoemde nadelen van benchmarking overtrokken noemen. Maar omgekeerd bepaalde voordelen waren er ook niet. Dat benchmarking zou werken als 'early warning' is niet gebleken.

We moeten dus aan de hand van dit beperkt aantal casus vaststellen dat de kwaliteit van benchmarking overschat wordt: blijkbaar is het formuleren van best practices verre van vanzelfsprekend; blijkbaar is verre van duidelijk wat een bestuur na afloop van een benchmarkonderzoek aan consequenties kan of moet verbinden aan de uitkomsten. Uit het voorgaande komt voorts een bevestiging naar voren van wat Helder (2001: 14) opmerkt naar aanleiding van ervaringen met benchmarking in de gemeenten Barneveld, Heusden, Noordoostpolder en Soest:

*'Een benchmark in zijn pure vorm is vaak moeilijk te realiseren'.*

Tal van nadelen die aan benchmarkonderzoek worden toegedicht, blijken onvindbaar of zwaar overtrokken, mede doordat benchmarking niet is (geworden) wat we verwachten. Deze stellingname heeft echter een nadeel. Het aantal casus waarop deze stellingname is gebaseerd is nog te beperkt om te kunnen generaliseren. We zullen dus verder moeten gaan met onderzoeken.

## **5 Veronderstellingen achter benchmarking**

Laten we nog eens verder graven. Achter benchmarking gaan veronderstellingen schuil. We noemen er een paar.

*1 Appels met appels vergelijken:* Een eerste veronderstelling is dat het mogelijk is om vergelijkbare overheidsorganisaties op te sporen waartussen een vergelijking echt zinnig is. Bij benchmarking moet een onderzoeker immers appels met appels vergelijken. Het heeft over het algemeen bij voorbeeld geen zin om prestaties van de grootstedelijke gemeente Amsterdam ten aanzien van wegonderhoud te vergelijken met pakweg de kleine en veelkernige plattelandsgemeenten Littenseradiel (29 kernen) of Sud-West Fryslan (69 kernen). Er wordt wel geprobeerd om echt vergelijkbare organisaties op te sporen, maar over het algemeen gebeurt dat op indirecte criteria, zoals een vergelijkbare gemeentegrootte qua aantallen inwoners. Soms zoeken vergelijkbare gemeenten elkaar wel op, zoals drugsgelateerde grensgemeenten.

*2 Commensuratie:* Het verrichten van een internationaal vergelijkend onderzoek naar pensioenstelsels is geen eenvoudige kost en het rapporteren hierover en begrijpen van de uitkomsten evenmin. Hoe als beleidsvoerders of journalisten over een dergelijke omvattende vergelijking te communiceren? Benchmarking heeft het voordeel dat eenvoudige, compacte en gestandaardiseerde indicatoren die weinig interpretatie behoeven, gebruikt worden ter vergelijking. Ook in geval van pensioenstelsels.

Denk aan het aantal geflitste hardrijders op snelwegen als indicator voor het halen van een bepaald prestatiedoel dat is afgesproken tussen minister en politiekorpsen. Denk aan de vermindering in het aantal buitenlandse drugskopers in coffeeshops als indicator voor een succesvolle(re) politiek tot reductie van drugsverkoop.

Kwaliteiten worden zo gereduceerd tot kwantiteiten, tot 'behapbare' verschillen in aantallen. Dat heet 'commensuratie' (Espeland & Suader, 2007; Espeland & Stevens, 1998, 2008). Die aanpak is aantrekkelijk want zo is een complexe werkelijkheid eenvoudiger in beeld te krijgen. En de interpretatie is op het oog eenvoudig. Daardoor kunnen beleidsmakers, journalisten en actieve burgers moeilijke zaken begrijpen, complexiteiten van zich afschudden en zich er in principe makkelijker op beraden. Dat zou de besluitvorming kunnen vereenvoudigen (Peeters en Verschraegen, 2013: 135). Er is volgens ons echter ook een *commensuratieparadox*. Die houdt in dat wat benchmarking in aanzet is, een reductie van complexiteiten – van kwaliteit naar kwantiteit genoemd – , in werkelijkheid niet optreedt: de complexiteit wordt helemaal niet kleiner gemaakt maar vergroot. Waardoor? Doordat er in tijdreeksen wisseling in scores optreden die de onderzoeker niet kan verklaren of doordat onbekend is wat achter de scores aan oorzaken schuilgaat. Emile Cleuver en Jos Blank (2006) laten in hun beschouwing over verschillen in kostenstructuur bij politiekorpsen dat de paradox bestaat.

*3 De gemeente als opdrachtgever.* In de benchmarkliteratuur (o.a. Camp, 1992; Keeley, 1996) worden processtappen beschreven. Die gaan uit van het gegeven dat de organisatie zelf kiest voor benchmarking , daarvan de omvang bepaalt, zelf zijn te benchmarkingpartners uitzoekt en behoorlijk wat invloed heeft of moet hebben op de indicatoren die worden gemeten, de informatie die moet worden aangeleverd, enzovoorts. Met andere woorden: het initiatief ligt bij de organisatie, (i.c. de gemeente) die wenst te benchmarken.

De praktijk bij gemeenten is echter geheel anders. Niet de gemeente die wenst te benchmarken staat 'aan het roer' maar de benchmark staat centraal. Er liggen panklare benchmarks 'op de plank' die uitgevoerd kunnen worden als er voldoende deelnemers zijn. De adviesbureaus die de benchmarks aanbieden, schrijven gemeenten aan dat 'ze weer mee kunnen doen'. De te benchmarken gemeente hebben op de inhoud van de benchmark geen invloed. KING heeft in het verlengde daarvan ook benchmarks gecertificeerd. Deze staan dus vast.

Bij meer diepgaande onderzoeken als bestuurskrachtmetingen en visitaties is de opdrachtgevende rol van de gemeentebesturen veel groter dan bij benchmarks. Zij stellen opgavenprofielen vast, ze zijn betrokken bij de normenkaders, enzovoorts. Dit bevordert de bestuurlijke en politieke betrokkenheid. Iets wat juist een gemis is bij benchmarking. En wellicht voor een deel kan verklaren waarom benchmarkresultaten sommige besturen wel bereiken maar heel vaak de gemeenteraden niet.

*4 De kern van de zaak indiceren.* Achter benchmarking gaat ook de veronderstelling schuil dat de prestatie-indicatoren die vergeleken worden altijd *de kern van de zaak moet raken* (Bouckaert e.a., 2011). Die veronderstelling kan echter in vraag gesteld worden.

Wie de kwaliteit van dienstverlening door een gemeentelijke afdeling burgerzaken op vergelijkende wijze analyseert, bijvoorbeeld op het punt van snelheid van documentverstrekking moet beseffen dat burgerzaken hier en daar een afzonderlijke afdeling is maar elders is opgegaan in een bredere afdeling publiekszaken met een gevarieerd en complexer productassortiment dan alleen het verkrijgen van een uittreksel uit het bevolkingsregister of het verkrijgen van een paspoort of identiteitsbewijs. Het probleem van wat er achter de vlag van vergelijking schuilgaat, was jaren terug al gebleken bij het moeilijk vergelijken van gemeentebegrotingen.

Dit aandachtspunt van de kern van de zaak betrekken in een benchmark moet overigens niet onderschat worden. Bij het overwegen van benchmarking van organisaties in de geestelijke gezondheidszorg of ziekenhuizen kan overigens onder professionals, die gevraagd worden om medewerking bij de informatieverstrekking wel ophef ontstaan omdat ze zich verbazen of zelfs ergeren aan de 'platte' indicatoren. Een van ons heeft dit zelf meegemaakt. Psychiaters zijn een

voorbeeld van een categorie professionals die griffelen bij het vergelijken van de productie waarbij aantallen behandelingen van cliënten of patiënten telde. Deze variabele is sterk te manipuleren. Men zei: 'hoe wil je het hebben?'. Als het moet schrijven we drie behandelingen achter elkaar uit. En deze variabele 'aantallen behandelingen' zegt weinig over het oplossen of redresseren van geestelijke nood of andere problemen, waar het echt om gaat.

*5 Vergelijking van cijfers in meerjarenperspectief:* Een veronderstelling houdt in dat de benchmark-onderzoeker over de relevante data beschikt en vergelijking in de tijd per indicator voor prestaties en kosten, bij een doelmatigheidsanalyse, mogelijk is. Aan deze meetvoorwaarde wordt lang niet altijd voldaan (Blank, Felsö en De Groot, 2011: 9).

*6 Analyse van oorzaken nodig:* Een zesde veronderstelling slaat op de analyse van oorzaken voor overeenkomsten en verschillen op indicatoren tussen organisaties, bij voorbeeld gemeenten (Blank, Felsö en De Groot, 2011). Die analyse is nodig om te achterhalen aan welke knoppen een overheidsbestuur moet proberen te draaien om tot verbeteringen te komen. Een benchmarkstudie naar gemeentelijk afvalbeheer toont dit aan. Doelmatigheidsverschillen tussen gemeenten bleken veel meer afhankelijk te zijn van de stedelijkheidsgraad van gemeenten dan van de organisatie van de afvalophaal en verwerking.

*7 Goede praktijken zijn transporteerbaar naar andere organisaties:* Een zevende veronderstelling betreft 'best practices'. Stel dat het lukt om 'best practices' te traceren dan is een best practice nog niet zo maar transplanteerbaar (overdraagbaar) naar een andere gemeente. Achter benchmarking gaat in feite de veronderstelling schuil dat een contextverschil in relevante kenmerken irrelevant is. Alsof een hartchirurg zegt dat elk excellent kloppend hart van welke burger dan ook is over te plaatsen naar een patiënt die een nieuw hart zal krijgen. Uit onderzoek is bij voorbeeld bekend dat niet elke in principe capabele leidinggevende geschikt is om willekeurig welke organisatie succesvol te leiden. Er moet 'fit' zijn: de leidinggevende moet op een bepaalde plaats en tijd passen bij een organisatie met bepaalde kenmerken. Zo is het op meer gebieden. 'Best practices' zijn niet zonder meer overdraagbaar, zo bleek uit een toekomstgerichte visitatie van Vlaamse grote steden in opdracht van de Vlaamse overheid waarbij een van ons sterk betrokken was. Deze opvatting staat nog los van een andere kanttekening bij 'best practices'. Zo stellen Herweijer en Noordam (2005: 2) dat benchmarking wel leerzaam kan zijn, maar dat het niet verstandig is om voor elk te benchmarken kwaliteitsaspect te mikken op de hoogste score. Het optimaliseren van het ene aspect (bijvoorbeeld efficiency) kan dan ten koste gaan van een ander aspect (bijvoorbeeld rechtmatigheid). In die optiek is het optrekken aan de 'best practice' dus geen vereiste.

We gaan ervan uit dat deze veronderstellingen juist blijken te zijn: appels zijn met appels te vergelijken; de stap van kwaliteit naar kwantiteit met onzekerheidsreductie als gevolg, treedt op en bevordert begrip; de kern van een benchmarkobject is door de keuze van indicatoren te vatten; er is een meerjarenperspectief mogelijk; een oorzakenanalyse maakt deel uit van benchmarking; goede praktijken zullen overdraagbaar zijn. De werkelijkheid is echter niet zelden een andere (Blank, Felsö en De Groot, 2011; Hulst en De Groot, 2011; Bouckaert e.a., 2011; Peeters & Verschraegen, 2013). Toegespitst op gemeentelijke benchmarking blijkt het volgende.

De meeste gemeentelijke taakgebieden en ambities en daarmee het meeste gemeentelijk beleid blijkt helemaal niet aan benchmarking onderworpen te worden omdat het zich blijkbaar niet kwantitatief laat vangen, andere waarden dan kwantiteiten een rol spelen en specifieke aanleidingen om benchmarking te starten blijkbaar ontbreken. Lettend op begrotingen blijkt dat in tal van grote en middelgrote gemeenten gepoogd wordt vanuit outputsturing een breder beleidspectrum in kengetallen of andere cijfers te vangen (Ter Bogt, 2005). Voor zover benchmarking wel voorkomt, blijkt dat de kernvariabelen in benchmarkings niet altijd gevonden worden, cijfers over een reeks jaren soms ontbreken, een oorzakenanalyse in een benchmarking vaak achterwege blijft en goede

praktijken vaak niet weergegeven worden of incidenteel wel weergegeven worden maar andere organisaties niet op zoek gaan naar wat die beste praktijken werkelijk inhouden (Helder, 2001; De Groot, 2004; Ter Bogt, 2005; Lapperre, 2007; Ananias, 2007).

## **6 Uitkomsten van benchmarking: enkele casus**

Benchmarking kan inderdaad gericht zijn op financieel-economische aspecten van overheidshandelen, waaronder doelmatigheid, zoals de derde mythe aangeeft. Maar benchmarking is niet uitsluitend gericht op doelmatigheid. We geven enkele voorbeelden van een dergelijk onderzoek gericht op gemeenten: over servicegerichtheid; burgerzaken; afvalbeheer; sportzaken; werk en inkomen; gemeentelijke overhead. De betrokken onderzoeksorganisaties zijn Berenschot, SGBB/BMC, Stimulansz en IPSE. Andere sectoren waar gemeentelijke benchmarking plaatsvond zijn onder meer: rioleringszorg (Langeveld e.a., 2006); bouw- en woningtoezicht (Lapperre, 2007); gemeentelijk ondernemingsklimaat (Ecorys, 2011).

### ***Eerste voorbeeld: benchmark servicegerichtheid***

Er is benchmarkonderzoek gedaan naar de kwaliteit van dienstverlening. Onderzoek volgens *'De staat van de gemeente'* laat zien dat gemeenten op basis van gegevens over het omgaan met burgers in verschillende rollen, gemeenten burgers in de rol van klant over het algemeen voldoende of beter tegemoet treden (Abma en Korsten, 2009).

Basile Lemaire (2012) is anders te werk gegaan. Hij heeft als *'mystery shopper'* alle gemeenten van Nederland benaderd met een verzoek om informatie over zijn eventuele vestiging in de gemeente. Hoe stellen gemeenten zich dan op: toegankelijk?; reageren ze snel?; geven ze de juiste informatie?; en denken ze met de burger mee? In zijn hoedanigheid van Bas Burgemeestre was Lemaire op zoek naar de servicegerichtheid van gemeenten, die hij operationaliseerde naar toegankelijkheid, snelheid van reageren en het geven van de juiste informatie. Blijkt daaruit dat gemeenten weten hoe ze met een burger moeten omgaan? Van alle gemeenten scoort 44 procent op de drie gebieden gemiddeld een voldoende of hoger, 34 procent scoort matig of slecht, en 22 procent van de gemeenten antwoordt niet, niet inhoudelijk en niet tijdig. Van de vier grootste gemeenten scoort alleen Rotterdam een voldoende. De gemeenten in de drie Randstedelijke provincies scoren gemiddeld het laagste. Alle resultaten samen geven het beeld dat een aantal gemeenten teveel naar binnen gericht is en niet goed in staat is om zich te verplaatsen in de behoeften van de vragende burgers. De servicegerichtheid van deze gemeenten zou volgens deze *'mystery shopper'* voor verbetering vatbaar zijn.

Deze benchmark is geen benchmark waartoe een aantal gemeenten samen opdracht gaven. Dus is niet na te gaan wat gemeenten als opdrachtgevers hebben gedaan met de resultaten van de benchmark van de opdrachtnemer.

### ***Tweede voorbeeld: benchmark burgerzaken***

Gemeentelijke organisaties verstrekken rijbewijzen, uittreksels uit het bevolkingsregister en andere documenten. De kostendoelmatigheid van de gemeentelijke afdelingen burgerzaken is in het kader van een benchmark vergeleken. Daaruit blijkt dat over de periode 2005-2008 de dienstverlening van het gemeentelijke bureau burgerzaken *productiever* is geworden. Deze verbetering gaat samen met een forse toename van het aantal afgegeven documenten.

In dit onderzoek is niet specifiek gekeken naar de invloed van het aantal vestigingen of het aanbod van digitale dienstverlening op de scores. De betekenis van het feit dat de dienstverlening van burgerzaken meer of minder geïntegreerd is met andere dienstverlening van de gemeenten onder de vlag publiekszaken bleef ook buiten beeld.

De gemiddelde doelmatigheidsscore in 2008 bedraagt volgens de benchmark burgerzaken 86,2 procent; een kwart van de gemeenten realiseert de maximale doelmatigheidsscore van 100 procent.



De doelmatigheid blijkt zich daarmee, gelet op de score van 86 procent, op een hoog niveau te bevinden. Maar de ene gemeente werkt doelmatiger dan de andere. Met de doelmatigheidsscore wordt dus duidelijk hoeveel ruimte een gemeente nog heeft voor verbetering. Voor *driekwart* van de gemeenten is er ruimte om de doelmatigheid te verbeteren richting 100 procent, aannemend dat die maximale score haalbaar is (Van Hulst en De Groot (2011)).

Fusiegemeenten scoren gemiddeld lager op doelmatigheid van burgerzaken. Gemeenten die na 2005 zijn heringedeeld scoren 79 procent, 7 procentpunt minder dan andere gemeenten. 'Vrij goed' overigens. Volgens de onderzoekers komt dit percentage mogelijk door tijdelijke aanpassingseffecten, zoals het blijven aanbieden van diensten van burgerzaken op meerdere locaties in de fusiegemeenten. Overigens lijkt het (nadelige) effect van herindeling op de doelmatigheidsscore na verloop van tijd weg te ebben: gemeenten die tussen 2000-2004 zijn gefuseerd, scoren gemiddeld 84 procent.

Hoewel pas gefuseerde gemeenten dus tijdelijk minder doelmatig blijken, ervaart twee derde van de Nederlandse gemeenten schaalvoordelen voor de dienstverlening van de afdeling burgerzaken. Een schaalvoordeel houdt in dat wanneer er op grotere schaal geproduceerd wordt, de kosten minder snel toenemen dan de productie. De onderzoekers stellen vast dat 20 procent van de Nederlandse gemeenten een optimale schaalgrootte heeft voor de dienstverlening van burgerzaken; bij 15 procent is er sprake van schaalnadelen.

#### ***Derde voorbeeld: benchmark afvalbeheer***

In 2011 is een benchmarkstudie van de TU Delft (in samenwerking met de UTwente) gepubliceerd naar gemeentelijk afvalbeheer (Felsö, De Groot en Van Heezik, 2011). Het betrof een onderzoek naar de productiviteit van het gemeentelijk afvalbeheer in de periode 2001-2009 onder ruim 400 gemeenten. Welke uitkomsten leverde dat op? Na een flinke initiële daling in de periode 2001-2004 herstelt de productiviteit en is die in 2009 circa 1 procent lager dan in 2001, aldus de drie auteurs.

De kosten per product verschillen tussen gemeenten aanzienlijk, maar deze verschillen vloeien grotendeels voort uit *omgevingskenmerken*, niet uit beleidskeuzen of organisatiekenmerken. Zo bepaalt de mate van stedelijkheid in grote mate de inzamelmethode en *daarmee* ook de kosten. Afvalbeheer in landelijke gebieden blijkt om en nabij 7% goedkoper dan in sterk stedelijke gebieden. Na correctie voor de omgevingsfactoren, wijken de kosten per product van de gemiddelde gemeente niet meer dan ruim 3 procent af van de goedkoopste gemeente met dezelfde kenmerken. Ondanks verschillen in aanpak tussen gemeenten, zijn de verschillen in doelmatigheid opmerkelijk genoeg gering. De gemiddelde doelmatigheid bedraagt volgens een indicator 96,7 procent.

Het is niet nodig te blijven hangen bij een verschil in stedelijkheid. Volgens deze benchmarkstudie naar afvalbeheer is uitbesteden zinnig, zij het beperkt zinnig. Over de periode 2001-2008 genomen, zijn gemeenten die afvalinzameling uitbesteden aan een privaat bedrijf gemiddeld het meest kostendoelmatig. Dat betekent dat het streven naar bedrijfsmatigheid volgens NPM, waarmee we de rapportage begonnen, wel degelijk enige werking had. Maar daar staat tegenover: gemeenten die de dienst door een buurgemeente laten uitvoeren zijn het minst kostendoelmatig. Gemeenten die een eigen inzamelingsdienst hebben, in een samenwerkingsverband participeren of een overheidsbedrijf inschakelen, liggen hier tussenin.

Zoals gesteld, de verschillen in doelmatigheid zijn *gering*. Dat is het hoofdbeeld uit het onderzoek. De organisatie van afvalbeheer maakt blijkbaar niet zoveel uit omdat factoren die *niet* met organisatie te maken hebben, de verschillen hoofdzakelijk bepalen. Gemeenten die afvalinzameling uitbesteden aan een private partij doen dit circa 1 procent goedkoper, let wel 'slechts één procent', dan gemeenten die dat niet doen. Dat is een uitkomst die de hele beweging naar bedrijfsmatigheid *relativeert*.

Dat gemeenten met meer dan 100.000 inw., een grens die in 2012 is voorgesteld door het kabinet-Rutte II en door minister Ronald Plasterk is uitgewerkt in plannen voor de komst van nieuwe gemeenten, meer doelmatig zijn in het afvalbeheer kon voor althans dat beleid niet aangetoond worden, aldus de benchmarkspecialisten Blank en De Groot (2012).

#### ***Vierde voorbeeld: benchmark sport***

Is de sportdeelname hoog? Wordt veel geld gespendeerd aan accommodatieonderhoud? Onderzoeksbureau SGBO is in 2007 een benchmark gestart met 26 deelnemende gemeenten om antwoord hierop te zoeken. De sportbenchmark is ontwikkeld voor en door gemeenten. De benchmark is een jaarlijks terugkerende activiteit, zodat gemeenten hun gegevens over meerdere jaren kunnen vergelijken. Het doel van deze benchmark is te komen tot een betere uitvoering van het sport(accommodatie)beleid. Met behulp van een internetvragenlijst vullen gemeenten gegevens in. Dit zijn zowel (financiële) cijfers als vragen over beleid. Op basis van de cijfermatige gegevens berekent SGBO kengetallen. De deelnemers aan de benchmark ontvangen een rapport waarin de kengetallen van de deelnemende gemeenten met elkaar worden vergeleken. Daarin komt ook naar voren op welke punten een gemeente goed 'scoort' en op welke minder. In kringbijeenkomsten vindt bespreking plaats van de resultaten en van de verschillen tussen gemeenten. Het verhaal achter de cijfers, het identificeren van 'good practices' en het leren van elkaar om te komen tot verbetering van het eigen beleid, staan centraal in de kringbijeenkomsten.

De gemeenten zijn ingedeeld in drie kringen: kring grote gemeenten (>120.000 inwoners), kring middelgrote gemeenten (tussen 60.000-120.000 inwoners) en de kring kleine gemeenten (tussen 20.000-45.000 inwoners). Wat bleek? Vrijwel alle gemeenten gebruiken een sportnota als basis voor beleidontwikkeling. Er zijn lichte accentverschillen waar te nemen in de doelstellingen van het beleid. Grote gemeenten richten zich iets meer op doelgroepen als gehandicapten, niet-westerse allochtonen en mensen met een laag inkomen. Ook topsport is voor deze gemeenten meer een beleidsspeerpunt, evenals de tevredenheid van de inwoners over de sportmogelijkheden en over de accommodaties. De kleinere gemeenten staan dicht op de sportvereniging en maken zich vooral hard voor de continuering van de vereniging en het vrijwilligersbeleid.

Sportbeleid is autonoom beleid en dus kunnen gemeenten zelf prioriteiten stellen over de inzet van eigen middelen. Dit is ook terug te zien in de vergelijking van de uitgaven van gemeenten. Deze lopen zeer uiteen, zowel wat betreft de hoogte van de uitgaven per hoofd van de bevolking als de verdeling van de middelen. De grote gemeenten geven verreweg het meeste uit aan de accommodaties. De middelgrote gemeenten zetten juist weer in op subsidies aan verenigingen en investeren relatief veel in sportprogramma's.

Het afzetten van de uitgaven tegen de inkomsten om zo de netto-uitgaven van de gemeenten in beeld te krijgen leidt tot het volgende beeld. De middelgrote gemeenten blijken dan de hoogste netto-uitgaven per inwoner te hebben, waarschijnlijk omdat zij meer investeren in het inhoudelijk sportbeleid dan de andere gemeenten.

Het gebruik van de accommodaties, de zogenaamde bezettingsgraadcijfers, is belangrijk om te bepalen in welke wijken of kernen sprake is van overbezetting of juist van leegstand. De bezetting van de sporthallen blijkt door de week onder schooltijd hoger te zijn dan 's avonds; respectievelijk 75 procent en 62 procent. De bezetting in het weekend ligt een stuk lager; rond de 40%. De gymlokalen worden eveneens meer overdag (65 procent) gebruikt dan in de avonduren (46%) met een zeer beperkt gebruik in het weekend. Het is opvallend dat het gebruik van de gymzalen onder schooltijd lager is dan van de sporthallen. Kennelijk geven veel scholen hun gymnastiekonderwijs in een sporthal. Er zijn geen duidelijke verschillen waar te nemen in bezettingsgraadcijfers tussen kleine en grote gemeenten.

De benchmark bevatte nog veel meer gegevens, zoals over lidmaatschappen en sportdeelname van verschillende leeftijdscategorieën. Voldeed de benchmark? De deelnemers zien deze als een waardevol instrument omdat cijfers naar boven komen en duidelijk wordt voor welke keuzes gemeenten eigenlijk staan.

#### ***Vijfde voorbeeld: benchmark werk en inkomen***

De vereniging van gemeentelijke diensten Werk en Inkomen Divosa startte begin 2013 de benchmark werk en inkomen, die wordt aangeprezen als ‘makkelijker en overzichtelijker’. Dat kenmerkt de nieuwe Divosa benchmark die vanaf 1 april 2013 de oude benchmarks WWB en Re-integratie vervangt. Gemeenten kunnen hiermee - vrijwel zonder handmatige invoer van gegevens - de effectiviteit van hun aanpak meten en vergelijken met andere gemeenten.

Divosa riep begin 2013 alle sociale diensten op aan de nieuwe benchmark mee te gaan doen. Niet alleen omdat leden hiermee hun resultaten kunnen spiegelen aan andere gemeenten, maar ook omdat de benchmark helpt zichtbaar te maken wat het resultaat is van alle inspanningen. Zeker in een periode waarin fors moet worden bezuinigd, loont het te kunnen aantonen hoe groot de toegevoegde waarde van sociale diensten is voor de samenleving. Bovendien helpt de benchmark de sector zich verder te verbeteren.

De nieuwe Divosa Benchmark is ontwikkeld in nauwe samenwerking met Stimulansz en SGBO/BMC en geeft eenduidig inzicht op basis van indicatoren als het aantal klanten dat uitstroomt naar werk en de kosten per klant.

Om tegemoet te komen aan de verschillen tussen gemeenten is de benchmark modulair opgebouwd. Naast een gratis variant zijn er basis- en verdiepende modules. Gemeenten die meer indicatoren willen vergelijken (dus de verdieping zoeken en meer te weten willen komen over bepaalde thema's zoals participatie, handhaving en jeugd), zijn de verdiepende modules bedoeld.

De Verdiepende module WWB is ook per april 2013 beschikbaar. Hier krijgt de gemeente naast een meer uitgebreide maandelijkse rapportage de mogelijkheid tot benchlearning in de vorm van bijeenkomsten met andere gemeenten. Ook een advies- of managementgesprek in de gemeente behoort tot de mogelijkheden. Benchlearning wil zeggen dat niet alleen wordt uitgelegd wat de verschillende cijfers voor de gemeente betekenen, maar wordt ook gekeken naar het verhaal achter de cijfers. Dit alles kan de nodige discussiestof opleveren voor gemeenten. Gemeenten kunnen op deze manier ook veel van elkaar leren.

De deelnamekosten bestaan uit een vast bedrag per deelnemer plus een bedrag per inwoner. Naast deelname aan de Divosa Basisbenchmark kunnen gemeenten afzonderlijk lid worden van een aanvullende module naar keuze. Binnenkort wordt u verder geïnformeerd over de prijzen die gaan gelden voor de nieuwe benchmark 2013.

#### ***Zesde voorbeeld: benchmark overhead gemeenten***

Hoe verhoudt de omvang van de overhead van de gemeente zich tot die van andere gemeenten? Deze vraag staat centraal in de Benchmark Overhead Gemeenten, die is ontwikkeld door Berenschot. Deze is een aanvulling op de Benchmark Ambtelijk Apparaat. De onderzoekers vergelijken gemeenten onderling op de omvang van de overhead. Sinds de start in 2002 heeft ruim tweederde van alle Nederlandse gemeenten aan deze benchmark deelgenomen. De benchmark heeft sinds 2006 jaarlijks een kwaliteitskeurmerk gekregen van het Kwaliteitsinstituut Nederlandse Gemeenten (KING).

De benchmarkresultaten geven informatie over de omvang van de overhead in vergelijking met andere gemeenten. Tevens bieden rapportages zicht op de salarislasten, de uitbestedingskosten en de zwakke/sterke kanten van de overhead. Gemeenten kunnen onderzoeksresultaten gebruiken bij overleg over de overheadomvang tussen college, ambtelijke organisatie en raad. De benchmark geeft niet het ‘finale oordeel’, maar biedt wel objectieve vergelijkingsinformatie (verder Huijben en Geurtsen, 2008).

### **Algemeen beeld**

Welk algemeen beeld komt nu naar voren uit al die benchmarking van gemeenten? Op basis van literatuurstudie wordt een schets gegeven. Meer onderzoek op dit vlak is nodig.

*Waardering:* Benchmarking wordt als kwaliteitsinstrument over het algemeen gewaardeerd (Ten Tije e.a., 2010). Maar sommige benchmarks zijn 'oliebollentests' zonder (veel) leercomponenten, menen Ten Tije e.a. (2010). Er wordt vaak ook een gebrek aan maatwerk gesignaleerd. Dat benchmarks vaak de klantbehoefte in relatie tot dienstverlening meten, blijkt niet juist maar gemeentelijke actoren signaleren op dit vlak wel een behoefte; meer benchmarking naar klantbehoeften wordt nodig geacht (Ten Tije e.a., 2010). Een doorontwikkeling van benchmarking naar (meer) benchlearning wordt mogelijk geacht.

*Welke gemeentelijke benchmark is de populairste?:* Waarstaatjegemeente.nl blijkt de populairste benchmark, aldus Ten Tije e.a. (2010: 28). Het betreft hier een doorontwikkeling van 'De staat van de gemeente', een integrale benchmark die het brede spectrum van gemeentelijke taken door middel van zestig indicatoren in beeld tracht te krijgen, volgens de pretenties van de initiërende Vereniging van Gemeentesecretarissen (VGS) en de opstellers (Aardema en Korsten, 2005).

*Inhoudelijke scoring:* Het algemene beeld dat na toepassing van de benchmarks in vele gemeenten naar voren komt, is dat veel gemeenten op een aantal rollen van burgers, waarnaar de scores geordend zijn 'boven de streep zitten', behalve op de rol van kiezer. Het vertrouwen van kiezers in gemeenteraden en colleges van B&W blijkt in een aantal kanten onvoldoende of 'op het randje' of 'dun voldoende'. Excellente gemeentebesturen en gemeentelijke organisaties bestaan niet; er blijken geen foutloze gemeenten te signaleren of gemeenten die op alle fronten goed of meer scoren (meer in: Abma en Korsten, 2009).

*Algemeen beeld niet altijd mogelijk:* Een analyse van een twaalftal veel gebruikte en geciteerde benchmarks van elektronische dienstverlening leidt tot de conclusie dat daar geen algemeen, inhoudelijk en coherent beeld van gemeentelijke dienstverlening uit op te maken is (Ten Tije e.a., 2010). De benchmarks verschillen onderling te sterk. Veel van deze benchmarks meten het aanbod en gebruik van dienstverlening, veel minder de kwaliteit en behoefte.

### **Kanttekening**

Vergelijking van deze benchmarkings over riolering, sport en andere onderwerpen geeft aanleiding tot enkele kanttekeningen.

Ten eerste dat de veronderstelling juist blijkt dat het nodig is om aan te geven waaruit verschillen tussen organisaties of productie te verklaren zijn. Dat blijkt nadrukkelijk bij afvalbeheer. Verschillen tussen gemeenten blijken deels te verklaren uit een verschil in stedelijkheid, een voor gemeentebesturen niet bestuurlijk 'manipuleerbare' variabele. Door de tracering van deze verschilverklarende factor wordt een grens gesteld aan de gemeentelijke ambities. Dat een gemeentebestuur zo maar de sprong naar de beste gemeente op het vlak van afvalbeheer kan maken is een valse illusie. Maar ook blijkt dat de organisatiewijze van afvalbeheer toch uitmaakt.

Een tweede bijzonderheid betreft de SGBO-benchmark van sport. Het betrekken van ambtenaren uit de deelnemende gemeenten bij de bezinning op 'de cijfers' via kringbijeenkomsten lijkt in twee opzichten positief. Zo wordt de kennisname van resultaten bevorderd en neemt de betrokkenheid toe, wat later van belang kan zijn om de bevindingen uit de benchmarking te benutten voor bezinning op gemeentelijke ambities. Een ander positief punt is dat de ambtenaren meedenken over de interpretatie van de scores in het licht van streefcijfers. Blijkbaar is het bij benchmarking nodig om aandacht te besteden aan optimalisatie van de bruikbaarheid van de rapportage(s) en de gebruiks- of benuttingscondities.

## 7 Bespreking van benchmarkrapportages: gebruiken, leren en verbeteren?

In vorige paragrafen is het onderwerp 'leren' al een aantal keren langs gekomen. Ook in een van de mythes. Het wordt daarom tijd om hierop nader in te gaan.

### **Benchmarking leidt niet tot radicale beleidsbreuken**

De vraag is op te werpen of bij benchmarking sprake is van *'single-loop learning'* of misschien van *'double-loop learning'*?

'Eén cirkel leren' (*'single-loop learning'*), zoals de Vlamingen zeggen, impliceert dat de waarden, referentiekaders en doelstellingen rond een organisatietaak en opgave een gegeven zijn en als gevolg van een benchmarking *niet* ter discussie komen. Hierbij ligt het eventueel verbeteren als gevolg van benchmarking dus in lijn met de bestaande strategie en is verbetering iets op het vlak van instrumentatie van beleid.

Bij 'dubbel cirkel leren' (*'double-loop learning'*) worden ook de waarden, strategie, doelstellingen en referentiekaders achter beleid ter discussie gesteld. Hier is sprake van een radicalere bezinning op beleid (Argyris & Schön, 1978; Bouckaert e.a., 2011: 108). Wat leert onderzoek naar verbeteringen als gevolg van benchmarking ten aanzien van leren? Alle aanwijzingen uit verricht empirisch onderzoek geven grond aan de waarschijnlijkheid dat benchmarking *zelden* leidt tot 'dubbel cirkel leren' (Ter Bogt, 1999, 2000, 2001a, b, 2005; Ananias, 2007; Lapperre, 2007). Met betrekking tot de politie is zelfs vastgesteld: 'Een organisatie onder druk zetten op zijn output leidt niet zonder meer tot een radicale verbetering van kwaliteit'. Wat draagt ertoe bij dat de kans op 'double loop learning' klein is? Als bij benchmarking in samenspraak met onderzochte organisaties wordt gekozen voor prestatie-indicatoren zal een organisatie *zelf* er niet toe neigen om de kans te vergroten om de eigen werkwijze ter discussie te stellen (Van der Zee, 2005: 146). 'De indicatoren zullen hoofdzakelijk gericht zijn op het presteren van "meer van hetzelfde", wat goed zichtbaar is in de kwantitatieve doelstellingen in de bewuste prestatiecontracten'. De kans op radicale verbeteringen van prestaties wordt zo vermeden (Van der Zee, 2005: 146).

Crises, zoals een bankencrisis, woningcrises of ernstige breuk in het consumentenvertrouwen in de samenleving, leiden veel eerder tot radicale veranderingen in beleid. De kans dat door benchmarking wel sprake is van 'één cirkel leren' is groter dan dat sprake is van 'twee cirkel leren'. Wat weten we hierover nog meer? We vatten leren verderop op als benutting ('utilization') van benchmarkinformatie en laten daarom het werk van Argyris c.s. verder voor wat het is (cf Korsten, 1983).

### **Geen breuk maar wel bezinning?**

Goede prestatie meting levert relevante, actuele, geldige, betrouwbare en begrijpelijke informatie op die tot bezinning en discussie kan leiden over daadwerkelijke verbeteringen in beleid en werkwijzen. Goede benchmarkrapporten zijn een opstap hiervoor en daarmee een hulpmiddel, maar ze zijn op zich niet genoeg om onmiddellijk een verbetering te veroorzaken. Ook voor een incrementele verandering (stukje bij beetje) is meer nodig. 'Measurement by itself will *not* improve performance', zegt Callahan (2007: 70).

Wat is dan zoal nodig? Daarover is veel te doen in de literatuur (o.a. Ter Bogt, 2000, 2005; Askim, 2007 en 2008; Askim, Johnson & Christophersen, 2008; Van Dooren en Van de Walle, 2008). Zijn bestuurders en politieke representanten wel ontvankelijk voor de inhoud van een benchmark? Nemen ze er kennis van of worden deze studies overgelaten aan anderen (ambtenaren of fractiespecialisten)? Heeft een rapport bestuurlijke prioriteit? En op dat moment? En stel dat men een rapport bespreekt, wat dan? De beleidskunde leert dat wie wil verbeteren op zich weer moet overzien welke verbeteringsrichtingen er mogelijk zijn en wat de consequenties zijn van een

bepaalde keuze. Ex ante evaluatie? Een cultuur gericht op reflectie is ook van belang. Dit zijn slechts enkele variabelen die inwerken op benutting of gebruik van bevindingen uit benchmarkrapporten.

### **Ontvankelijkheid voor benchmark bij bestuurders**

Onderzoek van Ter Bogt (2002, 2005) naar gemeentebestuurders laat zien dat de ontvankelijkheid voor de inhoud van verantwoordingsdocumenten, zoals benchmarkrapporten, beperkt is hoewel de portefeuille ertoe kan doen. Vaak wordt ook overschat dat bestuurders stukken grondig lezen. Dat valt in werkelijkheid nogal eens tegen. Zoals een wethouder van een middelgrote stad enkele jaren terug een keer opmerkte: 'Gelukkig leest de burgemeester altijd alle stukken en altijd heel goed. Dat stelt me gerust'. Of en wat wethouders doen, komt vaak in interactie met relevante derden, zoals ambtenaren, en via mondelinge conversatie tot stand. Wethouders vragen nogal eens 'kan ik er wat mee?' of 'moet ik er wat mee?' of 'waarbij helpt het mij?' Dan leidt de inhoud van een benchmarkrapport niet vanzelfsprekend tot een positief antwoord. Daarvoor schieten benchmarkrapporten, waartoe in de praktijk opdracht is gegeven, inhoudelijk en kwalitatief te vaak te tekort (Ananias, 2007; Lapperre, 2007). En er wordt gezondigd tegen de regel, die Martens e.a. (2002) formuleerden: dat op een benchmarkrapport altijd een nadere analyse moet volgen waardoor de benchmark naar de unieke situatie van de eigen gemeente getrokken wordt. Vaak ontbreekt, althans in de gemeentelijke praktijk, een dergelijk afzonderlijk, verbindend rapport.

### **Witte raven**

Desalniettemin, bepaalde benchmarkrapporten worden hier en daar wel instrumenteel en direct gebruikt door overheidsbesturen, bij voorbeeld bij waterschappen en het zuiveringsbeheer (Tillema en Van Helden, 2003; Brouwer en Van Helden, 2005; Van Helden e.a., 2005; Tillema, 2006) en gemeenten (Ter Bogt en Van Helden, 1994; Ter Bogt, 2001a en b; Ter Bogt, 2005) maar dat gebeurt niet zonder meer op grote schaal en intensief. We formuleren de hypothese dat rapporten over doelmatigheid daarbij meer de aandacht trekken dan rapporten over de kwaliteit van dienstverlening.

Tot benutte rapporten behoren bij voorbeeld goed interpreteerbare benchmarkrapporten die aan eisen voldeden en duidelijk maken dat een gemeente in de top zit van geproduceerd afval per huishoudens. Een willekeurig gekozen voorbeeld daarvan is het gemeentelijk beleid van de gemeente Nuth van rond 2008-2009 om mensen aan te zetten om minder grof afval te produceren, grof vuil meer te scheiden van ander vuil en dit afzonderlijk aan te bieden op milieuparken. Uit een benchmark was gebleken dat de inwoners van de gemeente Nuth volgens het college van B&W en de gemeenteraad te hoog scoorden in een benchmark op het vlak van de omvang van restafval.

Stappen we over van het ambitieuze twee-cirkel-leren naar de eenvoudige werkelijkheid, dan blijkt het volgende.

### **Nieuwsgierigheid bevredigen**

Een benchmark bevredigt vrijwel altijd het nieuwsgierigheidsgevoel. Er wordt door de meest ambtelijk betrokkenen dan ook vaak wel kennisgenomen van de belangrijkste uitkomsten. En het komt zeker voor dat kringbijekomsten plaatsvinden over bevindingen nog voor een rapport uitkomt, zoals de gemeentelijke sportbenchmark laat zien.

Maar het hoofdbeeld geeft voorsnog aanleiding tot de stelling dat het *expliciet* en *zichtbaar* gebruik van benchmarkbevindingen voor een bezinning op wel of niet voortzetting van bestaand (gemeentelijk) beleid beperkt voorkomt of geheel afwezig is (Ananias, 2007; Lapperre, 2007; Blank, Felsö en De Groot, 2011;).

Het vormen van een ambtelijke werkgroep na het uitkomen van een benchmarkrapport komt vermoedelijk slechts in een minderheid van gevallen voor. Begrijpelijk, want

- a) de bevindingen in benchmarkrapporten zijn moeilijk te interpreteren. Ze bevatten ook niet altijd streefcijfers of andere normen (dus heerst er mogelijk onduidelijkheid over wat echt nastrevenswaardig of onvoldoende is?);
  - b) 'best practices' vinden we vaak niet terug in benchmarkrapporten;
  - c) rapporten geven zelden aanbevelingen over wat te doen;
  - d) een basisrapport wordt niet gevolgd door een rapport per organisatie (waarin nader ingegaan wordt op de eigen organisatie, scores verklaard worden en aangegeven wordt welke vervolgplicht en bestuurlijke ambitie tot de mogelijkheden behoort).
- We moeten dus oog hebben voor niet – gebruik (non-use), zoals Van Dooren, Bouckaert en Halligan (2010: 131 e.v.) stellen.

### **Benuttingspotentie**

Maar hebben benchmarks dan toch niet een zekere potentie tot benutting? Potentieel kunnen benchmarks in de bestuurlijke en politieke arena's inderdaad comfort verschaffen. Benchmarkrapporten zijn via een tweede trap van de tweetrapsraket – een nadere analyse - geschikt te maken om een opstap te zijn naar discussie in de politiek-bestuurlijke arena (b.v. gemeenteraad) over de kwaliteit van wat beoordeeld is. Als een rapport ook van een begeleidende oplegnotitie van – in dit voorbeeld - B&W is voorzien, maar dit is in de praktijk over het algemeen *weinig* het geval, neemt de betekenis toe. Immers, dan zien raadsleden dat een college van B&W de uitkomsten van belang acht en de raad daarvan deelgenoot wenst te maken. Een van ons deed de ervaring in een gemeentebestuur op dat in de commissie financiën gedurende een periode van vier jaar nooit naar de in de begroting opgenomen kengetallen werd gevraagd. Raadsleden hadden geen enkele interesse in enig cijfer. De burgemeester die financiën in portefeuille had, gaf ook nooit toelicht op enig kengetal. De politiek-bestuurlijke interesse in kengetallen was in deze gemeente van bijna 20.000 inwoners kortom volkomen nihil.

### **Benchmarks geen bananenschillen**

Laten we een stapje verder zetten. Bevat een benchmarkrapport wel eens inhoud die leidt tot een politiek-bestuurlijk conflict? Bestuurders van gemeenten of provincies behoeven – in lijn met het voorgaande - nooit af te treden als gevolg van uitkomsten uit een benchmarking. Benchmarkrapporten bevatten blijkbaar *geen* politieke splijtstof. Er is sinds 1995 nog nooit een burgemeester of een wethouderscollectief of een heel college gestruikeld op een van de uitkomsten van een benchmarkrapport of het achterhouden van een dergelijk rapport. Blijkbaar zijn dergelijke rapporten niet echt bedreigend. Ze vormen geen bananenschil voor bestuurders (Korsten en Aardema, 2006; Korsten, 2010b; Aardema, Korsten, Riezebos en Van Dam, 2011).

Benchmarking van overheidsorganisaties, zoals van gemeenten en waterschappen, leidt ook zelden tot prikkelende verhalen in dag- en weekbladen. Toen in 2010-2012 conflicten in de pers werden uitgemeten, zoals over de hbo-instelling InHolland ('fraude' met diploma's) en de woningcorporaties Amarantis (financiële problemen; gebrekkig toezicht) of Vestia (financieel probleem) was een benchmark niet de aanleiding (De Jong, 2013).

Van discussie met de buitenwereld, zoals met klanten, huurders of ondernemingen, is bij het uitkomen van benchmarkrapporten ook zelden sprake. De stelling '*benchmarking leidt tot druk van "stakeholders" om tot verbetering te komen*' is zelden of nooit sprake in de publieke sfeer (Francis & Holloway, 2007). Dat hoeft nog geen definitief oordeel over nut en noodzaak van benchmarking te impliceren.

### **Verklaring**

'Benchlearning is een nog ver verwijderd ideaal'. Aldus de benchmarkspecialisten Jos Blank, Flóra Felső en Hans de Groot (2011: 9; zie ook Van Dooren & Van de Walle, 2008). Dat benchmarkrapporten weinig leiden tot *expliciete* ambtelijke en bestuurlijke bezinning in brieven,

notities en rapporten, en in gesprekstof en discussie daarover in gemeenteraden en in provinciale staten heeft verschillende redenen.

Een eerste reden is dat vaak geen apart rapport per opdrachtgevende overheidsorganisatie verschijnt terwijl dat wel een 'eis' is (Martens e.a., 2002: 39). Het zijn veelal aanbieders van benchmarks die zich aanbieden om een benchmarking uit te voeren en deelnemers aan een benchmark zoeken. Deelname van zeg 15 gemeenten leidt dan tot één rapport, niet tot 15 afzonderlijke rapporten of 15 afzonderlijke rapporten en een overall-rapport.

Het gevolg van gebrekkige of afwezige entreebesprekingen tussen opdrachtgever en opdrachtnemer en een lage want tussen alle deelnemers gedeelde prijs is dat gemeentelijke sectoren die aan bod komen genoeg nemen met één rapport en eigenlijk *zelf* nog een vertaalslag moeten maken in de vorm van een eigen rapportage in termen van 'wat betekent de analyse en wat betekenen de uitkomsten voor ons als gemeente?' Die vertaalslag wordt vaak niet op papier gemaakt. Misschien bestaat het ene rapport wel uit niet veel meer dan een set tabellen en grafieken.

Een tweede reden voor zwak of afwezig gebruik van rapporten en gebrekkige benchlearning is dat opdrachtgevende gemeenten niet of niet intensief genoeg met een opdrachtnemer overleggen voor de start van het onderzoek over de invulling van de opdracht, over de aanpak van de rapportage en de vertaalslag. Anders gesteld, er is nauwelijks sprake van maatwerk. Daardoor ontstaat zekere afstand tot de opdracht en rapportage bij de opdrachtgevers en blijft betrokkenheid bij de uitkomsten en nieuwsgierigheid uit.

Een derde reden is dat colleges van B&W zelden veel weet hebben van de deelname van de gemeente aan een analyse. De deelname is een zaak van een hamerslag onder het motto 'baat het niet dan schaadt het niet'. De deelname is zelden het gevolg van een strategisch besluit om juist aan bezinning te doen en zelden het gevolg van een politieke prioriteit uitgesproken in een coalitieakkoord of geaccentueerd door een aangenomen raadsnotie. Er is rond dit onderwerp benchmarking weinig druk op de ketel. Daarbij moet opgemerkt worden dat er uitzonderingen zijn. De publicatie van kaarten van onderwijsinstellingen door het dagblad Trouw werd vaak nauwlettend door schoolbestuurders en directeuren gevolgd vanuit het idee 'hoe staan wij ervoor?'

Een vierde reden: voor wethouders en burgemeesters is het door middel van benchmarking behalen van een substantieel competitief voordeel op omliggende gemeenten geen argument van enige betekenis. Bovendien, een benchmark afvalbeheer, een sportbenchmark of een benchmark dienstverlening heeft die betekenis (concurrentiekracht) volstrekt niet omdat een lage score zelden echt laag is. Dat blijkt uit door ons gevoerde gesprekken met burgemeesters en wethouders, participatie in enkele collegevergaderingen van Heerlen en Eindhoven en ervaringen als informateur door een van ons. Bestuurders hebben over het algemeen aandacht voor andere prioriteiten en referentiekaders (Tops, Korsten en Schalken, 1994; Korsten en Tops, 1998; Ter Bogt, 2002; Korsten en Aardema, 2006; Abma en Korsten, 2009; Aardema e.a., 2011).

De keer dat een gemeente Schiedam evident slecht scoorde in *'De staat van de gemeente'* en er enige vragen vanuit de gemeenteraad over werden gesteld reageerde B&W van Schiedam schriftelijk dat het onderzoek gebaseerd was op data van enkele jaren terug. Daarmee werd de indruk gewekt dat de situatie nu veel beter was, waarvoor overigens geen bewijs werd aangedragen. Daarmee was ook voor de raadsmeerderheid de kous af.

De vijfde reden is dat een benchmarkrapport wel cijfers geeft maar lang niet altijd normen en daarom moeilijk te interpreteren is. Wat moet je dus aan met een vergelijking? Zitten alle gemeenten in feite op een aspect of onderdeel van gemeentelijk handelen onder de streep of collectief er boven? De rapporteurs van benchmarkrapporten laten de lezers in het ongewisse. Ja, maar vertegenwoordigers van gemeenten of provincies of waterschappen kunnen toch in de laatste fase



vóór het eindrapport verschijnt geconfronteerd worden met de scores en gaan meedenken over de interpretatie? Inderdaad, maar daarvan komt het in de praktijk zelden.

Een zesde reden is dat benchmarkrapporten vaak uitgaan van een brede definitie van benchmarking, als vergelijking en beoordeling. Deze rapporten bevatten dan geen 'best practices', en daarmee dus geen handreikingen voor gemeenten of andere overheidsorganisaties die echt willen verbeteren. In termen van Mark van de Vall, die onderzoek deed naar de benutting van 120 onderzoeksrapporten in drie sectoren: de 'implementaire validiteit' van benchmarkrapporten is niet erg sterk. Direct gesteld: benchmarkrapporten bevatten zelden concrete handelingsaanwijzingen.

Een zevende reden voor gebrekkige benchlearning is dat ambtenaren en/of bestuurders helemaal geen zin hebben in 'leren' of in 'verbeteren'. Een organisatie die bepaalde prestaties niet haalt, kan proberen het roer om te gooien maar kan ook de volgende tegenstrategieën volgen waardoor veranderingen tegengehouden worden: a) aangeven dat de veronderstelling dat prestaties altijd te verhogen zijn, niet opgaat; wat goed is, kan misschien wel niet beter (De Bruijn, 2002); b) de kwaliteit van de indicatoren voor presteren kraken; wie een benchmark rapport de grond in boort, hoeft het natuurlijk niet op te volgen; c) een volgende strategie is zeggen dat de gegevens waarop de benchmark gebaseerd is verouderd zijn; d) aangeven dat de organisatie over andere, recentere cijfers beschikt die wel positief zijn (Van der Zee, 2005: 146).

Het kan niet anders dan dat benchmarking zo gezien niet goed kan beantwoorden aan de ambitie om daadwerkelijk bij te dragen aan een expliciete bezinning in ambtelijke werkgroepen of teams en in bestuurlijke gremia (als het college van B&W en de gemeenteraad) aan verbeteringen in beleid. Toch gaat het benchmarken door.

### **Conclusie**

Hoewel een benchmark een potentieel geschikt instrument is voor zekere bezinning op de kwaliteit van het gemeentelijk handelen zitten in de dagelijkse praktijk weinigen in Nederland, specifiek in gemeenten, echt op een benchmarkrapport te wachten. De urgentie om rapporten te benutten is beperkt. Instrumenteel gebruik van benchmarking is niet afwezig maar komt ook niet op grote schaal voor; we noemen het daarom beperkt (Ter Bogt, 2005). College en politieke fracties roepen er slechts incidenteel om maar doen er weinig mee. De kansen op sturing die men heeft worden niet benut. Maar we dienen te beseffen dat er ook beelden zijn over decentrale overheidsbesturen uit andere landen, zoals Noorwegen (met KOSTRA) en Engeland (met Best Value en CPA/CAA), die tot meer optimisme aanleiding geven (Van Dooren en Van de Walle, 2008: 138; Van Dooren, Bouckaert en Halligan, 2010). Daar moet wel bij worden aangetekend dat het bij beide genoemde landen gaat om een staatsvorm waarbij de hogere overheid via wetten en voorschriften de lagere overheden periodiek om kengetallen kan vragen en ook verbeteringen kan afdwingen. Nader onderzoek is gewenst.

Askim stelt: 'Many scholars have demonstrated that organisations keep measuring even though leaders make *limited use* of the performance information. Many call this a paradox' (Askim, 2008: 138). Blijven vergelijken is blijkbaar toch uitdagend, ook al gooit het bestuur beleid niet om als gevolg van benchmarking. De nadruk op de ambitie van 'deugdelijk bestuur' en 'good governance' in huidige tijden maakt dat deelname aan benchmarking voortgaat. En benchmarking is een vorm van actieve verantwoording naar controleurs en burgers.

## **8 Valkuilen bij benchmarking**

Voor we concluderen dat benchmarking niet brengt wat ervan verwacht wordt (best practices, verklaring daarvan, leren, verbeteren) nog iets over valkuilen.

### **Valkuilen op een rij**

Benchmarkingonderzoek vertoont nogal eens tekorten. Die zijn op te sporen door te bezien of benchmarking wel toegepast worden volgens de eisen die aan dergelijk onderzoek gesteld worden. Maar daarop wordt niet gewacht. Vanuit de praktijk kunnen we al waarnemen waarom benchmarking niet steeds cum laude slaagt of überhaupt slaagt. Daarom de vraag: wat zijn nu valkuilen en meest voorkomende fouten (pitfalls) bij benchmarking? De volgende bleken te traceren.

1. De valkuil om een groot benchmarksysteem te laten ontstaan waarin organisaties participeren met prestatie-indicatoren voor alle taakvelden.
2. De valkuil om alles te willen reduceren tot cijfers, of de beperktheid van het kwantitatieve.
3. De valkuil van beperkte relevantie van indicatoren.
4. De valkuil van de gebrekkige afdekking van het meetobject.
5. De valkuil van de afwezige beleidsdoelstellingen.
6. De valkuil van de ongenormeerde vergelijking.
7. De valkuil van de verbeteringsverwachting.
8. De valkuil van het abusievelijk uitvergroten van kleine verschillen.
9. De valkuil van de onbegrijpelijkheid.
10. De valkuil van de uitblijvende tweetrapsraket.
11. De valkuil van niet geactiveerd gezond verstand.

Deze valkuilen worden toegelicht. Deze valkuilen moeten bij voorkeur vermeden worden want het zijn negatieve verschijnselen. Daarom zijn aan de benoemde valkuilen ook suggesties verbonden. Worden deze niet vermeden dan zullen benchmarkrapporten achter blijven bij de verwachtingen.

### **1 Een groot interbestuurlijk systeem met benchmarking als kern is disfunctioneel**

Benchmarking is doorgaans een vorm van op besluitvorming gericht (praktijk)onderzoek. Meestal is bij benchmarking sprake van een vergelijking tussen 5-20 organisaties. Na 2000 is bij het ministerie van BZK echter, in de lijn van het streven naar meer verantwoording (de VBTB-operatie) overwogen om alle ruim vierhonderd gemeenten op te nemen in een omvattend interbestuurlijk benchmarkingsysteem, waarbij het totale gebied aan gemeentelijke opgaven, taken en uit medebewind voortvloeiende of daarmee verbonden ambities door middel van indicatoren zou worden geïndiceerd op kwaliteit. Dat zou een systeem zijn waarin honderden kernvariabelen zouden moeten worden opgenomen. De hoop was dat het Centraal Bureau voor Statistiek een aantal cijfers zou kunnen leveren. Maar die poging is nooit van de grond gekomen en ook onwenselijk vanwege allerlei complicaties (Rob, 2002). Een belangrijke reden is dat als er overeenkomsten en verschillen tussen gemeenten zijn op het vlak van wegonderhoud, dienstverlening of huisvuilophaal van ruim vierhonderd gemeenten bekend kunnen zijn, men nog niet weet wat daarvan de echte verklaringen, redenen of achtergronden zijn. Wat achter de scores schuil gaat, is niet bekend terwijl dat wel van belang is om te weten, om de scores en scoreverschillen te kunnen duiden. Een benchmark die niet gepaard gaat met een verklarende analyse van verschillen is niet erg zinvol en misschien zelfs zinloos (Blank e.a., 2011: 5; Van Dooren, Bouckaert & Halligan, 2010).

Dat een gemeente in absolute zin veel budget kwijt is aan wegonderhoud, zegt weinig als niet bekend is hoeveel kilometers weg een gemeente moet onderhouden. Het totale budget voor wegonderhoud zegt dus weinig. Een beter vergelijkingsgetal is het budget voor wegonderhoud per kilometer weg die voor gemeentelijk onderhoud in aanmerking komt. Maar ook dat is niet genoeg want er kan sprake zijn van een jaarlijks verschil in investering in wegonderhoud. Je moet dus ook over een cijfermatig meerjarenbeeld beschikken. Taakuitvoering en kenmerken van beleid, zoals in relatie tot bij voorbeeld de Wet maatschappelijke ondersteuning (Wmo), laten zich ook niet eenvoudig in een beperkte set indicatoren vangen. Wegonderhoud en de Wmo zijn dan nog maar enkele taakgebieden voor gemeenten, in werkelijkheid gaat het om een groot taakgebied (Bordewijk en Klaassen, 2000).

Dit verschijnsel - dat het gehele, omvattende beleid van in een benchmarksysteem betrokken gemeenten zich moeilijk kwantitatief, in cijfers, laat 'betrappen' en verklaren - heeft ook het gebruik van kengetallen in gemeentelijke begrotingen in de periode 1995-2010 geremd (Bordewijk en Klaassen, 2000). Zou een meer eenvoudige benchmarking wel mogelijk zijn? Als de valkuil van 'het grootse' voorkomen wordt. De Rob vindt benchmarking zinvol en relevant als een kleiner aantal deelnemers deelneemt en verschillen tussen scores ook verklaard kunnen worden: geen interpretatiekwaliteit zonder zicht op achtergronden. Wie een interbestuurlijk werkend benchmarksysteem wenst op te zetten dat ook verklarende variabelen bevat, zal te maken krijgen met een sterk uitdijend systeem; het aantal variabelen zal zeer sterk toenemen. Dat brengt ons bij de tweede valkuil.

## **2 De beperktheid van het kwantitatieve**

Wat is de aard van de cijfers die in benchmarkonderzoek feitelijk aan de orde komen? Meetsystemen zijn doorgaans slechts kwantitatief gericht. Daarmee is een beperking van benchmarking een feit. 'Men zal zelden de hele realiteit op een kwantitatieve wijze kunnen vatten' (Bouckaert e.a., 2011: 121).

Veel benchmarkrapporten kennen kwantitatief-beschrijvende scores maar *geen* verklarend deel en dat is zeer betreurenswaardig. Het waarom van een positie in de betere helft van vergeleken gemeenten op het vlak van rioleringszorg of e-government achterhaal je meestal in een benchmarkrapportage niet. Evenmin in een rapportage over andere organisaties, zoals waterschappen (Van Helden e.a., 2005; Ter Bogt e.a., 2010). Terwijl verklaringen wel van groot belang zijn om tot verbeteracties te *kunnen* komen (Blank e.a., 2011: 5). Men moet weten of een doelmatigheidsverschil in afvalbeheer veroorzaakt wordt door de stedelijkheid of door de organisatiewijze van het afvalbeheer, of door beide en in welke mate. De valkuil van het *alleen* voorzien in hitlijsten, *zonder* zich te bezinnen op oorzaken van hoge of lage scores, moet dus voorkomen worden (cf Bouckaert e.a., 2011).

## **3 De relevantie van indicatoren is afwezig**

Benchmarking impliceert vergelijking van eenheden in de vorm van indicatoren, die zijn te beschouwen als standaarden of gemeenschappelijke maatstaven. Deze indicatoren zouden min of meer relevante, accurate, volledige en actuele weergaven moeten geven van of over een verschijnsel dat object van benchmarking is (Callahan, 2007).

Dat betekent dat als een benchmarking plaatsvindt van rioleringskwaliteit een indicator ook relevant moet zijn en iets wezenlijks moet zeggen over de *essentiële* kwaliteit van rioleringen: het systeembereik van rioleringen (welk deel van een gebied wordt wel of niet bereikt?); de opvang- en verwerkingscapaciteit per uur; de kwaliteit van de rioolbuizen in termen van dichtheid, slijtage en levensduur; de capaciteit van riolen op piekmomenten, zoals stortbuizen of bij andere wateroverlast; de kwaliteit van riolering in bepaalde bijzondere gebieden, zoals op bedrijventerreinen.

In de benchmarkpraktijk komt het voor dat de verantwoording van de variabelenkeuze of indicatorenkeuze in een rapport *achterwege* blijft (Bouckaert e.a., 2011). Het gevolg kan zijn dat een organisatie op bepaalde variabelen positief scoort maar dat de organisatie zich in de luren laat leggen omdat dit niet de *essentiële* en *relevante* indicatoren zijn maar *marginale* indicatoren zijn. Het is zinloos met een criminaliteitsbenchmark van steden te komen waarbij de aantallen tasjesberovingen de enige indicator is voor de omvang van criminaliteit.

## **4 Validiteitsprobleem: de valkuil van de gebrekkige afdekking**

Bij het ontwikkelen van een meetsysteem in het algemeen en dus ook van benchmarking moet aandacht worden besteed aan de eis van geldigheid of validiteit. Een onderzoeker moet de goede

dingen meten. Dat wil zeggen dat een meetobject, zoals rioleringszorg of afvalbeheer voldoende afgedekt moet zijn als deze object van benchmarking zijn. Lukt dat steeds? Wie de kwaliteit van een ziekenhuis beoordeeld begint niet als eerste over de kwaliteit van de keuken maar over iets anders, zoals de kwaliteit van de chirurgen of succesvolle operaties. Een valkuil bij benchmarking blijkt in de praktijk de *overschatting* van de afdeckingsgraad van een indicator (Bouckaert e.a., 2011: 123).

Dat betekent dat soms ten onrechte verondersteld wordt dat een indicator het hele (bredere) meetobject afdekt. Zo is het onvoldoende om diefstal van handtassen als indicator voor stedelijke criminaliteit te nemen en daarmee te volstaan. Want dan ontsnappen andere vormen van criminaliteit aan de meting, zoals onder meer scooter- en autodiefstal, proletarisch winkelen en winkelinbraken. Handtasdiefstal past wel bij straatroof en straatroof is dan weer een onderdeel van het bredere fenomeen 'stedelijke criminaliteit'. Wat is dus een indicator van wat? Wie veiligheid als object van een benchmark neemt moet aandacht besteden aan fysieke en sociale veiligheid. De onderzoeker stuit op gebrek aan gegevens maar is verplicht op de een of andere manier ook moeilijk meetbare aspecten in beeld te brengen. Een reductie van het meetobject tot de indicator zelf in geval het verwerven van meer gegevens teveel werk wordt geacht, is een ongewenste handigheid.

Soms wordt een indicatorset overigens gestructureerd in een model, zoals een 'balanced scorecard' (Martens e.a., 2002: 72; Bouckaert c.s., 2011: 128). Ervaringen met de toepassing hiervan wijzen uit dat het onmogelijk is de balanced scorecard-aanpak in gemeenten gemeentebreed toe te passen. De balanced scorecardmethode is voornamelijk geschikt voor sterk productiegerichte organisaties of organisatiedelen, die eenvoudige producten afzetten en niet geschikt voor gebruik van alle sturings- en verantwoordingsvraagstukken.

### **5 De doelstellingen van overheidsbesturen zijn onbekend**

Een andere valkuil is dat het hanteren van eenzelfde indicator voor een object dat onderwerp is van benchmarking, zoals rioleringen, veronderstelt dat de doelstellingen van beleidsmakers (bestuurders) van zeg verschillende gemeenten gelijk zijn. In werkelijkheid kunnen de beleidsdoelstellingen echter verschillen.

Stel de ene gemeente heeft riolering gelegd met een levensduur van 40 jaar en een andere met een duur van 60 jaar. Misschien is op het moment van meting de periode van 40 jaar in een gemeente met een stelsel met de levensduur van 40 jaar al bijna verstreken en in de andere niet. Dan is een lagere score bij de gemeente met de riolering die gold voor 40 jaar meer begrijpelijk dan als de andere gemeente met een zelfde slijtage en andere kwaliteitsaspecten voor de dag komt. Zelfde cijfers maar er gaan andere doelstellingen en andere producten achter schuil.

Dat lees je lang niet altijd in verslagen van benchmarks. Het gevolg hiervan is dat je niet zomaar kunt spreken van een reële vergelijking omdat in één gemeente in dit voorbeeld de riolering is afgeschreven. Advies aan een onderzoeker: kijk dus als onderzoeker altijd naar wat er schuil gaat achter een onderwerp of object van benchmarking. Advies aan een opdrachtgever: dwing de onderzoeker om de doelstellingen achter beleid en productkenmerken (levensduur, kwaliteit) mee te nemen in de interpretatie van overeenkomsten en verschillen tussen scores op indicatoren.

### **6 De valkuil van de ongenormeerde vergelijking**

Dicht tegen de genoemde valkuil van ontbrekende doelstellingen aan zit een andere valkuil. Het betreft de valkuil van de ongenormeerde vergelijking, van een vergelijking *zonder* streefcijfers. De redenering luidt hiervoor als volgt. Er vindt in een benchmark een vergelijking van scores plaats. Stel op het vlak van de doelmatigheid van afvalbeheer zit de gemeente X in de middenmoot van pakweg 40 andere gemeenten. Omdat er dus gemeenten zijn die qua doelmatigheid beter zijn wordt, zonder dat sprake is van normering over wat qua doelmatigheid in afvalbeheer wenselijk of haalbaar is, gesteld dat er nog doelmatigheidsvoordeel te behalen is. Bij *gebrek aan normering* geldt de beste organisatie als referentiepunt. Dan is dus geen sprake van normering op basis van 'inhoud' of

politieke sturing van wat nastrevenswaard is bij de organisatie van afvalbeheer, maar wordt het politiek bestuur door de benchmark op een idee gebracht: streef naar meer doelmatigheid want andere gemeenten bereikten al meer doelmatigheid. Wat hier miskend wordt, is dat misschien alle deelnemers aan de benchmark al op een hoog doelmatigheidsniveau zitten. En wat buiten beschouwing blijft is, of het voor een gemeente wel mogelijk is om de organisatie te wijzigen. Misschien heeft een gemeentebestuur wel een contract met een looptijd van 30 jaar gesloten dat niet makkelijk te wijzigen is dan tegen vergoeding van een flinke compensatie voor het openbreken van het contract.

### **7 De valkuil van de verbeteringsverwachting**

Benchmarking is gericht op verbeteren maar geeft de opdrachtnemende onderzoeker ook zicht daarop? Deelnemers aan een benchmark mogen verwachten dat een benchmarkrapport aangeeft of een toekomstige verbetering in scores voor hun organisatie mogelijk is en in welke richting die zou moeten gaan. Maar in werkelijkheid leest men hierover misschien niets omdat geen streefcijfers of een andere normering en of 'best practices' vermeld worden. De lezer van een benchmarkrapport heeft *geen idee of het totaal aan scores onder een kritische grens zit of juist er boven*. Het politiek bestuur van zeg een deelnemende gemeente is in zo'n geval dus onthand.

We zien hier dat het belangrijk is voor politieke sturing die kan volgen op benchmarking, dat sprake is van normering in de vorm van streefcijfers. Als er sprake is van zowel streefcijfers als feitelijke cijfers valt aan te geven of er een kloof is tussen beide en als ook blijkt dat er sprake is van substantieel betere organisaties en 'best practices' is een (cijfermatige) aanzet voor verbetering aanwezig. Dat brengt ons bij een volgende valkuil: is er werkelijk veel verschil tussen de scores?

### **8 De valkuil van het abusievelijk uitvergotten van kleine verschillen**

Een specifieke valkuil betreft de verschillen tussen scores. 'Benchmarking can be a powerful tool and useful for identifying red flags, but caution should be taken when comparing small differences' (Callahan, 2007: 65). Het is mogelijk op grond van vergelijkingen te komen tot een waarschuwing van een deelnemer die het er echt bij liet zitten als gevolg van slechte scores. Maar het is ook mogelijk dat een rangorde van scores ontstaat waarbij de verschillen klein zijn en alle deelnemende benchmarkpartijen, bijvoorbeeld gemeenten, voldoende of meer scores. De valkuil is dan dat toch het beeld kan opdoemen dat de hoogste in de rangorde *ver superieur* is boven de laagste in rangorde en alleen de hoogste de echte excellente of goed praktijk representeert terwijl dit niet juist is. Klaas Abma en Arno Korsten (2009) stotten op dit verschijnsel toen ze deelnemers aan *'De staat van gemeenten'* op zes burgerrollen vergeleken. De scores lagen op een aantal rollen dicht tegen elkaar aan en waren 'voldoende' of hoger. De suggestie dat er grote kwalitatieve beleidsprongen voorwaarts te maken waren door gemeenten na de benchmark ging voor een groot aantal deelnemers dus niet op. Advies: bekijk als onderzoeker goed de grootte van de verschillen in scores.

### **9 De valkuil van de onbegrijpelijkheid**

In de literatuur wordt als eis gesteld dat benchmarkrapporten met scores begrijpelijk moeten zijn (Callahan, 2007; Bouckaert e.a., 2011: 122, 128). Deze eis van begrijpelijkheid impliceert dat aan lezers en andere gebruikers van rapporten duidelijk moet worden gemaakt welke streefcijfers of andere normen aan de orde zijn, waarom welke indicatoren per meetobject gekozen zijn, wat deze afdekken, en wat verschil tussen organisaties of in de tijd veroorzaakt. In de praktijk van gemeentelijke benchmarking wordt duidelijk dat aan deze eis niet steeds wordt voldaan.

De analyse van de kostenstructuur van politiekorpsen zette onderzoekers voor raadsels. In een periode van vier jaar bleek de gemiddelde efficiëntie van politiekorpsen licht afgenomen. Waarom? En het politiekorps Rotterdam-Rijnmond was in een bepaalde jaar een van de minst efficiënte korpsen

terwijl het drie jaar later een van de meest efficiënte was. Waarom? De oorzaak in het beleid zoeken was onjuist. De onderzoekers kregen er nog geen greep op (Cleuver en Blank, 2006: 63).

Aan de begrijpelijkheid wordt soms wel wat gedaan, via additionele bijeenkomsten na het verschijnen van een benchmark. De vraag is of dat niet aan de late kant is.

Neem de benchmark ict-kosten die Berenschot uitvoert. Vraagt een gemeentebestuur zich af of de ict-activiteiten aan de dure kant zijn? En of het goedkoper kan zonder in te leveren op kwaliteit? Is een gemeente op zoek naar relevant vergelijkingsmateriaal, dan is de benchmark over ict-kosten een uitkomst. De Benchmark ict-kosten geeft inzicht in de kosten van beheer en onderhoud van werkplekken, applicaties en netwerkverbindingen, en in taken in de informatievoorziening en hun plaats in de organisatie. De onderzoekers willen bij gebleken behoefte in workshops aangeven wat de benchmarkuitkomsten betekenen in de context van uw organisatie.

### **10 De valkuil van de uitblijvende tweetrapsraket**

'Measurement by itself will not improve performance', zegt Callahan (2007: 70). Uitkomsten uit een benchmarkonderzoek kunnen niet goed gebruikt worden

- als de indicatoren te beperkt zijn voor wat men wil waarnemen (geldigheidsvraagstuk),
- als geen cijfers over meerdere jaren beschikbaar zijn (trend of wisselingen in scores niet zichtbaar), en
- als hoge en lage scores niet verklaard worden (scores beïnvloedbaar door bestuur?),
- als de analyse en rapportage daarover niet goed te begrijpen is.

Zoveel is intussen wel duidelijk. Omdat elke organisatie uniek is, moeten bevindingen uit een benchmarkonderzoek beschouwd worden als een indicatie waarop een *nadere analyse* van de specifieke situatie moet volgen. Want een benchmark kan nooit het enige onderzoek zijn waarop grote of kleinere beleidsverandering worden gestoeld, zeggen we Martens e.a. (2002: 39) na. Een benchmarkanalyse is in feite de eerste trap van een wenselijke tweetrapsraket. Anders gesteld, er moet altijd een verbinding gelegd worden tussen vergelijking van organisaties en de positiebepaling van de eigen organisatie. Het gaat om antwoord op de volgende vragen: 'waar staan wij, waarom staan we waar we staan en wat betekenen de bevindingen voor ons?').

### **11 De valkuil van niet geactiveerd gezond verstand**

Een benchmark bevat vooral gegevens over het verleden van organisaties. Maar een bestuurder of manager is toekomstgericht, althans hoort dit te zijn. De valkuil kan hierin bestaan dat een bestuurder veel hoop vestigt op een verbetering van de benchmarkonderzoeken, om zo een plausible stuurrichting in beeld te krijgen. Echter, een visie is voor een bestuurder of manager misschien wel belangrijker dan door middel van benchmarkonderzoek te weten wat in het verleden is gebeurd (Van der Ende, 1991: 37). Het is ook denkbaar dat deze leidinggevendenden verder komen door het gezond verstand te gebruiken of een team samen te stellen dat een toekomstig beleid mee uitstippelt.

1. De valkuil om een groot benchmarksysteem te laten ontstaan waarin organisaties participeren met prestatie-indicatoren voor alle taakvelden.
2. De valkuil om alles te willen reduceren tot cijfers, of de beperktheid van het kwantitatieve.
3. De valkuil van beperkte relevantie van indicatoren.
4. De valkuil van de gebrekkige afdekking van het meetobject.
5. De valkuil van de afwezige beleidsdoelstellingen.
6. De valkuil van de ongenormeerde vergelijking.
7. De valkuil van de verbeteringsverwachting.
8. De valkuil van het abusievelijk uitvergooten van kleine verschillen.
9. De valkuil van de onbegrijpelijkheid.
10. De valkuil van de uitblijvende tweetrapsraket.

### 11. De valkuil van niet geactiveerd gezond verstand.

Wat nu van de genoemde valkuilen te denken? Een aantal valkuilen is serieus te nemen, zoals de beperktheid van het kwantitatieve, de ongenormeerde vergelijking, de gebrekkige indicering van een te meten object of concept, het uitvergroten van kleine verschillen in rangorde, het fenomeen van het ontbrekende tweede deel van de tweetrapsraket, en de valkuil van de onbegrijpelijke rapportage. Ze ondergraven de verwachtingen rond benchmarking (vergelijken, rangordenen, best practices aanwijzen, echt verbeteren).

De valkuil van het niet geactiveerd gezond verstand treedt in de werkelijkheid van het openbaar bestuur in relatie tot benchmarking maar beperkt op. Er gaat achter deze valkuil de veronderstelling schuil dat benchmarking zo dominant en zo gewichtig is dat bestuurders in het verleden blijven zitten. Dat impliceert een enorme onderschatting van bestuurders. Zo dom zijn de meesten niet.

### **Methodische valkuilen op een rij**

Er bestaan ook specifieke methodische valkuilen. Enkele zijn er al genoemd. Welke zijn toe te voegen? Bouckaert en Balk (1991), Bouckaert (1995: 388), Bouckaert en Van Dooren (2004) en Van Dooren, Bouckaert & Halligan (2010) noemen er enige. Ze duiden die aan als *pathologieën*. Wij spreken van valkuilen. Het gaat om valkuilen die leiden tot *valse klanken* in de uitvoering van de benchmarkmuziek ('measurement noise'). Het is gewenst de valkuilen te vermijden opdat een benchmark ons niet op het verkeerde been zet. De literatuur spreekt dan ook over '*diseases*' en '*cures*'. Het gaat om de volgende valkuilen:

1. De valkuil van Pangloss.
2. De valkuil van het onmogelijke.
3. De valkuil van inferioriteit of zwaarmoedigheid.
4. De 'valkuil' van de opgeblazen ballon.
5. De valkuil van de verwaarlozing van input.
6. De atrofieziekte.
7. De valkuil van gepercipieerde toename.
8. De valkuil van vervuild cijfermateriaal.
9. De valkuil van de te korte horizon.

We lichten deze toe.

### **1 De valkuil van Pangloss**

Deze ziekte of valkuil ('Pangloss disease') bij benchmarkonderzoek houdt in dat een onderzoeker iets gaat meten wat er helemaal niet kan zijn. Dit beeld komt van Dr. Pangloss uit Voltaire's *Candide* (Bouckaert, 1995). Deze valkuil is duidelijk te maken aan de hand van een voorbeeld. Het is onmogelijk bruggenbouw via benchmarking op bouwefficiëntie te beoordelen als er maar een brug in de wereld bestaat. Immers, als er maar een brug bestaat, valt er niet te vergelijken, laat staan een 'best practice' op te sporen. Realistisch? Er is inderdaad maar één opera van Sydney.

### **2 De valkuil van het onmogelijke**

Deze valkuil ('the impossibility disease') lijkt op de vorige maar verwijst niet naar het unieke maar naar het maximale van iets. Het is onmogelijk om een situatie te onderzoeken waarbij sprake is van lucht en geen lucht. Lucht is immers een collectief goed; geen mens is ervan uitgesloten; lucht is voor iedereen beschikbaar; lucht is overal. Er zijn geen situaties denkbaar om mensen lucht te onthouden. Anders gesteld: er is altijd lucht. Let op zo'n valkuil bij benchmarking. Overigens is er wel te onderscheiden naar zuiverder en minder zuivere lucht. In zoverre is het voorbeeld niet helemaal geschikt. Ander voorbeeld: wie in het kader van rioleringszorg in Nederlandse gemeenten ten aanzien van woningen meet of sprake is van rioleringsbuizen, meet een open deur; voor zover bekend bestaat riolering bij woningen zonder rioleringsbuizen niet. In een

benchmark van scholen vragen naar de beschikbaarheid van schoolgebouwen is dan ook zinledig. Elke school heeft altijd wel minstens een schoolgebouw.

### **3 De valkuil van inferioriteit of zwaarmoedigheid**

Het is fout om de publieke sector op voorhand als inferieur of als superieur af te doen ten opzichte van de private sector, de markt en marktwerking (Bouckaert, 1995: 393). De publieke sector heeft eigen kenmerken. Een voorbeeld. Een brandweer een brand laten blussen, is bijvoorbeeld niet alleen te beoordelen op de snelheid van het spuiten met water of het beëindigen van de brand. Soms moet een brand 'begeleid' worden omdat deze goed moet uitbranden. Een brandweergroep moet immers ook altijd letten op de gevolgen van het brandweerwerk, zoals: is de kans groot dat een brand weer opblaait bij te snel het bluswerk willen afronden?; is er een kans dat de veiligheid op een industrieterrein in gevaar komt als het doven van de brand niet begeleid wordt? Een benchmarkonderzoeker moet dus wel weten wat vergeleken wordt en welke indicatoren gebruikt worden (Bouckaert, 1995: 394).

### **4 De 'valkuil' van de opgeblazen ballon**

Hierbij gaat het om het gevaar om output zo te meten dat op voorhand duidelijk is dat het resultaatniveau in werkelijkheid lager is. Bij voorbeeld dat men zegt dat de recherche zeer succesvol is in het oplossen van misdrijven terwijl tegelijk ruimschoots bekend is dat tal van misdrijven of overtredingen niet eens bij de politie worden gemeld en geregistreerd en dat sommige wel gemelde misdrijven niet worden onderzocht. Als misdrijven worden opgelost, gaat het slechts om wat gemeld is en daarvan weer de subcategorie 'waarnaar een onderzoek is gestart'. Dat is in absolute zin slechts een minderheid van gevallen. Het oplossingspercentage kan dus sterk vertekenen. Een benchmarkonderzoeker moet *niet* in deze valkuil lopen omdat dit leidt tot het verwijt van onbewust of bewust bedrog.

### **5 De valkuil van de verwaarlozing van input**

Bij deze valkuil gaat het erom dat een analyse van resultaten zonder rekening te houden met kosten verkeerd is. Het is namelijk mogelijk dat een concentratie op de omvang van productie of resultaten leidt tot veel hoge(re) inputkosten zonder dan je dit in de gaten hebt. Let dus altijd op de verhouding output/input. Anders gesteld, let er op of het mogelijk is met een gelijke input aan middelen meer resultaat te boeken.

### **6 De atrofie-ziekte**

Atrofie verwijst naar gebrek aan voeding. Deze valkuil houdt in dat een onderzoeker niet ziet dat een toename van input niet leidt tot meer resultaat. Meer middelen kan zelfs een contraproductief effect hebben, vergelijkbaar met de overtraineerde schaatser: harder getraind dan ooit in de hoop op excellent presteren maar de werkelijke resultaten gaan niet naar het oude niveau maar liggen daaronder. Een benchmarkonderzoeker moet hier oog voor hebben. Een lagere score van een partnerorganisatie dan de gemiddelde score kan het gevolg zijn van heel goede bedoelingen en inspanningen. Deze score hoeft niet altijd het gevolg te zijn van slecht beleid maar kan de consequentie zijn van 'het te goed willen doen' (Bouckaert, 1995: 397). De onderzoeker ontkomt er dus niet aan om op zoek te gaan naar het verhaal achter de scores. Een gezamenlijke bespreking met representanten van de partnerorganisaties van een concept-benchmarkanalyse kan soms zo een verhaal naar boven halen.

### **7 De valkuil van gepercipieerde toename**

Deze 'disease' betreft: 'more detailed yardsticks cause an increase of perceived amounts', aldus de Leuvense hoogleraar Geert Bouckaert (1995: 398). Deze valkuil houdt in dat de indruk kan ontstaan dat als je een meer *gedetailleerd* vergelijkingsmaatstaf hanteert er ook een *toename* van positieve resultaten ontstaat. In dat geval bestaat het gevaar dat je jezelf voor de gek houdt. Het hoeft niet maar er kan sprake zijn van een fopperceptie. Wie als mensenrechtencentrum het aantal martelingen in een dictatoriaal bestuurd land telt, moet misschien niet van positieve progressie in



het voldoen aan mensenrechten spreken in geval het aantal gemartelde burgers van enkele honderden naar 215 precies getelden loopt, enkele jaren later. Misschien was eerder ook al sprake van om en nabij 215 gemartelden. Advies: de onderzoeker moet om deze kuil heen lopen.

### **8 De valkuil van vervuild cijfermateriaal**

In een organisatie is een onderscheid te maken in input (invoer), throughput (doorvoer/ verwerking), output (resultaat) en outcome (effect). Wie als onderzoeker output relateert aan input krijgt zicht op de productiviteit en doelmatigheid. Maar wie in een benchmark werkt met een mix aan gegevens op het vlak van input, activiteiten, resultaten in termen van kwantiteiten of effecten verontreinigt de vergelijking. Dan is sprake van wat men 'polluting measures' noemt (Bouckaert, 1995: 399). Het gevolg is verwarring. Wie iets weet van de omvang van de input, weet nog niets van de omvang aan resultaten (outputs). En wie zicht heeft op prestaties of resultaten (outputs) weet nog niet of ook gewenste effecten (outcomes) bereikt worden. En wie het bereiken van resultaten en effecten gemeten heeft, weet nog niet of de burgers dit ook waarnemen of zo beleven. Een benchmarkonderzoeker moet dus goed nadenken over het opsporen en gebruiken van *significante* informatie. Bij het ontwerpen van de benchmark moet daar op gelet worden. De indruk bestaat dat deze valkuil in gemeenten zeker voorkomt (Haselbekke, 1998).

### **9 De valkuil van de te korte horizon**

Een benchmarkonderzoeker moet geen gegevens gebruiken die te snel na het in gang zetten van beleid zijn verzameld. Wie wil weten of de verkeersveiligheid vergroot is, moet niet het aantal verkeerslachtoffers nagaan twee dagen na de plaatsing van borden om een meer verkeersluwe situatie op zeg 30 punten te verkrijgen en twee dagen na het aanbrengen van andere verkeersnelheidsremmers. Het is gewenst later te meten en dit ook meerdere malen, met tussenpozen, achter elkaar te doen. De druk om daarentegen snel te meten, is doorgaans groot omdat begrotingen een jaarcyclus kennen en politici en bestuurders in een beperkte periode tussen verkiezingen actief zijn en 'met resultaat thuis willen komen'. Deze druk moet echter weerstaan worden, omdat valse succesnummers op den duur geen 'evergreen' zullen blijken te zijn. Advies: kies voor meerjarencijfers. Ook om deze valkuil wordt niet steeds heen gelopen.

### **Valkuilen vermijden in ontwerpprocessen**

Aandacht voor deze potentiële valkuilen bij benchmarking en bij andere manieren van prestatiemeting, is nadrukkelijk gewenst, meent een expert als de Leuvense hoogleraar Geert Bouckaert. Hij stelt:

'Reducing the occurrence of measurement diseases should be a major concern in the public productivity measurement design process' (Bouckaert, 1995: 406).

Het vermijden van de genoemde valkuilen is nodig om te komen tot *betekenisvol meten*, tot wat Bouckaert (1995: 405) noemt 'meaningful measurement' in de praktijk van prestatiemeting.

Een logische vervolgvraag die valt te stellen, luidt dan: Is de discussie over de genoemde en andere methodische valkuilen nu iets dat *in de praktijk* van benchmarkonderzoek ook bediscussieerd en vermeden wordt of speelt dit in de praktijk van concrete benchmarking niet of nauwelijks? Dergelijke raadgevingen zijn zeker besteed aan 'eredivisie'-onderzoekers van de OECD, het Sociaal en Cultureel Planbureau (SCP) en het Centraal Planbureau (CPB). Onderzoekers en auteurs als Jos Blank hebben hiervoor oog. Aandacht voor deze methodische valkuilen hebben we 'in de tweede divisie van benchmarkpraktijken' veel minder aangetroffen. Benchmarkrapporten maken doorgaans geen melding van de noodzaak van het vermijden hiervan of getroffen maatregelen. Bestuurlijke gremia (gemeentebesturen, colleges van bestuur van een roc of riagg) die opdracht geven tot benchmarking of formeel hiervoor verantwoordelijk zijn, bemoeien zich niet echt hiermee. Leden van bestuurlijke gremia of van raden van toezicht die benchmarks af en toe onder ogen krijgen, bemoeien zich ook niet met methodische valkuilen. Dat is begrijpelijk omdat dergelijke methodische valkuilen vermijden ver ligt van hun taakopdracht en hun expertise (Haselbekke, 1998; Ananias, 2007; Lapperre, 2007;

Goodijk, 2012; Douma, Stevens en In 't Veld, 2012). De stelling luidt daarom dat ten aanzien van methodische valkuilen bij de opzet van benchmarking voor gemeentelijke (partner)organisaties sprake is van drie gescheiden werelden: de wereld van de methodisch specialisten (die valkuilen aanwijzen), de wereld van uitvoerders van benchmarkonderzoek (die valkuilen zouden moeten vermijden) en de wereld van beslissers over en gebruikers van benchmarkonderzoek (die valkuilen niet kennen maar wel kunnen aandringen op kwaliteitsrijke benchmarking, ook in methodisch opzicht).

## 9 Enkele conclusies

Benchmarking is een potentieel interessant instrument voor verantwoording en leren maar beantwoordt niet aan de verwachtingen. Deze conclusie geldt voor veel benchmarkonderzoek en zeker ook voor benchmarking in de context van het lokaal bestuur. Tevens stellen we vast dat verder onderzoek nodig is. Dit wordt toegelicht.

Benchmarking is een instrument om via hoofdzakelijk kwantitatieve vergelijking tussen organisaties (of delen daarvan of productclusters) en rangordening van scores, te zoeken naar beste of goede praktijken en op basis daarvan te komen tot verbeteringen. Benchmarking richt zich op prestaties en daarmee is het een onderzoeksaanpak die past binnen prestatie management. Het instrument heeft voordelen en nadelen. Tot de voordelen behoort de gerichtheid op verbetering en verantwoording. Benchmarking is ook een basis om met elkaar in gesprek te gaan, dus een signaleringsinstrument voor het opsporen van zwakke plekken en sterke plekken in processen in en producties van organisaties. Tot de nadelen rekenen we dat dit instrument, zeker als het gaat om gemeentelijke benchmarking, meer de burgers als klant in beeld brengt dan burgers als participant, partner in samenwerking, onderdaan of wijkbewoner (cf Hiemstra, 2003: 78; uitzondering in Abma en Korsten, 2009).

Benchmarking brengt over het algemeen niet netwerkproductiviteit in beeld terwijl gemeenten voor tal van taken en opgaven afhankelijk zijn van mede-overheden en andere actoren (De Rynck, Voets en Van Dooren, 2010). Daarom kan de volgende stelling in een rapport als *'Rotterdam vergelijkenderwijs'* (2002: 2) niet goed onderbouwd worden. De stelling luidt: 'Alle processen en functies van een organisatie kunnen in beginsel in aanmerking komen voor een benchmarkonderzoek'. Die stelling miskent dat ook Rotterdam functioneert in netwerken als de veiligheidsregio. Prestaties van netwerken zijn wel te benchmarken, maar daar moet dan speciale aandacht aan worden besteed. Verder, we kennen geen benchmarks waarin werkelijk alle gemeentelijke opgaven en taken worden meegenomen. Logisch, niet alles is kwantificeerbaar.

Benchmarking is bovendien vaak gericht op de kwaliteit van machinebureaucratieën, dus op grote productiegerichte organisaties met min of meer gestandaardiseerde operationele processen in eenvoudige en betrekkelijk stabiele omgevingen. Of op wat daarbij in de buurt komt. Denk aan processen waarbij aantallen een rol spelen, zoals bij afvalbeheer. Benchmarking is als manier van prestatie meting *minder* geschikt voor het opsporen en beoordelen van prestaties van cocreatie, van 'ad hoc' organisaties, 'high performance organizations', politieke partijen of kerkelijke organisaties. Immers, die prestaties zijn vaak uniek en/of er is sprake van geheimhouding of de prestatie is niet te kwantificeren of zo een organisatie wil niet beoordeeld worden in termen van producties en prestaties.

Benchmarkonderzoek is gericht op verantwoording naar bestuur en samenleving maar in werkelijkheid worden benchmarks zelden door burgers geraadpleegd. Ze worden doorgaans ook niet actief richting burgers op transport gezet.

Tal van aspecten van benchmarking zijn behandeld, maar welke bevindingen levert onze analyse op over de mythen rond benchmarking? Dat was het perspectief dat in deze beschouwing centraal stond. Benchmarking maakt over het algemeen de verwachtingen niet helemaal waar. De keizer heeft minder kleren aan dan gewenst en verwacht.

### **Eerste mythe**

De eerste mythe luidde: Organisaties zoeken zelf ten behoeve van benchmarking naar partnerorganisaties, waarmee een vergelijking mogelijk is. Vanuit de benchmarkpraktijk zoals deze bij waterschappen en gemeenten voorkomt, blijkt echter dat benchmarking door externe aanbieders wordt aangeboden; van het zelf zoeken van partnerorganisaties waarmee een vergelijking mogelijk is, komt daarom meestal weinig of niets terecht. Soms zijn overigens alle organisaties in een benchmark betrokken, zoals voor roc's geldt; dan valt er ook niet uit te zoeken. Maar dat is een uitzondering. De belangrijkste reden voor de constatering van een afwijking van de mythe is een andere. Benchmarking is sterk *aanbodgedreven*. Waterschappen en gemeenten nemen heel vaak deel aan benchmarking door *aanmelding en inschrijving*. Benchmarking blijkt in het openbaar bestuur ook een collectief proces te zijn waarbij de individuele deelnemers prestatiegegevens verzamelen in het kader van een aanpak zoals het onderzoeksbureau die voor ogen staat. De uitkomsten worden centraal verwerkt door een extern bureau. De uitkomsten in de vorm van tabellen en grafieken worden daarna aan de deelnemers ter beschikking gesteld, zodat deze de resultaten van de 'rangordening' verder kunnen analyseren.

### **Tweede mythe**

De tweede mythe luidde: Benchmarking leidt tot het opsporen van 'best practices' en bezinning daarop in benchmarkrapportages. Op dit vlak is op grond van onze huidige inzichten sprake van teleurstelling. In veel benchmarkrapporten komt een aanduiding van excellente of goede praktijken namelijk helemaal *niet* expliciet voor. Meergenoemde benchmark over de riagg's kende geen beste praktijken en de raad van toezicht heeft het daarover dan ook niet gehad. Logisch in zoverre dat het benchmarkrapport ook geen visie op succesvolle therapieën over het helpen van ggz-geïndiceerde cliënten kon bevatten. Er is ook (nog) geen studie te vinden die aanwijst wat aan goede praktijken is overgenomen door ziekenhuizen, thuiszorginstellingen, politiekorpsen, rechtbanken, universiteiten, waterschappen of gemeenten. Toch komt navolging (imitatie) op zich wel voor, zo leert de organisatiekunde, maar het bewijs is niet gevonden dat positieve imitatie in sterke mate het directe of indirecte gevolg is van uitkomsten van benchmarking.

### **Derde mythe**

De als derde genoemde mythe hield in: Benchmarking draagt bij aan overheidsorganisaties die beleid (in casu prestaties, producten, processen) door benchmarking daadwerkelijk verbeteren en zich daarmee een lerende organisatie tonen. Deze mythe is geïnterpreteerd als: er komt 'één cirkel leren' en 'dubbel cirkel leren' voor (Argyris en Schön, 1978). Voor dubbel cirkel leren, waarbij waarden en doelstellingen achter beleid ter discussie komen, zijn echter geen harde aanwijzingen verkregen in benchmarkstudies die handelen over afzonderlijke sectoren, als die van gemeenten, waterschappen, en de genoemde mbo of riagg (waarin via een lidmaatschap van de raad van toezicht geparticipeerd werd).

Voor zover er wél sprake is van leren, gaat het vermoedelijk eerder om 'één cirkel leren': leren binnen de bestaande (beleids)kaders. Wat houdt dat leren eigenlijk in?

We hebben leren gezien als het benutten van benchmarkinformatie. Benchmarkonderzoeken over burgerzaken, afvalbeheer, werk van sociale diensten en sport – die werden behandeld - zetten het bestaande gemeentelijk beleid van deelnemers aan de benchmarking over het algemeen voort en dus niet op de kop. Van benchmarking wordt door ambtenaren aan opdrachtgeverskant bijna altijd wel kennis genomen. Tenslotte bevredigt benchmarking ook de nieuwsgierigheid. Maar dat is een heel zwakke mate van benutting van bevindingen. Is er meer? Benchmarkrapporten worden wel ter

kennis gebracht van besturen maar zelden expliciet besproken in een gemeenteraad of in algemene besturen van waterschappen. Expliciete bezinning in ambtelijke werkgroepen ter voorbereiding op de bestuurlijke koers, anders dan in genoemde kringen, is schaars.

Om welke onderzoeksfuncties gaat het bij bezinning dan? Benchmarking dient menigmaal ter legitimering van bestaand beleid of reeds ingezette of overwogen marginale aanpassingen in beleid of werkwijzen. Anders gesteld, benchmarkuitkomsten dragen als deze al het bestuurlijk benut worden, meer bij aan incrementele veranderingen (verandering bij stukjes en beetjes) dan aan synoptische. Conclusie: de mythe zal blijven bestaan maar dat benchmarking tot daadwerkelijke verbeteringen in beleid of werkwijzen leidt, moet als algemeen beeld van veel vraagtekens worden voorzien. Relativering past. Grote verwachtingen op dit punt worden niet waargemaakt. Uiteindelijk blijft het een bestuurlijke kwestie: welke benutting is goed genoeg?

#### ***Vierde mythe***

De vierde genoemde mythe hield het volgende in: Verbeteren is in elk geval financieel-economisch gericht. Benchmarking van vergelijkbare organisaties draagt bij aan meer bedrijfsmatig functionerende overheidsorganisaties. Benutting van benchmarkrapporten zou dan, in onze operationalisatie, moeten leiden tot meer doelmatige uitvoering van taken. De reactie hierop luidt: hier is – in aansluiting op de tweede mythe - nog geen substantiële aanwijzing voor gevonden, hoewel benchmarking van zuiveringsschappen wel heeft bijgedragen aan stappen voorwaarts. Maar, benchmarking doet ook geen afbreuk aan bedrijfsmatige stappen voorwaarts, zeker niet als de valkuilen bij benchmarking vermeden worden. Hoe meer benchmarking is ingebed in een lerende cultuur en expliciete stimulansen van bestuurders en topambtenaren om te willen leren geventileerd worden, en hoe beter de kwaliteit van de toepassing van benchmarking, hoe meer kans op een bijdrage aan bedrijfsmatig functionerende overheidsorganisaties.

Het voorgaande impliceert dat geen bevestiging is te vinden voor de stelling dat benchmarking in het openbaar bestuur al tot veel verbeteringen in termen van het radicaal herzien van beleid heeft geleid (cf Blank, Felsö en De Groot, 2011: 9). Het herzien van beleid als gevolg van benchmarking kan ook nauwelijks worden bereikt omdat benchmarking geen beleidsevaluatie is, cijfers uit benchmarks (nog) zelden geplaatst worden in het licht van wat een overheid wilde bereiken en omdat rapporten door controlerende organen, als gemeenteraden, zelden nadrukkelijk wordt gevraagd. Ergo: de politieke urgentie en prioriteit is niet groot genoeg.

Rapporten leiden ook zelden tot het evidente beeld dat een organisatie het op een bepaald terrein zeer slecht doet. De evidentie hiervan ontbreekt in benchmarkrapporten. Zou dat wel het geval zijn dan zou de druk om te verbeteren groot of groter zijn.

Benchmarking geeft volgens de interpretatie van ambtenaren wel vaak een (diffuse) bevestiging van bestaand beleid ('we doen de goede dingen redelijk goed'). Of er wordt gezegd: benchmarking 'versterkt de richting van de verandering die toch al werd ingeslagen'. Zo gezien vervult benchmarking nogal eens een legitimatiefunctie. Tot heftige politiek-bestuurlijke debatten of controverses geeft benchmarking zelden of nooit aanleiding.

#### ***Algemeen beeld aangevuld***

Deze ervaringen bevestigen het volgende beeld dat uit praktijkervaringen in brede zin naar voren komt en zeker ook lijkt te gelden voor gemeenten uit de jaren 1990-2010. Deze praktijkervaringen van grofweg twee decennia benchmarking in het Nederlandse openbaar bestuur op decentraal niveau hebben we gerubriceerd naar de kenmerken voor benchmarking zoals we onder 3 hebben aangegeven.

Als eerste kenmerk het (rang)ordenen van uitkomsten ('scores') op vergelijkingspunten:

- Prestaties van organisaties die aan een vergelijkingsoperatie deelnemen, worden op een rij gezet in een kwantitatieve vergelijking (tabellen, grafieken of andere manieren). Op deze wijze is een rangordening mogelijk van prestatiemeting.
- Benchmarkonderzoek richt zich meer op prestaties (output) dan op gewenste effecten (outcomes).

Als tweede kenmerk het vergelijken van organisaties, delen of aspecten daarvan in kwantitatief opzicht door middel van scores:

- Een vergelijking van organisaties wordt in principe door veel betrokken bestuurlijk en ambtelijk leidinggevendend zinnig gevonden, maar van de gekozen indicatoren is lang niet altijd duidelijk of ze de kern van een prestatieveld of productie(cluster) betreffen en voldoende dekkend zijn.

Als derde kenmerk het trachten te begrijpen van de scores ('het verhaal achter de cijfers'):

- Bij benchmarking gaat het over het algemeen, uitzonderingen daargelaten, voornamelijk of uitsluitend om kwantitatieve vergelijkingen tussen organisaties en geen kwalitatieve, die rekening houden met of aangevuld worden met verklaringen van substantiële verschillen in scores tussen organisaties,
- Verklaringen voor scores ('het waarom?') ontbreken veelal in rapportages en
- De interpretatie van uitkomsten is in benchmarkrapportages afwezig of uitermate sober.

Als vierde kenmerk het zoeken naar 'best practices':

- Expliciete normeringen van scores over wat goede of slechte scores zijn of streefcijfers zijn, ontbreken vaak in benchmarkrapporten.
- 'Best practices' worden vaak niet in benchmarkrapporten vermeld. Als 'best practices' wel worden vermeld, en ook geloofwaardig zijn ('geen toeval'), is onduidelijk of die over te nemen zijn in een andere context.
- Aanbevelingen ontbreken in veel benchmarkrapporten.

Als vijfde kenmerk het verwerken van de uitkomsten in rapportages:

- De deelname aan een benchmark kan doorgaans op een zekere ambtelijke en bestuurlijke welwillendheid rekenen, hoewel bestuurders vaak niet nadrukkelijk betrokken worden bij een reflectie op de aard en inhoud van de opdracht tot benchmarkonderzoek.
- Het benchmarkonderzoek mondt vaak uit in één benchmarkrapport voor alle opdrachtnemers (lees ook: deelnemers) samen en niet voor elk apart.
- Een vertaalslag van de kwantitatieve vergelijking, en de bevindingen die uit deze vergelijking voortkomen, en de verklaring van verschillen is nodig naar de eigen organisatie om ambtelijke en bestuurlijke betekenis te krijgen. Deze (contextgebonden) vertaalslag wordt vaak niet door een opdrachtgever bij de opdrachtverlening geëist. Veel opdrachtnemers komen zelf in het kader van de opdrachtverlening niet met een suggestie op dit vlak en stellen ook later geen apart rapport op als vanzelfsprekend onderdeel van de opdracht.

Als zesde het verbeteren van de organisatie in de richting van de betere organisaties of betere praktijk:

- Benchmarking komt in het openbaar bestuur, ook in en tussen gemeenten, volop voor en is vaak (qua doel) gericht op verbetering van de bedrijfsvoering bij overheidsorganisaties.
- Veel benchmarkrapporten bevatten geen expliciete aanbevelingen aan onderzochte organisaties over hoe procesmatig en inhoudelijk te komen tot een beleidswending of verbetering van beleid bij stukjes- en-beetjes.
- De lezers en andere gebruikers van benchmarkrapporten nemen van de inhoud wel kennis, maar de rapportage leidt niet tot een zware en expliciete reflectie in bestuurlijke kringen.

Uit het beeld dat hier geschetst wordt blijkt met name de smalle opvatting van benchmarking, zoals we deze in 3 hebben onderkend, dominant.

### **Aanbeveling**

De keizer ('benchmarking') heeft minder kleren aan dan gedacht. Indien veel van de genoemde valkuilen bij benchmarking daadwerkelijk voorkomen, is de betekenis van benchmarking in de praktijk heel klein. Maar het kan anders. Voorkom daarom in opdrachtgever-opdrachtnemerrelaties dat tal van valkuilen bij benchmarking optreden en positieve voorwaarden juist wel aanwezig zijn.

Daarmee mondt deze beschouwing uit in een aanbeveling. Die luidt: opdrachtgevers moeten serieuzer werk maken van hun opdrachtgeverrol als ze in benchmarking een opstap zien naar 'leren' en 'verbeteren' in hun eigen organisatie. Daarmee sluiten we aan op een suggestie die Helder (2001; zie ook De Bruijn, 2001; Blank, 2010) eerder al eens deed. Serieuzer werk maken van de opdracht voor benchmarkonderzoek en de entreebesprekingen betekent in elk geval

- dat een opdrachtgever de verwachtingen over een benchmark naar mede-opdrachtgevers en vooral naar een opdrachtnemer van het benchmarkonderzoek moet specificeren,
- moet nagaan hoeveel en welke relevante gemeenten in de vergelijking meegenomen kunnen of moeten worden,
- moet bijdragen aan het verkrijgen van meerjarencijfers,
- een oorzakenanalyse van verschillen tussen scores onderdeel moet maken van de benchmark,
- een afzonderlijk rapport voor de eigen organisatie moet vragen met cijfers, interpretaties en 'best practices' die voor de eigen gemeente relevant zijn
- en sterk betrokken moet blijven in een begeleidingscommissie van de benchmarking gedurende het onderzoek.

Dit vereist dus dat de opdrachtgevers van te voren goed hun doel moeten hebben geformuleerd wat zij beogen te bereiken met benchmarking en met de uitkomsten willen gaan doen. Gaat het om slechts het verkrijgen van inzicht? Of wil een opdrachtnemer werkelijk komen tot prestatieverbetering? Dit heeft consequenties voor het type benchmark.

Benchmarking is nu veel te veel verworden tot een kwestie van 'voor een dubbeltje op de eerste rang willen zitten'. Opdrachtgevers krijgen suboptimale benchmarkrapporten omdat ze niet voldoende willen betalen voor kwalitatief hoogstaande benchmarking en daarmee voor 'leren', of zelf niet voor ogen hebben wat met benchmarking te bereiken.

Benchmarking komt voort uit meer bedrijfsmatig willen werken, streeft over het algemeen ook meer of betere bedrijfsmatigheid na, maar heeft zelf ook kwalitatief opzicht last van die bedrijfsmatigheid, namelijk van te weinig budget willen reserveren voor benchmarking.

### **Literatuur**

Aardema, H. en A.F.A.Korsten, *De staat van de gemeente*, Inaxis, Den Haag, 2005.

Aardema, H., A. Korsten, K. Riezebos en M. van Dam, *De vallende wethouder*, Ministerie van BZK, Den Haag, 2011.

Aardema, H., M.J.G.J.A. Boogers en A.F.A. Korsten, Vallende wethouders – Een verkenning van de vertrekredenen van onvrijwillig teruggetreden bestuurders op lokaal niveau, in: *Bestuurswetenschappen*, jrg. 66, 2012, nr. 2, pp. 13-34.

Aardema, H., W. Derksen, M. Herweijer en P. de Jong (red.), *Meerwaarde van de bestuurskunde – Liber amicorum voor prof.dr. Arno F.A. Korsten*, Boom Lemma, Den Haag, 2010.

Abma, K. en A.F.A. Korsten, Burgers beoordelen hun gemeente. Waar zijn gemeenten volgens hun burgers goed in en waarin minder goed?, in: *Tijdschrift voor Public Controlling*, april 2009c, pp. 35-39.

Abma, K. en A.F.A. Korsten, *Gemeenten in rapportcijfers*, Uitgeverij Eburon, Delft, 2009.

Abma, K. en A.F.A. Korsten, Normenkaders bij bestuurskrachtmetingen - Een beeld van de ideale gemeente?, in: *Bestuurswetenschappen*, jrg. 62, 2008, nr. 6, pp. 42-67.

Abma, K., *Beoordelen van gemeenten*, Wolf Legal Publishers, Nijmegen, 2012.

Albeda, H. en A.L. Buinink, *Benchmark responsesnelheid 8 gemeenten*, Stichting Rekenschap, Den Haag, 2004.

Allers, M.A. e.a., *Atlas van de lokale lasten 2013*, COELO, Groningen, 2013.

Allers, M.A. en C. Hoeben, *Kerngegevens belastingen grote steden 2013*, COELO, Groningen, 2013.

Ammons, D.N., *Municipal Benchmarks. Assessing Local Performance and Establishing Community Standards*, M.E. Sharpe, New York, 2012 (third edition; 1996: first edition).

Anand, G. & R. Kodali, Benchmarking the benchmarking models, in: *Benchmarking - An International Journal*, vol. 15, 2008, nr. 3, pp. 257-291.

Ananias, J.H.G.H., *Benchmarking bij Sociale Diensten- een analyse van kwaliteit en benutting van vier benchmarks bij gemeentelijke sociale diensten van 2000 – 2007*, Open Universiteit Nederland, Heerlen, 2007 (doctoraalscriptie).

Ankersmit, F. en L. Klinkers (red.), *De tien plagen van de staat – De bedrijfsmatige overheid gewogen*, Van Gennep, Amsterdam, 2008.

Argyris, C. & D. Schön, *Organizational Learning - A Theory of Action Perspective*, Addison-Wesley, Reading, MA, 1978.

Argyris, C., Double loop learning in organizations, in: *Harvard Business Review*, vol. 55, 1977, nr. 5, pp. 115-125.

Askim, J., Determinants of Performance Information Utilization in Political Decision Making, in: Dooren, W. van, & S. van de Walle (eds.), *Performance Information in the Public Sector- How it is used*, Palgrave MacMillan, Londen, 2008, pp. 125-140.

Askim, L., A. Johnson & K.A. Christophersen, Factors behind organizational learning from benchmarking: Experiences from Norwegian municipal benchmarking networks, in: *Journal of Public Administration Research and Theory*, vol. 18, 2008, nr. 2, pp. 297-320.

Askim, L., *Local Government by Numbers – Who Makes Use of Performance Information, When and For What Purposes?*, University of Oslo, Oslo, 2007.

Baakman, N.A.A. (red.), B.L. Becker, A.M.B. Michels en J.G.A. van Mierlo, *Overheid en onderneming – Een inleiding*, OU, Heerlen, 2001.

Barzelay, M., *The New Public Management – Improving Research and Policy Dialogue*, University of California Press, Berkeley, 2001.

Behn, R.D., *Rethinking democratic accountability*, Brookings, Washington, 2001.

Behn, R.D., Why measure performance? - Different Purposes Require Different Measures, in: *Public Administration Review*, 2003, vol. 63, nr. 5, pp. 586-606.

Bekkers, V., *Beleid in beweging, Boom/Lemma*, Den Haag, 2012.

Bentlage, F.A. e.a., *De excellente overheidsorganisatie*, Kluwer, Deventer, 1998.

*Bestuurswetenschappen*, jrg. 64, 2010, nr. 4, pp. 8-28.

Blank, J. en H. de Groot, Paspoort dunder door gemeentelijke schaalvergroting, in: *MeJudice*, 26 november 2012.

Blank, J., F. Felsö en H. de Groot, Beter benchmarken, in: *TPC* (tijdschrift public controlling), dec. 2011, pp. 4-9.

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 productiviteitsstructuur van verpleeghuizen*, SCP/Vuga, Rijswijk, 1996.

Blank, J.L.T., A.C.M. Dumaij en B.L. van Hulst, *Ziekenhuismiddelen in verband – Een empirisch onderzoek naar productiviteit en doelmatigheid in de Nederlandse ziekenhuizen 2003-2009*, IPSE, Delft, 2011.

Blank, J.L.T., Benchmarken of de kunst van het vergelijken, in: *ESB*, 29 mei 1998, pp. 432-435.

Blank, J.L.T., *Principes van productiviteitsmeting*, Shaker Publishing, Maastricht, 2010.

Bogt, H.J. ter, P. Jansen en S. Tillema (red.), *Met het oog op de publieke zaak*, RUG, Groningen, 2010.

Bogt, H.J. ter, Bronnen van controlinformatie voor politieke bestuurders, in: *Maandblad voor Accountancy en Bedrijfseconomie*, jrg. 75, 2001a, nr. 12, pp. 525-538.

Bogt, H.J. ter, De doelmatigheid van outputgerichte managementinformatie voor wethouders, in: *Beleidswetenschap*, jrg. 16, 2002, nr. 2, pp. 114-144.

Bogt, H.J. ter, en J.G. van Helden, Kwaliteit van prestatiemeting bij gemeentelijke overheden, in: *Bestuurskunde*, 1994, nr. 1, pp. 2-13.

Bogt, H.J. ter, Gebruik van outputinformatie door gemeentebestuurders, in: *Overheidsmanagement*, 2000, nr. 7/8, pp. 183-187.

Bogt, H.J. ter, *Neo-institutionele economie, management control en verzelfstandiging*, Labyrint, Cappelle a.d. IJssel, 1997.

Bogt, H.J. ter, Performance evaluation styles in governmental organizations: how do professional managers facilitate politicians' work?, in: *Management Accounting Research*, vol. 14, 2003, nr. 4, pp. 311-332.

Bogt, H.J. ter, Politicians and output-oriented performance evaluation in municipalities, in: *The European Accounting Review*, vol. 10, 2001b, pp. 621-643.

Bogt, H.J. ter, Politieke bestuurders, productbegrotingen en prestatiebeoordelingen bij gemeenten, in: *Maandblad voor Accountancy en Bedrijfseconomie*, jrg. 73, 1999, nr. 12, pp. 671-683.

Bogt, H.J., ter, *Managementvernieuwing bij de overheid - Mooie woorden of echte daden?*, Groningen, 2005 (oratie).

Bogt, H.J., ter, Prestatiegegevens: nuttig voor sturing van gemeentelijke taken, in: *Public Controlling*, juni 2004, pp. 52-58.

Boivard, T. & E. Löffler (eds.), *Public Management and Governance*, Routledge, Londen, 2003.

Bongers, F.J. en D. Verlet, Indicatoren en beleid: doe meer met ongeveer, in: *Bestuurswetenschappen*, jrg. 67, 2013, nr. 1, pp. 36-55.

Bordewijk, P. en H. Klaassen, *Wij laten ons niet kennen – Een onderzoek naar het gebruik van kengetallen bij negen grotere gemeenten*, VNG Uitgeverij, Den Haag, 2000.

Borst, R.T., C.J. Lako en M.S. de Vries, Is prestatie management in de publieke sector toepasbaar?, in: *Bestuurswetenschappen*, jrg. 66, aug. 2012, nr. 4, pp. 14-35.

Bouckaert, G. en T. Auwers, *Prestaties meten in de overheid*, Brugge, 1999.

Bouckaert, G. & J. Halligan, *Managing performance. International comparisons*, Routledge, Londen, 2008.

Bouckaert, G. & J. Halligan, Performance and performance management, in: B. Guy Peters & J. Pierre (eds.), *Handbook of Public Policy*, Sage, Londen, 2006, pp. 443-460.

Bouckaert, G. & W. Balk, Public Proactivity Measurement: Diseases and Cures, in: *Public Productivity and Management Review*, vol. 15, 1991, nr. 2, pp. 229-235.

Bouckaert, G. & W. van Dooren, Performance Measurement and Management in Public Sector Organizations, in: Boivard, T. & E. Löffler (eds.), *Public Management and Governance*, Routledge, Londen, 2009, pp. 151-165.

Bouckaert, G., A. Hondelghem, J. Voets, S. op de Beeck en E. Cautlaert (red.), *Handboek overheidsmanagement – Overheid in beweging*, Vanden Broele, Brugge, 2011.

Bouckaert, G., B. De Peuter en W. van Dooren, *Meten en vergelijken van lokale bestuurlijke ontwikkeling: een monitoringsysteem voor het lokaal bestuur in Vlaanderen*, Die Keure, Brugge, 2003.

Bouckaert, G., Improving Public Performance, in: Halachmi, A. & G. Bouckaert (eds.), *The Enduring Challenges in Public Management*, Jossey-Bass Publ., San Francisco, 1995, pp. 379-413.

Bouckaert, G., Prestaties en prestatie management in de publieke sector, in: *Tijdschrift voor Economie en Management*, 2006, nr. 3, pp. 237-265.

Bouckaert, G., *Productiviteit in de overheid*, VCOB, Leuven, 1990 (diss.).

Bouckaert, G., Renewing Public Sector Performance Management, in: Jann, W., M. Röber & H. Wollmann (Hrsg.), *Public Management: Grundlagen, Wirkungen, Kritik*, Sigma, Berlijn, 2006, pp. 119-133.

Bowerman, M., G.A.J. Francis e.a., The Evolution of Benchmarking in UK Local Authorities, in: *Benchmarking: An International Journal*, vol. 9, 2002, pp. 429-449.

Boyne, G.A., K.J. Meier e.a. (eds.), *Public Sector Performance – Perspectives on Measurement and Management*, Cambridge UP, Cambridge, 2006.

Brans, M., A.P.M. Giesbers en A.J. Meijer, Alle ogen op ziekenhuizen gericht? De effecten van openbaarheid van prestatiegegevens, in: *Bestuurswetenschappen*, jrg. 62, 2008, nr. 2, pp. 32-53.

Brinkerhoff, R.O. & D.E., Dressler, *Productivity Measurement – A Guide for Managers and Evaluators*, Sage, Londen, 1990.

Brouwer, N. en G. J. van Helden, Benchmarking en prestatieverbetering in de publieke sector. Ervaringen met de bedrijfsvergelijking zuiveringsbeheer, in: *Management Accounting*, jrg. 79, 2005, nr. 11, pp. 573-581.

Bruijn, H. de, Afrekenen op prestatie leidt tot bureaucratie, in: *NRC*, 3 jan. 2003.

Bruijn, H. de, Maak spanningen management & professionals vruchtbaar, in: Brink, G. van den, Th. Jansen en D. Pessers (red.), *Beroepszeer*, Boom, Amsterdam, 2005, pp. 143-155.

Bruijn, H. de, *Managers en professionals*, Sdu, Den Haag, 2008.

Bruijn, H. de, *Managing Performance in the Public Sector*, Routledge, Londen, 2007.

Bruijn, H. de, Outputsturing in publieke organisaties, in: *M&O*, mei/juni 2002, nr. 35-22.

Bruijn, H. de, *Prestatiemeting in de publieke sector*, Lemma, Utrecht, 2001.

Brunings, D., *Benchmarking bij Nederlandse gemeenten – Van meten tot leren*, Erasmus Universiteit, Rotterdam, 2011.



Bullivant, J.R.N., *Benchmarking for Continuous Improvement in the Public Sector*, Longman, Essex, 1994.

Callahan, K., *Elements of Effective Governance*, Taylor & Francis, Boca Raton, 2007.

Camp, R.C., *Benchmarking – Het zoeken naar de beste werkmethode die leiden tot superieure prestaties*, Kluwer, Deventer, 1992.

Castenmiller, P., M. Meesters en B. Pluut, *De achtergronden van gemeentelijke prestaties*, Zenc, dec. 2006.

Charbonneau, E. & F. Bellavance, Blame Avoidance in Public Reporting: Evidence from a Provincially Mandated Municipal Performance Measurement Regime, in: *Public Performance & Management Review*, vol. 35, march 2012, nr. 3, pp. 399-422.

Cleuver, E. en J. Blank, Productieve politie – Een empirische analyse van de kostenstructuur van de Nederlandse politie, in: *Beleidswetenschap*, jrg. 20, 2006, nr. 3, pp. 52-67.

Commissie-Bovens, *Wil tot verschil – Gemeenten in 2015*, VNG, Den Haag, 2006.

Commissie-Kuiper, *Verbinding verbroken? Onderzoek naar de parlementaire besluitvorming over de privatisering en verzelfstandiging van overheidsdiensten*, Eerste Kamer, Den Haag, 2012.

Commissie-Meijerink, *Recht van spreken: Interdepartementaal beleidsonderzoek naar de bedrijfsvoering rechtspraak*, Ministerie van Justitie, Den Haag, 1999.

Cörvers, R.J.M., P. Glasbergen en A.F.A. Korsten, Netwerksturing bij natuurontwikkeling, in: *Bestuurswetenschappen*, juni 2003, pp. 205-228.

Dijke, A.J., *Van Benchmarken naar Benchlearnen. Een onderzoek naar het leren van gemeenten van de benchmark Work First (2007)*, Erasmus Universiteit, Rotterdam, 2009.

Doorduyn, W.A., e.a., *Rotterdam vergelijkenderwijs – Benchmarks voor de verbetering van de bedrijfsvoering*, Gemeente Rotterdam, Rotterdam, 2002.

Dooren, W. van, en S. van de Walle (eds.), *Performance Information in the Public Sector- How it is used*, Palgrave, Londen, 2008.

Dooren, W. van, G. Bouckaert & J. Halligan, *Performance Management in the Public Sector*, Routledge, Londen, 2010.

Dorsch, J.J. & M.M. Yasin, A Framework for Benchmarking in the Public Sector, in: *International Journal of Public Sector Management*, vol. 11, 1998, nr. 2/3, pp. 91-115.

Douma, Th., R. Stevens en R. in 't Veld, *Naar (h)erkende kwaliteit in het mbo*, Audit Akademia, Amersfoort, 2012.

Ecorys, *Benchmark Gemeentelijk Ondernemingsklimaat: Thematische Rapportage, eindmeting*, Ministerie van Economische Zaken/ECORYS, Den Haag/Rotterdam, 2010.

Elias, J., J.Th. van Ieperen e.a., *Doelmatigheid in de thuiszorg*, Bakkenist Management consultants, Amsterdam, 1997.

Ende, D.A.P.W. van der, Bespreking van 'G. Bouckaert, Productiviteit in de overheid', in: *Openbare Uitgaven*, 1991, nr. 1, pp. 33-37.

Engbersen, R. e.a., *Nederland aan de monitor*, NIZW Uitgevers, Utrecht, 1997.

Espeland, W.N. & M. Sauder, Rankings and Reactivity: How Public Measures Recreate Social Worlds, in: *American Journal of Sociology*, vol. 113, 2007, nr. 1, pp. 1-40.

Espeland, W.N. & M.L. Stevens, A Sociology of Quantification, in: *European Journal of Sociology*, vol. 49, 2008, nr. 3, pp. 401-436.

Espeland, W.N. en M.L. Stevens, Commensuration as a Social Process, in: *Annual Review of Sociology*, vol. 24, 1998, nr. 1, pp. 313-343.

Felsö, F., H. de Groot en A. van Heezik, *Benchmark gemeentelijk afvalbeheer – Een empirisch onderzoek naar de productiviteit en kostendoelmatigheid*, IPSE Studies 1011-6, TU Delft, 20-11.

Felsö, F., J. Wildschut, H. de Groot en Th. Niaounakis, *Contractvoorwaarden en de kostendoelmatigheid van gemeentelijk afvalbeheer*, IPSE-studie- 2012, TUDelft, dec. 2012.

Ferber, G., Benchmarking: wanneer is het best?, in: Berg, C. e.a. (red.), *Moderne bedrijfsvoering bij de overheid*, Sdu, Den Haag, 2003, pp. 193-208.

Francis, G. & J. Holloway, What have we learned? Themes from the literatuur on best practice benchmarking, in: *International Journal of Management Reviews*, vol. 9, 2007, nr. 3, pp. 171-189.

Frey, B.S. & M. Osterloh, Onderzoeksevaluaties: verborgen kosten, twijfelachtige voordelen en betere alternatieven, in: Jansen, Th., G. van den Brink en J. Kole (red.), *Beroepstrots*, Boom, Amsterdam, 2009, pp. 194-218.

Freytag, P.V. & S. Hollensen, The process of benchmarking, benchlearning and benchaction, in: *The TQM Magazine*, vol. 13, 2001, nr. 1, pp. 25-33.

Gerrichhauzen, J., A.F.A. Korsten en H. Feijen, De lerende organisatie in een ambtelijke cultuur, in: *Bestuurswetenschappen*, jrg. 56, 2002, nr. 3, pp. 221-236.

Goodijk, R., *Falend toezicht in semipublieke organisaties?*, Van Gorcum, Assen, 2012

Gravesteyn, M., F. Evers, C. Wilderom en M. Molenveld, *Leren van presteren op de werkvloer via zelfontwikkelde prestatie-indicatoren*, in: *M&O*, sept./okt. 2011, pp. 62-82.

Grit, K., *Economisering als probleem*, Van Gorcum, Deventer, 2000.

Groot, H. de (red.), *Benchmarking in de publieke sector*, Sdu, Den Haag, 2004.

Groot, H. de, en R. Goudriaan, *De produktiviteit van de overheid*, Academic Service, Schoonhoven, 1991.

Hakvoort, J. & H. Klaassen, *Benchmarking in non-profit organisaties*, in: *Beleidsanalyse*, 1999, nr. 3, pp. 11-20.

Hakvoort, J.L.M. en H.L. Klaassen, *Bedrijfsvoeringstechnieken voor overheid en non-profitorganisaties*, Sdu, Den Haag, 2004.

Haselbekke, A. e.a., *Prestaties tellen*, VNG, Den Haag, 1990.

Haselbekke, A.G.J., *Doelmatigheid en doeltreffendheid van gemeentelijk beleid*, in: Korsten, A.F.A. en P.W. Tops (red.), *Lokaal bestuur in Nederland*, Samsom, Alphen a.d. Rijn, 1998, pp. 554-567.

Helden, G. van, *BBI in de praktijk*, Groningen, 1999.

Helden, G.J. van, *De bedrijfsmatige gemeente als lege huls?*, in: *B&G*, jrg. 28, 2001, nr. 9, pp. 26-31.

Helden, G.J. van, & Å. Johnsen, *A Comparative Analysis of the Development of Performance-based Management Systems in Dutch and Norwegian Local Government*, in: *International Public Management Journal*, 2002, nr. 5, pp. 79-95.

Helden, G.J. van, & H.J. ter Bogt, *The Application of Businesslike Planning & Control in Local government: A Field Study of Eight Dutch Municipalities*, in: *Local Government Studies*, vol. 27, 2001, nr. 1, pp. 61 -86.

Helden, G.J. van, en S. Tillema, *In Search of a Benchmarking Theory for the Public Sector*, in: *Financial Accountability & Management*, vol. 21, 2005, nr. 3, pp. 237-261.

Helden, G.J. van, en E. Jansen, *New public management bij gemeenten*, in: *Overheidsmanagement*, jrg. 15, 2002, nr. 3, pp. 64-68.

Helden, G.J. van, en N. Brouwer, *Benchmarking en prestatieverbetering in de publieke sector: ervaringen met de bedrijfsvergelijking zuiveringsbeheer*, in: *MAB*, jrg. 79, 2005, nr. 11, pp. 574-583.

Helden, G.J. van, en S. Tillema, *Naar een benchmarkingtheorie voor de publieke sector*, in: *Tijdschrift voor Openbare Financiën*, jrg. 36, 2004, nr. 3, pp. 99-111.

Helden, G.J. van, *Normering prestatie-indicatoren, begin- of eindpunt van doelmatigheid*, in: *B&G*, september 1997, pp. 8-12.

Helden, G.J. van, S. Tillema en R. van Esch, *Benchmarking in de Publieke Sector. Ervaren met de bedrijfsvergelijking zuiveringsbeheer*, in: *Economienblad*, jrg. 28, 2005, juni, pp.13-16.

Helden, G.J. van, S. Tillema, M. Kuppens en J.W.C. Dekking, *Werkt benchmarking bij waterschappen?*, in: *Bestuurskunde*, jrg. 14, maart 2005, nr. 2, pp. 30-39.

Helden, G.J. van, en N. Brouwer, *Benchmarking en prestatieverbetering in de publieke sector: ervaringen met de bedrijfsvergelijking zuiveringsbeheer*, in: *Maandblad voor Accountancy en Bedrijfseconomie*, jrg. 79, 2005, nr. 11, pp. 574-583.

Helder, S., *Op weg naar structurele inzet benchmark-instrument*, in: *Bank & Gemeente*, maart 2001, pp. 12-14.

Herweijer, M. en F.M. Noordam, *Inleiding*, in: M. Herweijer e.a. (red.), *Alles in één keer goed. Juridische kwaliteit van bestuurlijke besluitvorming*, Kluwer, Deventer, 2005.

Hiemstra, J., *Presterende gemeenten*, Kluwer, Deventer, 2003.

Hood, C., *A Public Management For All Seasons?*, in: *Public Administration*, vol. 69, 1991, nr. 1, pp. 3-19.

Hood, C., R. Dixon & C. Beeston, *Rating the Rankings: Assessing International Rankings of Public Service Performance*, in: *International Public Management Journal*, vol. 11, 2008, nr. 3, pp. 298-328.

Hood, C., *The 'New Public Management' in the 1980s. Variations on a Theme*, in: *Accounting, Organisations and Society*, vol. 20, 1995, nr. 2/3, pp. 93-109.

Hoogwout, M., *Het ontwerpen van overheidsbenchmarks. Een krachtig management- en verantwoordingsinstrument*, in: *Public Controlling*, 2004, juni, pp. 32-42.

Huijben, M. en A. Geurtsen, *Heeft iemand de overhead gezien? Een beproefde methode om de overhead te managen*, Academic Service, /Sdu, Den Haag, 2008.

Hulst, B.L. van, en H. de Groot, *Benchmark burgerzaken*, IPSE Studies 2011-7, TU Delft, 2011.

Jansen, R.W.J., *Prestatie-onderzoek naar Nederlandse overheidsbedrijven*, UM, Maastricht, 1993.

Jochoms, M.P.C.M. en A. Sey, *Naar intelligent presteren: uitdagingen voor leiderschap bij sturing met prestaties bij de politie*, in: *Tijdschrift voor Veiligheid*, jrg. 6, 2007, nr. 1, pp. 21-35.

Jong, R. de, *De balans verstoord- Een rapport over de corporatiesector ten behoeve van de Parlementaire Enquête Woningcorporaties*, Aedes, Den Haag, februari 2013.

Jurriëns, J. en J. Sylvester, *Verzelfstandigde overheidsorganisaties – Wat heeft verzelfstandiging ons geleerd? Een bedrijfskundige benadering*, in: *M&O*, mei/juni 2005, pp. 42-60.

Kaashoek, B., G. Ongena en J. Raab, Netwerken die werken? Netwerkanalyse als instrument voor beleidsevaluatie, in: *Bestuurswetenschappen*, jrg. 63, 2009, nr. 5, pp. 55-73.

Karsten, L. en K. van Veen, *Managementconcepten in beweging: tussen feit en vluchtigheid*, Van Gorcum, Assen, 1998.

Keehley, P. (ed.), *Benchmarking for best practices in the public sector – Achieving performance breakthroughs in federal, state and local agencies*, San Francisco, 1996.

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.

Knaap, P. van der, en J. van Hofwegen, Naar een resultaatgericht sturingsmodel voor de Belastingdienst, in: *Bestuurskunde*, 1999, nr. 4, pp. 145-155.

Kool, D. de, *Monitoring in beeld*, Erasmus Universiteit Rotterdam, Rotterdam, 2007.

Korsten, A. en M. Schoenmaker, *Sterke colleges – De kracht van colleges van burgemeester en wethouders in 37 adviezen*, Reed Business, Den Haag, 2008.

Korsten, A., De Nederlandse overheidsmanager en de 'nieuwe zakelijkheid', in: Bouckaert, G., A. Hondeghem en R. Maes (red.), *De overheidsmanager*, KULeuven, 1994, pp. 123-162.

Korsten, A.F.A. en H. Aardema, *De vallende burgemeester*, Ministerie van BZK, Den Haag, 2006.

Korsten, A.F.A. en M. Noordegraaf, Plaatsbepaling van publiek ondernemerschap, in: Noordegraaf, M., A.B. Ringeling en F.J.M. Zwetsloot (red.), *De ambtenaar als publiek ondernemer*, Uitgeverij Coutinho, Bussum, 1995, pp. 20-52.

Korsten, A.F.A. en M. Schoenmaker, Gezond bestuur en bestuurlijke probleemgemeenten: tijd voor een bestuurschout?, in: *Bestuurswetenschappen*, jrg. 65, 2011, nr. 3, pp. 28-50.

Korsten, A.F.A. en P.W. Tops (red.), *Lokaal bestuur in Nederland*, Samsom, Alphen a.d. Rijn, 1998.

Korsten, A.F.A. en R. Hoppe, Voortgang, vooruitgang en achteruitgang in de beleidswetenschap, in: *Beleidswetenschap*, jrg. 20, 2006, nr. 4, pp. 34-72.

Korsten, A.F.A., A.F.M. Bertrand, P. de Jong en J.M.L.M. Soeters (red.), *Internationaal-vergelijkend onderzoek*, Vva, Den Haag, 1995.

Korsten, A.F.A., B.L. Becker en T. van Kraaij, *Samenwerking beproefd*, Sdu uitgevers, Den Haag/Arnhem, 2006.

Korsten, A.F.A., *Deugdelijk bestuur*, Open Universiteit, Heerlen, 2010a.

Korsten, A.F.A., Gedwongen vertrek van burgemeesters, in: *Justitiële Verkenningen*, jrg. 36, 2010b, nr. 3, pp. 10-31.

Korsten, A.F.A., K. Abma en J.M.L.R. Schutgens, *Bestuurskracht van gemeenten – Meten, vergelijken en beoordelen*, Eburon, 2007.

Korsten, A.F.A., *Lokaal bestuur in verandering*, Open Universiteit, Heerlen, maart 2009.

Korsten, A.F.A., M. Schoenmaker, e.a., *Onder burgemeesters*, Boom Lemma, Den Haag, 2013.

Korsten, A.F.A., *Orde en veiligheid in het publieke domein*, Open Universiteit, Heerlen, sept. 2009.

Korsten, A.F.A., P. de Jong en C.J.M. Breed, *Regeren met programma's – Interdepartementale kabinetsprogramma's van het kabinet-Balkenende IV: voortgang en samenwerking*, Boom Lemma, Den Haag, 2010.

Korsten, A.F.A., Visiteren van gemeentebesturen, in: *Bestuurswetenschappen*, jrg. 58, 2004b, nr. 4, pp. 305-324.

Korsten, A.F.A., *Wat is goed genoeg? - Benutting van onderzoek in overheidsbeleid*, Kobra, Amsterdam, 1983.

Kuhry, B. en J.J.J. Jonker, *Maten voor gemeenten 2009*, SCP, Den Haag, 2009.

Lane, J.E., *New Public Management*, Routledge, Londen, 2000.

Langeveld, J., H. Herbermann en K. van Dijken, Benchmark rioleringszorg. Resultaten uitvoering cluster 1 en doorontwikkeling van instrument, in: *Rioleringswetenschap*, jrg. 6, 2006, nr. 23, pp. 9-26.

Lapperre - van Dijk, C.H.M., *Benutting van benchmarkonderzoek bij gemeenten. Case studie naar de benutting van benchmark Bouw- en Woningtoezicht 2005 bij gemeenten*, Open Universiteit Nederland, Heerlen, 2007 (doctoraalscriptie).

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.

Lemaire, B., *Benchmark servicegerichtheid Nederlandse gemeenten 2012*, juni 2012.

Luts, M., W. van Dooren & G. Bouckaert, *Internationale rangschikkingen gerangschikt- Een meta-analyse van de publieke sector*, Steunpunt, Leuven, 2008.

Maas, M., H. Witjes en I. Zaat, *Benchmarking bij overheid en non-profitorganisaties*, Elsevier, Den Haag, 1998.

Martens, M., K.J. Groen en B. van der Wal, *Publiek management in 65 modellen*, Berenschot, Utrecht, 2002.

Mintzberg, H., *Organisatiestructuren*, Pearson Education, Rotterdam, 2006.

Mintzberg, H., Structure in 5's: A Synthesis of the Research on Organizations Design, in: *Management Science*, vol. 26, 1980, nr. 3, pp. 322-341.

Moynihan, D.P., *The Dynamics of Performance Management*, Georgetown University Press, Washington, 2008.

Nelissen, N.J.M. en P.J.M. de Goede, Benchmarking: vergelijken en verbeteren, in: *Bestuurskunde*, jrg. 8, 1999, nr. 2, pp. 54-70.

Niekerk, N. van, De overheid vergelijken met een BV is natuurlijk klinkklare onzin, in: *Elseviers Weekblad*. 1 febr. 1986.

Noordegraaf, M. e.a., *De ambtenaar als publiek ondernemer*, Coutinho, Bussum, 1995.

Noordegraaf, M. en J. Sterrenburg, Administratieve lasten voor publieke professionals: ficties en feiten, in: Jansen, Th., G. van den Brink en J. Kole (red.), *Beroepstrots*, Boom, Amsterdam, 2009, pp. 115-129.

Noordegraaf, M., K. Geuijen en A. Meier (red.), *Handboek publiek management*, Boom Lemma, Den Haag, 2013.

Osborne, D. & P. Plastrik, *Banishing Bureaucracy*, Addison-Wesley, Reading, 1997.

Osborne, D. & T. Gaebler, *Reinventing Government*, Addison-Wesley, Reading, 1992.

Peeters, H. en G. Verschraegen, Government by numbers: risico's verbonden aan de internationale benchmarking en ranking van pensioensystemen, in: *Beleid en Maatschappij*, jrg. 40, 2013, nr. 2, pp. 133-155.

Plasterk, R., De gekte van streefcijfers, in: *De Volkskrant*, 24 okt. 2002.

PricewaterhouseCoopers en Berenschot, *Benchmarkonderzoek thuiszorg biedt aanknopingspunten voor instellingen en overheid*, Utrecht, maart 1999.

Pwc, *Benchmark middelbaar beroepsonderwijs 2011*, Mbo Raad, Gouda, 2012.

Pwc, *Benchmark middelbaar beroepsonderwijs 2012- Bouwsteen studiesucces*, Mbo Raad, Gouda, 2013.

*Raad voor het openbaar bestuur, Presteren door leren*, Den Haag, 2002.

Rynck, F. de, J. Voets en W. Van Dooren, Het beoordelen van de prestaties van netwerken: aanzet tot een analysekader, in: *Bestuurswetenschappen*, 2010, nr. 1, pp. 35-55.

Schoenmaker, M., *Bestuurlijk gedonder- Onderzoek naar bestuurlijke probleemgemeenten in Nederland, 1998-2010*, Wolf Legal Publishers, Nijmegen, 2011 (diss.).

Schutgens, J.M.L.R., A.J.C. Maessen en A.F.A. Korsten, *Beoordeling van gemeenten- Financiële verdiepingsonderzoeken en bestuurskrachtmetingen van gemeenten*, Eburon, Delft, 2009.

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 vergelijkingskringen-methode, in: *De Overheidsmanager*, jrg. 7, oktober 2000, nr. 10.

Sociaal en Cultureel Planbureau, *Op weg met de Wmo - Evaluatie Wet maatschappelijke ondersteuning 2007-2009*, SCP, Den Haag, 2010.

Social and Cultural Planning Office of the Netherlands, *Public Sector Performance*, SCP, Den Haag, 2004.

Stroobants, J. en G. Bouckaert, *Lokale benchmarking – Conceptualisering, buitenlandse praktijken en verkennende oefeningen voor benchmarking van gemeentelijke dienstverlening via niet-parametrische grensmethoden*, KU Leuven, 2012.

Taphoorn, R. en F. van Rooij, Rijksbrede benchmark – Ook voor gemeenten!, in: *B&G*, oktober 2004, pp. 17-19.

Teunisse, P.B.W., M. de Kiewit en H. Boshove, *De gezonde ondernemende gemeente – De bedrijfsvoering van gemeenten in 2010*, VNG Uitgeverij, Den Haag, 2005.

Thiel, S. van, en F.L. Leeuw, De prestatieparadox in de publieke sector, in: *Beleidswetenschap*, 2003, nr. 2, pp. 369-404.

Thiel, S. van, Lokale verzelfstandiging: trends, motieven en resultaten van verzelfstandiging door gemeenten, in: *Beleidswetenschap*, jrg. 16, 2002, nr. 1, pp. 3-32.

Tije, S. ten, W. Pieterse, J. van Dijk en J. Jansen, *Balans in benchmarking*, Universiteit Twente, Enschede, 2010.

Tillema, S. en G.J. van Helden, *Van vergelijking naar verbetering? Onderzoek naar de actiebereidheid van waterschappen naar aanleiding van de bedrijfsvergelijking Zuiveringsbeheer 1999*, Groningen, 2003.

Tillema, S., Gebruik van benchmarkinformatie in de publieke sector, in: *MAB*, vol. 80, 2006, nr. 6, pp. 300-308.

Tits, M. van, Benchmarks, toepassing in de jeugdgezondheidszorg, in: Boekenooien, M. van, P. van Hoesel, Y. Prince en C. Verheijen (red.), *Methoden van beleidsonderzoekers: creatief en oplossingsgericht*, Lemma, Den Haag, 2009, pp. 323-333.

Tjeenk Willink, H.D., De overheid als supermarkt - dat moet fout lopen, in: *NRC*, 21 maart 2002.

Tops, P.W., A.F.A. Korsten en C.A.T. Schalken (red.), *De wethouder*, Vuga, Den Haag, 1994.

Twist, M.J.W. van, M.C. den Boer, B.P.A. van Mill en I. Geut, *Beelden van bestuur*, Berenschot, Utrecht, Uitgeverij Lemma, Utrecht, 2002.

Vall, M. van de, *Sociaal beleidsonderzoek*, Samsom, Alphen, 1980.

- Veldheer, V., Hoe ontevreden zijn burgers eigenlijk over de overheid?, in: SCP, *Het theorema van Thomas*, Den Haag, 2003, pp. 42-48.
- Veldkamp, C.M.H., *Beelden van kwaliteit - Normanalyse als bruikbaar hulpmiddel bij kwaliteitszorg in verpleeghuizen*, Universiteit Twente, Enschede, 1998 (diss.).
- Vierbergen, R.A. en A.J.M. Brouwer, Bestuurlijk relevante prestatiegegevens in de publieke sector, in: *Overheidsmanagement*, 2002, nr. 1, pp. 4-7.
- Visitatiecommissie Vlaamse steden (Commissie-Korsten-Reynaert/IDEA), *Syntheserapport Visitaties stedenfonds 2011 - Over het Stedenfonds en het samenspel tussen de Vlaamse overheid en haar centrumsteden*, IDEA Consult, nov. 2011 (op [www.thuisindestad.be](http://www.thuisindestad.be)).
- Vogelzang, P., Efficiency politie is moeilijk in cijfers te vangen, in: *De Volkskrant*, 6 april 2002.
- Vollaard, B.A., Prestatieprikkels door verantwoording, in: *ESB*, 25 juli 2003.
- Vriend, G.K. de en A. Timmerman, *Benchmarking*, Kluwer, Deventer, 1995.
- Vries, J. de, en J. van der Togt, *Benchmarking in 9 stappen*, Kluwer, Deventer, 1995.
- Vries, M.S. de, en C.J. Lako, Prestatie-indicatoren voor effectief en robuust gemeentelijk beleid, in: *Bestuurswetenschappen*, jrg. 63, 2009, nr. 6, pp. 73-93.
- Waalewijn, Ph. e.a., Benchmarking in Nederland: op zoek naar het ideale proces, in: *Holland Management Review*, 1996, nr. 51, pp. 57-69.
- Wassenaar, M.C. en A.J.W.M. Verhagen, *Financiën van decentrale overheden*, Sdu, Den Haag, 2006.
- Watson, G.H., *Strategisch benchmarken*, Scriptum, Schiedam, 1998.
- Wittebrood, K., De prestatieparadox van politie en justitie, in: SCP, *Averechtse effecten*, Den Haag, 2002, pp. 63-68.
- WRR, *Publieke zaken in de marktsamenleving*, Amsterdam University Press, Amsterdam, 2012.
- Zee, H. van der (red.), *Op weg naar een effectieve overheid*, Sdu uitgeverij, 2005.
- Zifcak, S., *New Managerialism*, Open University Press, Buckingham, 1994.