



Het gras bij de buren

Benchmarking de maat genomen

Arno F.A. Korsten (redactie)
Klaas Abma
Anne-Douwe van der Meer

Update 15 augustus 2018

Inhoudsopgave

1 Inleiding

2 Benchmarking: achtergronden en eerste typering

3 Benchmarking: kenmerken, vormen, doelen, grenzen

4 Voordelen en nadelen van benchmarking

5 Ervaringen met verschillende benchmarkingtrajecten: verdiepende impressies

6 Veronderstellingen achter benchmarking nader bezien

7 Uitkomsten van benchmarking: benchmarks centraal

8 Uitkomsten van partiële benchmarking: gemeenten centraal

9 Uitkomsten van integrale benchmarking: de beste woongemeenten

10 Uitkomsten van integrale benchmarking: de bestuurlijke staat van gemeenten

11 Bespreking van benchmarkrapportages: gebruiken, leren en verbeteren?

12 Valkuilen bij benchmarking

13 Enkele conclusies over de keizer en de kleren

14 Discussie: enkele stellingen over benchmarking

Literatuur

Bijlage 1 - KING: het Kwaliteits Instituut Nederlandse Gemeenten

Bijlage 2: Overzicht van een aantal gemeentelijke benchmarks

Bijlage 3: Scores van burgers op het item 'goed bestuur' van de gemeente in 'De Staat van de gemeente'

Over de auteurs

Overzicht van figuren

Fig.1: Ranglijst der ranglijsten: het gaafste land, 2016

Fig.2: Inventarisatie van benchmarking van gemeenten in de periode 2000 tot en met 2007

Fig.3: Voorbeeld van 'best practice'-informatie over aan cliënten bestede zorgtijd in verzorgingshuizen

Fig. 4: Soorten performance-indicatoren

Fig. 5: Weergave van een deel van de woonaantrekkelijkheidsmap van een gemeente in de Randstad

Fig. 6: Vormen van benchmarking (bm)

Fig. 7: Kenmerken van het productieproces bij zeven onderzochte instellingen voor thuiszorg vlg. Elias e. a. (1997)

Fig. 8: Efficiënties van verpleeghuizen vlg. SCP-onderzoek

Fig. 9: Positie van Enschede in de rangorde van steden op het vlak van duurzame mobiliteit (2018)

Fig. 10: Plaats van de gemeente Rotterdam in de 'Atlas voor gemeenten 2012'

Fig. 11: Selectie van gemeenten uit 'De beste gemeenten van 2013' naar rangorde en afzonderlijke scores

Fig. 12: De beste 100 gemeenten uit het onderzoek 'beste gemeenten in 2013' en gemeenten uit de kelder van de ranglijst (plaats 300 en hoger)

Fig.13: Plaats van steden boven 100.000 inw. op de rangorde in 'beste gemeenten van 2013'

Fig.14: Aandeel bewoners van gebieden met een zeer negatieve, negatieve of matige leefbaarheidsscore in meerdere jaren van een aantal gemeenten, in rangorde per 1998

Fig.15: Deelnemende gemeenten aan 'De Staat van de Gemeente' in de periode 2005 t/m 2007

Fig.16: Uitkomsten van 'De Staat van de Gemeente' in de vorm van scores op een schaal van 1-10 naar burgerrollen

Fig.17: Rapportcijfers op indicatoren voor alle gemeenten behorende bij de zes rollen

Fig.18: Ranglijst van gemeenten naar totaalscores op 'De Staat van de Gemeente'

Fig.19: Kwaliteit van gemeenten in rapportcijfers, naar gemeentegrootteklassen

Fig.20: Valkuilen bij gemeentelijk benchmarkonderzoek

Fig.21: Enkele methodische valkuilen bij gemeentelijk benchmarkonderzoek

Fig.22: Stellingen over benchmarking op een rij

Voorwoord

De titel van dit boek luidt *'Het gras bij de burens – Benchmarking de maat genomen'*. Dit boek wil ontmaskeren. Een Engelse spreuk luidt: het gras bij de burens is altijd groener. Is het gras bij de burens altijd groener? Functioneren andere gemeentelijke organisaties, politie-eenheden of ziekenhuizen beter dan de eigen gemeente, eigen politie-eenheid of het eigen ziekenhuis? Het boek wil de sluier van organisatievergelijking van prestaties aftrekken. Komt er bij vergelijking van organisaties altijd naar boven wie echt het beste presteert ('welk gras het meest groen is') en wat de geheimen zijn van de organisaties aan de top zo dat andere organisaties daarvan kunnen leren hoe ook die top te bereiken? Of is de werkelijkheid een andere? Komen we er misschien helemaal niet achter of en waarom het gras bij de burens groener is of lijkt? In elk geval heeft de langdurige zomer van 2018 geleerd dat het gras in de tuin bij heel veel inwoners met een eigen huis dor was. Met andere woorden, niemand had meer groen gras. Wat dan? Vergelijking lijkt dan bijna een dwaze handeling. Een kwestie van afwachten?

Recent kwam iemand in een Journaalrapportage voor die na de langdurige zomer van 2018 het bleke en verdorde gras met een verfstof groen had gespoten om het gras uit het voortuintje beter te doen ogen. Een truc van een makelaar die het huis wil verkopen en een leuke kleurenfoto wil voor de brochure of website? Je kunt het ook een privégrapje noemen of nieuws uit de journalistieke 'komkommertijd'.

Dit soort verschijnselen willen we uitbannen, en moeten we zelfs vermijden wil benchmarking een geloofwaardig middel zijn in het verkeer tussen organisaties, werknemers en burgers. Het doel van benchmarking is immers om echt de beter presterende organisatie te ontdekken en te bekijken of als je de top van de rangorde in beeld hebt, waarop de top het beste is en wat je moet doen om ook de top te bereiken.

Rangorde van sterren

Een alternatieve titel zou kunnen luiden: *'Ranking the Stars- Over rijkheid en leegheid van benchmarking'*. De argumentatie voor deze titel blijft dan vergelijkbaar met die voor de eigenlijke titel. Deze studie gaat namelijk over het *vergelijken* van prestaties en *rangordening* daarvan, zodat de *beste* praktijken komen bovendrijven en het *waarom* ervan duidelijk wordt. Zo ontstaat een oriëntatiepunt voor de toekomst. Benchmarking is heel interessant als concept. Het is nuttig te weten hoe organisaties het doen in vergelijkend opzicht. Wie is waarom de betere en is daarvan te leren? Dat verwijst naar rijkheid. Organisaties zouden van elkaar kunnen leren, heet het. Maar benchmarking is mogelijk ook een 'leeg concept' of overschat concept als mensen in organisaties gaan proberen vlug te scoren op taken of opdrachten waarop snel te scoren valt ('cherry picking') en de complexe taken laten liggen of enigszins verwaarlozen. Dat verschijnsel kan zich ook voordoen.

Deze tweede titel kwam bij ons naar voren toen enkele dagen terug Menzis een plan voor de sector van de geestelijke gezondheidszorg lanceerde. 'Resultaat is de norm voor vergoeding door Menzis' kopte De Volkskrant op 13 augustus 2018. Deze zorgverzekeraar zou de kosten voor behandeling van depressies en angsten blijven vergoeden, maar niet langer alleen kijken naar het aantal behandelingen. Menzis wilde ook het behandelresultaat in beeld krijgen. Vergelijking zou kunnen leiden tot een positieve prikkel voor organisaties uit de wereld van de psychiatrie en overige ggz-instellingen die sneller resultaat boeken en achterblijvende behandelaars misschien een zetje kunnen geven om een behandeling eerder of beter af te ronden. Deze maatregel moet leiden tot een meer doeltreffende therapie en op termijn lagere kosten. Want zo was de redenering: op termijn zullen de kosten voor de geestelijke gezondheidszorg nog meer de pan uitrijzen dan ze al deden en doen. Maar kan dit zo werken? Er kwam verzet vanuit de vereniging van psychiaters. Ze kruisten de degens met de representant van Menzis. Damiaan Denys zei vele bezwaren te kunnen noemen. Om voor- en nadelen te kunnen begrijpen, moet natuurlijk uitgetekend worden hoe het echt gaat met benchmarking. Wellicht zijn bepaalde nadelen te ondervangen.

Dit boek behandelt deze actuele thematiek van organisatievergelijking om te komen tot meer effectiviteit in taakuitvoering en doelmatigheid, met accent op benchmarking van gemeenten.

Derde invalshoek

Er is nog een derde invalshoek mogelijk. Die was oorspronkelijk aanleiding om aan dit project te beginnen. Dat zit zo. *'New Public Management'* (NPM) is een geheel van samenhangende managementconcepten dat in de jaren tachtig van de vorige eeuw en later is toegepast door (semi-) overheidsbesturen om te moderniseren door middel van hervormingen. Deze denkrichting wil(de) het succesvolle denken uit de private sector toepassen op de publieke sector. De basiswaarden van NPM waren prestatiegerichtheid en outputsturing, klantgerichtheid, bedrijfsmatigheid, doelmatigheid en effectiviteit. Een prototype van NPM in de publieke sector was verzelfstandiging in de vorm van de vorming van agentschappen (interne verzelfstandiging) en zelfstandige bestuursorganen (externe verzelfstandiging). Aan leidinggevend van deze organisaties, die op afstand van de overheid functioneerden, werd meer verantwoordelijkheid en handelingsvrijheid gegeven. Daarbij viel het woord beleidsvrijheid en 'contractmanagement'. De verwachting was dat verzelfstandigde organisaties beter zouden presteren door meer prestatiegerichtheid, bedrijfsmatigheid, klantgerichtheid en wendbaarheid.

Na verloop van tijd zijn niet alleen sterkten maar ook enkele zwakheden en neveneffecten zichtbaar geworden, waardoor het bureaucratisch overheidsmodel weer aan aanhang won. Zo zou een te grote nadruk op efficiency ten koste kunnen gaan van andere waarden, zoals steun voor beleid en rechtvaardigheid. Legitimiteit, solidariteit en rechtvaardigheid zijn waarden waarmee marktdenkers en verdedigers van ondernemingsgewijze productie niet of nauwelijks rekening houden. Is het NPM daarmee volledig overbodig en benchmarking misschien ook? Een overheid is wel voor een moment te zien als een bedrijf ('laten we eens doen alsof het een bedrijf is'), maar feitelijk is een overheid geen bedrijf dat op winst gericht is. Het accent op het meten van producten en prestaties kan bovendien ongewenste gedragseffecten hebben. Een negatieve consequentie is bijvoorbeeld dat de wegenpolitie nepcontroles uitvoert om makkelijk prestatiedoelstellingen te bereiken en de toegevoegde waarde te bewijzen. Een andere mogelijkheid is een verwijzing naar dalende criminaliteit zonder te vermelden dat burgers diefstallen niet of minder bij de politie melden. Het is makkelijk scoren als je de dalende aangiftebereidheid van burgers buiten beeld houdt. NPM roept *cijfermanipulatie* op en dat is een zwak punt. NPM heeft daardoor tegenwoordig minder aanhangers.

Sommigen benadrukken 'governance' als alternatief concept. Dat concept houdt in dat een overheidsorganisatie in netwerken functioneert van organisaties die deels of bijna geheel van elkaar afhankelijk zijn om bepaalde doelen te bereiken. Denk aan de veiligheidsketen waarbij organisaties op elkaar inhaken. Deze organisaties moeten beelden die ze hebben van bepaalde situaties delen en doelen en middelen verknopen om zo gezamenlijk verder te komen. 'Samen' is beter dan 'alleen'. Een overheid is voor tal van taken geen centrale organisatie die alles zelf bedenkt en realiseert, maar een organisatie die andere organisaties, burgers en groepen nodig heeft en zich verknoot met de samenleving, niet slecht om meer draagvlak te verkrijgen maar juist om effectief te zijn.

'Governance' is dus een nieuwe filosofie geworden in een wereld van afhankelijkheden, waarin weinig organisaties alle doorzettingsmacht in eigen hand hebben. Deze netwerkfilosofie is nuttig maar ook weer eenzijdig omdat er ook taken bestaan die een overheid wel degelijk zelfstandig kan en moet uitvoeren. Denk aan het cluster van burgerzaken, zoals de levering van paspoorten.

Sommigen vullen het concept NPM en 'governance' verder aan door te spreken over het '*Post-New Public Management*'-tijdperk, of van de '*Digital-era Governance*'. Ze benadrukken de opkomst van moderne informatie- en communicatietechnologie, bij voorbeeld voor het uit de grond stampen van nieuwe participatiemogelijkheden voor burgers in het kader van de deliberatieve democratie.

Vanuit NPM is benchmarking benadrukt maar netwerkdenkers zijn niet per se tegen. Benchmarking of organisatievergelijking die is gericht op rangordening, het zoeken naar goede praktijken en verbetering heeft de tand van de tijd doorstaan. Wie voorstander is van het vergelijken van producten en processen tussen gemeenten, ziekenhuizen of universiteiten hanteert het uitgangspunt dat van andere organisaties te leren is. En misschien is zelfs het uitgangspunt dat het 'gras bij de burens wel eens groener kan zijn dan bij onszelf'.

Burgers vergelijken graag

'Het gras bij de burens is altijd groener'. Dat is het uitgangspunt van dit boek. Deze wijsheid impliceert dat burgers graag vergelijken en kijken naar hoe het beter kan dan bij henzelf. 'Gras' staat voor een hele boel producten, processen of benaderingen die elders beter kunnen zijn. Zo zijn de (medische) prestaties van een ziekenhuis te vergelijken met die van andere ziekenhuizen. En is vergelijking en rangordening mogelijk van prestaties van woningcorporaties, politie-eenheden, opleidingen of hele universiteiten, hogescholen en gemeentelijke organisaties. Noem maar op.

Wat benchmarking is

Wie het heeft over vergelijken, rangordenen en verbeteren heeft het in feite over organisatievergelijking. Vroeger heette deze vergelijking bedrijfsvergelijking, tegenwoordig benchmarking. Dit boek gaat over benchmarking. Benchmarking is een benadering voor prestatiemeting, die erop gericht is om *systematisch* op zoek te gaan naar *het betere* bij de *eigen* organisatie en bij *andere* organisaties. Benchmarking omvat inventarisatie en vergelijking van prestaties en zo mogelijk effecten, rangordening naar kwaliteit, verklaring van verschillen tussen prestaties, het zoeken van 'beste praktijken' en het zetten van stappen naar verbetering.

Is dat wel een goede aanpak? Waarom? Wat wordt daarmee beoogd? Is die aanpak overal op toepasbaar of worden bepaalde opgaven en taken niet belicht? Wat zijn de voor- en nadelen van benchmarking? Welke valkuilen dreigen? Hoe kun je de praktijk van benchmarking beoordelen? Is benchmarking een middel om te zorgen voor doorbraken in bestaande praktijken die verbetering behoeven en is het zelfs een kantelpunt?

Dat zijn vragen die in dit boek aan de orde komen. Het boek bevat veel voorbeelden uit de dagelijkse praktijk. Speciale aandacht wordt besteed aan benchmarking van gemeenten.

Schimmert/Wommels/Harderwijk, 15 augustus 2018

Arno Korsten, Klaas Abma en Anne-Douwe van der Meer

1 Inleiding

1.1 De unieke prestatie

Marketingmensen die werkzaam zijn bij een organisatie benadrukken graag hoe bijzonder hun organisatie is en de geleverde prestaties wel niet zijn. Ze sluiten aan op de strategisch koers die de leidinggevenden kozen op basis van de eigen identiteit en een vergelijking met de concurrentie.

Bij veel van wat wetenschappers en anderen onderzoeken, staat ook het bijzondere geval centraal. Onderzoekers maken een analyse van bijvoorbeeld de Deltawerken, van 'de Rotterdamse aanpak van sociale vernieuwing' of 'het lokaal veiligheidsbeleid in Rotterdam sinds de opmars van Pim Fortuyn in 2002'. Die keuzen worden verdedigd vanuit het unieke. Dat is het - om bij Rotterdam te blijven - door het gemeentebestuur gekoesterde en door onderzoekers als veronderstelling overgenomen geloof in de uniciteit en onvergelykbaarheid van het gekozen object. Denk aan argumenten als 'deze gemeente met zijn internationale haven', 'deze stad met zijn bijzondere ligging en zijn mentaliteit van geen woorden maar daden', 'deze stad als het echte experimenteergebied voor sociale innovaties'. Wie één geval onderzoekt, benadrukt het anders-zijn van de sociale of bestuurlijke context. Het onderzochte project of traject wordt gezien als *bijzonder*.

Dit exceptionalisme zien we ook terug in historische studies van Nederland en in andere studies. Nederland zou een uniek land zijn waar 'de eeuwige strijd tegen het water' is gevoerd, waar het zuilenstelsel vorm kreeg en ontzuiling optrad en waar het bestuurlijk polderen (consensusvorming gesymboliseerd in onder meer het 'akkoord van Wassenaar' van werkgevers en werknemers) is uitgevonden en gekoesterd. We zien het terug in studies van de verzorgingsstaat met de ontwikkeling van een eigen stelsel voor sociale zekerheid, dat nergens in de wereld zo ontwikkeld is.

Het bijzondere, zeldzame en unieke blijkt ook uit de typering '*Nederland is een gidsland*', alsof de regeringen van andere Europese landen niet het goede ten voorbeeld stellen. Nog wat. Sinds minister Max van der Stoep (kabinet-Den Uyl) lopen opeenvolgende ministers van Buitenlandse Zaken voorop in het wijzen van de weg naar '*het ware en het goede*' op het vlak van waarden en normen die politiek van belang zijn. Te noemen zijn de vrijheid van meningsuiting, de erkenning van mensenrechten, het toestaan van het homohuwelijk, gelijkberechtiging van mannen en vrouwen bij verkiezingen, beloning en op ander vlak. Regeringen van heel veel andere landen, zoals China, Noord- en Zuid-Korea en Indonesië worden in het diplomatiek verkeer 'geattendeerd' of zo nodig tot de orde geroepen. Buitenlands beleid is in belangrijke mate mensenrechtenpolitiek van 'met het vingertje wijzen' of eufemistischer 'aandacht vragen voor ...'.

Voor wie wil vergelijken, is het exceptionalisme echter al gauw *een mythe*. Wie wil vergelijken, gaat op zoek naar de evenknie, naar de parallel. Wie vergelijkt, ziet discussies over bijvoorbeeld schaalvergroting en gemeentelijke herindeling in tal van landen in West-Europa. Wie de Nederlandse politiek bestudeert, komt uit bij de politicoloog Arend Lijphart. Die heeft ons getrakteerd op vergelijking van politieke stelsels, waaronder '*consociational politics*' zoals het verzuilde stelsel met pacificatie aan de top tussen de politieke representanten van de zuilen. Er valt wel degelijk veel te vergelijken. Benchmarking is expliciet gericht op vergelijken maar er is meer zoals nog zal blijken.

1.2 Altijd en overal vergelijken

Wie dieper graaft, ontdekt dat mensen wel het unieke willen benadrukken maar ook graag vergelijken, ordenen en competities aangaan. Vergelijken is bij nader inzien evenzeer als de beklemtoning van het unieke overal zichtbaar. En bij vergelijken is ook de 'ranglijst', de 'rangorde' of het 'klassement' al snel in beeld. Louis van Gaal zei ooit als gevierde voetbaltrainer van Bayern München, staand op een balkon in München tegen enthousiaste fans: '*Wisst ihr es? Wir sind die Besten von Deutschland*'. Und '*wir sind auch die Besten von Europa*'.

De Rotterdamse profvoetbalclub Feyenoord is nationaal uniek maar vergelijkingen maken met giganten als PSV of Ajax doen veel mensen. De sportfanaat met interesse voor voetbal zal wekelijks

de uitslagen van de voetbalcompetities in het amateurvoetbal of betaald voetbal die hem of haar boeien, bijhouden en de eventuele opmars van de favoriete club op de ranglijst volgen. Deze liefhebber zal vermoedelijk ook een mening hebben over waarom een bepaalde ploeg aan kop gaat en een eigen verklaring hebben voor de plaats van de eigen club op de ranglijst, ook al is die verklaring speculatief. Ook sportdirecteuren van voetbalclubs vergelijken: welke talent aan te trekken en welke speler te verkopen?

Cynisme

Als vergeleken wordt, is de ironie of het cynisme niet ver weg. Hoeveel wedstrijden had de (ontslagen) trainer Danny Blind van het nationaal voetbalelftal na het in 2017 verloren duel tegen Bulgarije eigenlijk gewonnen, hoeveel duels verloren en hoeveel waren geëindigd in een gelijk spel?

Stadionmanagers zullen oog hebben voor de kwaliteit van de sportvelden. Zij zullen willen weten hoe lang de grasmat in de Amsterdam Arena het deze keer houdt. Ook hier is vergelijken en beoordelen aan de orde. Sinds de opening van de Amsterdam Arena in 1996 zijn er al meer dan zestig nieuwe grasmaten in de het stadion gelegd (volgens de stand begin 2017). Vooral in de eerste tien jaar volgden de verschillende groene maten elkaar in hoog tempo op. In 2006 stond de teller al op 45 nieuwe grasvelden. De laatste jaren leken de grootste problemen met de grasmat achter de rug. Was geleerd van aanleg van sportvelden zoals die elders had plaats gevonden? Het kan verkeren. In het seizoen 2011-2012 werd het veld in de Arena zelfs verkozen tot *beste* grasmat van de eredivisie. Dus het lukte toen zelfs om in de Arena 'het beste speelveld' neer te leggen. Wat is slecht en wat is goed?

Wie vergelijkt er nog meer?

De senior medewerker zorginkoop van een zorgverzekeraar zal ook vergelijken en beoordelen, bij voorbeeld wat de behandelingen en verrichtingen van specialisten in ziekenhuizen en zelfstandige behandelcentra (zbc's) kosten. Hij of zij zal als het even kan aansturen op een goede behandeling van een patiënt met een duidelijk medisch probleem, in een bepaald ziekenhuis of behandelcentrum tegen een - vergelijkend gezien - lage prijs.

De reiziger vergelijkt ook al gauw. Multatuli stelde ooit:

'Ik heb veel landen bezocht en bijverde mij overal acht te slaan op de publieke zaak. Welnu, ik verklaar nergens zulke totale abstinente van plichtsbesef, nergens zo'n walgelijke onbekwaamheid te hebben aangetroffen als bij 't stadsbestuur van Amsterdam.'

Vergelijken in de wetenschap

Vergelijken komt, zoals al gesteld is, ook in de wetenschap voor. Wetenschappers kijken of een vanuit de literatuur of andere bronnen gefundeerde hypothese te verwerpen of bevestigen is. We kennen zelfs specifiek internationaal-vergelijkend onderzoek naar allerlei onderwerpen, zoals hoeveel ambtenaren een land telt om zo de omvang van de overheidsbureaucratie in kaart te brengen (Korsten, Bertrand e.a., 1995).

Bestaan er ook theorieën met vergelijking als kern? Uiteraard. Een voorbeeld. Een socioloog constateert dat kleine of grotere groepen burgers vaak de eigen situatie met die van andere burgers en groepen vergelijken: wie heeft het in welk opzicht beter? De beroemde sociologen Robert King Merton en Samuel Stouffer zijn bekend geworden met een theorie hierover, *de relatieve deprivatietheorie*. Relatieve deprivatie is een subjectieve tevredenheid of ontevredenheid die *niet* wordt veroorzaakt door een objectieve situatie, maar door de relatieve positie van de een ten opzichte van de situatie van een (relevante) ander. Mensen zetten volgens deze theorie voortdurend de eigen situatie af tegen die van anderen. Ze willen bepaalde zaken die anderen bezitten binnen hun bereik brengen.

De positie van Amerikaanse militairen in de Tweede Wereldoorlog en in de Vietnam-oorlog is wel met deze theorie in de hand bekeken. Amerikaanse militairen die rond 1961-1968 (onder de

presidenten John F. Kennedy en Lyndon B. Johnson) in de jungle vchten tegen de guerrillagroepen van de Vietcong vergeleken zich met andere militairen die meer vrijheid en rust hadden in Saigon en andere steden in Vietnam. Ze vergeleken zich *niet* met de positie van militairen in de Verenigde Staten, terwijl die toch juist genoten van een veel meer comfortabele positie dan bijna alle strijdende Amerikaanse militairen in Vietnam. Blijkbaar vergapen burgers zich niet aan onbereikbare idealen of omstandigheden.

Het is ook zo dat de burgers zich in een woonwijk vaak vergelijken met de burens uit dezelfde of nabijgelegen straat, maar niet met onbereikbare miljonairs uit een stad of dorp verderop. Zo zien we dat de groep militairen in het thuisland en de meest welvarende groep miljonairs niet het referentiepunt waren. De mens vergelijkt zich blijkbaar niet graag met het onbereikbare. Verderop zal blijken dat het zoeken naar 'de beste praktijk' voor organisaties ook vaak een brug te ver is, een onbereikbaar ideaal.

Vergelijken van organisaties

De reiziger vergelijkt ook al gauw. Daarbij kan durf om te oordelen niet ontzegd worden. Multatuli stelde ooit: *'Ik heb veel landen bezocht en bejverde mij overal acht te slaan op de publieke zaak. Welnu, ik verklaar nergens zulke totale abstinentie van plichtsbesef, nergens zo 'n walgelijke onbekwaamheid te hebben aangetroffen als bij 't stadsbestuur van Amsterdam.'*

Vergelijken zit in de mens en dus vergelijken we ook organisaties. Op een feestje wordt wel gezegd: 'hoe gaat het in jouw bedrijf en in jouw gemeente?' Bedrijven bekijken hun positie en hun marktaandeel zelfs nadrukkelijk te midden van de concurrentie. Dan valt al gauw het woord 'concurrentiepositie'.

Als vergelijken geschiedt in sportcompetities dan spreken we van een klassement, klassering of rangorde. Dit boek gaat niet over voetbalranglijsten, niet over internationale ranglijsten van beste voetballers (FIFA Ballon d'Or), niet over een rangorde van wielrenners die de meeste punten verzamelden in erkende koersen (UCI World Ranking), en evenmin over een zelfde type rangorde onder tennissers (WTA-ranglijst), maar wel over het vergelijken van organisaties of delen daarvan in de vorm van organisatie-eenheden, producten of processen. Denk wel aan het vergelijken van zelfde studierichtingen die door verschillende universiteiten of hogescholen worden aangeboden. Hoe tevreden zijn studenten over aspecten van bijvoorbeeld de studie geologie in Amsterdam, Utrecht of een andere universiteitsstad? Hoe lang doen studenten gemiddeld in de ene stad over hun studie in vergelijking met dezelfde studie in andere steden en wat is dus het studierendement? Zijn de verschillen terug te voeren op selectie aan de poort, de organisatie van de studie, de studentbegeleiding, de kwaliteit van docenten en/of het didactisch studieconcept dat geldt?

Wie prestaties wil meten van organisaties, ze wil vergelijken en rangordenen, en daarna wil verklaren welke organisatie, opleiding of andere eenheid van handelen waarom op welk vlak de beste is en wil aangeven wat nodig is om de beste of in elk geval een betere te worden, doet aan *benchmarking*. In dit boek gaan we nader in op benchmarking, met als doel om te *verbeteren*.

1.3 Hoe Nederland op ranglijsten voor komt

Benchmarking is in feite nooit ver weg. Mark Rutte had het er als premier van Nederland regelmatig over, zowel op persconferenties als in de Tweede Kamer. Zonder het woord benchmarking overigens te noemen. Dan zegt hij dat Nederland zo'n '*gaafland*' is omdat Nederland in tal van 'rijtjes' hoog scoort en positief in de top staat in rangordes van de Europese Unie, de Wereldbank, het Internationaal Monetair Fonds (IMF), de Organisatie voor Economische Samenwerking – afgekort de OECD of OESO -, of Transparency International (met zijn corruptie-index). Anderen noemen Nederland een paradijs. En inderdaad, de champagne kan op tafel. Nederland deed en doet het internationaal op heel veel gebieden goed of zelfs excellent, zoals op een geluksindex. De

Nederlandse economie scoort goed in vergelijking met andere landen in de wereld. Sommigen zeggen zelfs: 'Nederland is het beste land ter wereld'.

Het is mogelijk aparte benchmarks van landen te nemen en te bespreken, maar het is ook mogelijk een aantal ranglijsten over elkaar te leggen en dan te bekijken welke beeld oprijst uit de 'ranglijst van ranglijsten'. We doen dat laatste hier, maar introduceren eerst kort enkele lijsten (zie ook FD, 3 juni 2016).

Het onderzoeksinstituut IMD stelt ieder jaar een 'World Competitiveness Scoreboard' op. Zo ook in 2016. Op basis van honderden indicatoren wordt de concurrentiekracht van landen met zoveel mogelijk harde cijfers over koopkracht en uitgaven voor onderzoek en ontwikkeling (R&D) in kaart gebracht. Nederland kwam in de recente editie op de achtste plaats van boven terecht, een zeer goede score.

Het World Economic Forum (WEF) brengt concurrentiekracht op een iets andere wijze in kaart ten behoeve van de 'Global Competitiveness Index', namelijk met meer aandacht voor subjectieve gegevens, zoals meningen van ondernemers. In deze tweede rangorde, de WEF-rangorde springt Nederland er met plaats vijf in de top vijf opnieuw heel goed uit. Zwitserland voert deze ranglijst uit 2016 aan, gevolgd door Singapore, de Verenigde Staten en Duitsland. Waar is Italië en Frankrijk?

Een derde meetinstrument is de 'Global Innovation Index', die jaarlijks door het bekende centrum voor managementstudies INSEAD, gevestigd in het Franse Fontainebleau, wordt berekend. De index meet de innovatiekracht van landen. Op deze lijst moet Nederland alleen landen als Zwitserland, het Verenigd Koninkrijk en Zweden voor zich dulden. Nederland is hier vierde.

Over innovatie gesproken. Volgens de jaarcijfers van het Europese Octrooibureau (EOB) staat Nederland in de wereldranglijst van krachtige innovatie- en technologie landen gelet op patentaanvragen per hoofd van de bevolking, in 2016 op de tweede plaats. Alleen Zwitserland staat in die ranglijst hoger. Na Nederland volgt Zweden, Denemarken, Finland, Duitsland, Oostenrijk, België, Japan, Frankrijk. In absolute aantallen staat Nederland op de zevende plaats in de wereld met 6889 aangevraagde octrooien in 2016, net achter China. Wie zorgt voor die octrooien? Onder meer bedrijven als de chemiereus DSM en Philips. Philips is het bedrijf met wereldwijd in 2015 en 2016 de meeste octrooiaanvragen. De sectoren waarin de meeste Europese octrooien worden aangevraagd zijn medische technologie en digitale communicatie- en computertechnologie. Uitgesplitst naar regio binnen Nederland scoort Brabant goed, want de helft van de aanvragen komen uit die provincie; deze provincie wordt gevolgd door Zuid-Holland en Limburg. Op de lijst van alle Europese regio staat de provincie Noord-Brabant vijfde. Conclusie: ook op dit vlak worden vergelijkingen gemaakt.

Zo zijn er meer mondiale indexen, zoals de 'Human Development Index' van UNDP. Deze index van een organisatie van de Verenigde Naties, brengt onder meer mondiaal de levensverwachting van burgers in kaart, het onderwijsniveau en het inkomensniveau van burgers. Op deze index komt Noorwegen op plaats een, Denemarken op vier, Nederland op de vijfde plaats terwijl Zweden op plaats dertien staat.

Aangezien de economie ten dienste staat van het levensgeluk van mensen is ook het 'World Happiness Report' van belang. De economen Richard Layard en Jeffrey Sachs hebben gewerkt aan het in kaart brengen van het levensgeluk van burgers. Ook anderen waren op dit gebied actief. Denk in dit verband aan mogelijkheden voor burgers om zelfstandig keuzen te maken in hun leven. In deze ranking uit 2016 staat Denemarken op plaats één, Zwitserland nu op plek twee, Noorwegen op vier, Canada op zes en Nederland op zeven. Een land als Duitsland komt op de veertiende plaats en de V.S. staan op negentien (fig. 1).

Andere geluksindexen verwijzen onder meer naar solidariteit onder burgers en buurtzorginitiatieven. In dit soort rijtjes scoren landen uit Noord- en Noordwest-Europa traditioneel hoog. Denk aan landen als Noorwegen, Denemarken, Zweden, Finland, IJsland, Nederland, Zwitserland, Canada. Een land als Polen komt hier niet bij de eerste vijftig voor. Voor meer informatie over definities van geluk, operationalisatie in indicatoren en landen scores kan men terecht in het werk van Ruut Veenhoven.

Het WHR 2017 laat zien dat Noorwegen aan de top staat 'as overall happiest country in the world'. 'All the top ten countries have high scores in the six categories'. Bijna alle Afrikaanse landen bungelen onderaan.

Fig.1: Ranglijst der ranglijsten: het gaafste land, juni 2016

Land	Rang	IMD	WEF	Insead	UNDP	Layard & Sachs	Totaal: optelsom horizontaal
Meetinstrument		WCS	GCI	GII	HDI	WHR	
Zwitserland	1	2	1	1	3	2	9
Nederland	2	8	5	4	5	7	29
Verenigde Staten	3	3	3	5	8	13	32
Denemarken	4	6	12	10	4	1	33
Zweden	5	5	9	3	13	10	40
Singapore	6	4	2	7	11	19	43
Noorwegen	7	9	11	20	1	4	45
Duitsland	8	12	4	12	6	14	48
Canada	9	10	13	16	9	6	54
Ierland	10	7	24	8	6	16	61
Luxemburg	11	11	20	9	18	17	75
Hongkong	12	1	7	11	12	75	106
WCS =	World Competitiveness Scoreboard						
GCI =	Global Competitiveness Scoreboard						
GII =	Global Innovation Index						
HDI =	Human Development Index						
WHR =	World Happiness Report 2016						

Het Centraal Bureau voor de Statistiek (CBS) beschikt over cijfers die indicatief zijn voor het geluksgevoel onder burgers. Vorig jaar noemde 88 procent van de Nederlandse volwassenen zich gelukkig. Het percentage is sinds 2013, toen de metingen begonnen, ongeveer gelijk. Een kleine drie procent is ongelukkig en de rest zegt gelukkig noch ongelukkig te zijn. Vooral gezondheid speelt volgens het CBS een belangrijke rol bij geluk. Van degenen die hun gezondheid beoordelen als slecht of zeer slecht is maar 58 procent gelukkig. Bij mensen die wel gezond zijn is dat 94 procent. Ook Nederlanders met een partner (92 procent gelukkig) zijn doorgaans gelukkiger dan singles. Bij mensen die hun partner hebben verloren door overlijden noemt 75 procent zich gelukkig, bij mensen die gescheiden zijn is dat percentage 82.

Nederland staat op de zesde plaats van de lijst van gelukkigste landen ter wereld. We zijn daarmee een plaatsje gestegen, ten koste van Canada. Het gelukkigst zijn de Noren, staat in het *World Happiness Report 2017*. Noorwegen lost Denemarken (2) af, dat de afgelopen jaren vaak op één stond. In de top-10 staan alleen maar westerse landen, met na Noorwegen en Denemarken respectievelijk IJsland, Zwitserland, Finland, Nederland, Canada, Nieuw-Zeeland, Australië en Zweden. In het rapport wordt geluk gedefinieerd aan de hand van factoren als: gelijke verdeling van welvaart, sociale verbondenheid, vertrouwen in de samenleving en de overheid, kwaliteit van de zorg en vrijheid.

Het rapport is opgesteld door het Sustainable Development Solutions Network, een initiatief van de Verenigde Naties dat regeringen wil stimuleren om het belang van hun burgers voorop te stellen.

Over Noorwegen staat in het rapport dat het land niet dankzij maar ondanks de olie-inkomsten zo gelukkig is. Er wordt op gewezen dat Oslo de inkomsten al jarenlang investeert voor de toekomst en niet direct uitgeeft. De lage olieprijs heeft volgens het rapport dan ook weinig effect op de Noren. De Verenigde Staten staan op plaats 14 en scoren negatief vanwege toenemende corruptie.

China is niet gelukkiger dan 25 jaar geleden, hoewel het sindsdien een spectaculaire economische groei heeft doorgemaakt. De onderzoekers wijzen op de toenemende werkloosheid en het ontbreken van sociale voorzieningen. Het slechtst af zijn landen in Afrika, met Zuid-Sudan, Liberia, Guinee, Togo, Rwanda, Tanzania, Burundi en als laatste in de rij de Centraal-Afrikaanse Republiek.

Als al deze indexen samen worden genomen en de plaatsen worden opgeteld, dan moet het land met het laagste totaal beschouwd worden als het beste land in de wereld. Dat blijkt Zwitserland te zijn met Nederland als goede tweede, de Verenigde Staten op drie en Denemarken op vier en Zweden op vijf. Vier Europese landen behoren wereldwijd dus tot de beste vier.

Die goede kwaliteit in landenvergelijkend opzicht is niet beperkt tot economie, innovatie, concurrentiekracht en geluksgevoel. Daarom kunnen nog enkele andere meetinstrumenten tegen deze bevindingen aan gelegd worden. Bekijk het *World Values Survey* (WVS).

Een ouder '*World Development Report*' van de Wereldbank heeft aandacht voor vijf kerntaken van de staat: het verschaffen van een juridische fundering aan de maatschappij; het onderhouden van een helder en consequent beleid en economische stabiliteit; het investeren in de basale infrastructuur; het beschermen van de zwakken in de samenleving. In 1997 voerden 69 landen deze vijf kerntaken matig tot slecht uit (NRC, 260697). Een overheid moet proberen op dit gebied doelgericht en effectief te zijn.

Nederland scoort ook hier redelijk goed en weet die positie later te consolideren. In rechtsstatelijk opzicht scoort Nederland op de zevende plaats en qua kwaliteit van de regelgeving op de tiende. Nederland is volgens allerlei indicatoren een gaaf land met betrekking tot de rechtsstaat en de kwaliteit van de regelgeving; het onderwijssysteem; de kwaliteit van natuurbehoud en van de voedselproductie; het stelsel van sociale zekerheid; het pensioenstelsel; de gezondheidszorg en ouderenzorg; het omgaan met laagland en de waterhuishouding.

Het rapport van de Wereldbank uit 2016 handelt over de enorme opmars van digitale technologie op dataverkeer, communicatie, handel en economische groei. Hoe staat het in elk land met de beschikbaarheid en toegankelijkheid van internet, het gebruik en non-gebruik in arbeid en vrijetijd en de veiligheidsrisico's. De aanname is dat hoe groter en omvangrijker het gebruik is, hoe groter de bijdrage is aan de arbeidsproductiviteit per hoofd van de bevolking en economische ontwikkeling. Het rapport laat onder meer door grafieken en kaarten zien dat er in de wereld grote verschillen zijn tussen landen in de mate van digitalisering en de invloed daarvan op e-commerce, handel, innovatie, doelmatigheid.

Het rapport zegt ook: 'The internet can make governments more capable and responsive'. Maar er zijn ook schaduwzijden. Internet leidt tot ongelijkheid tussen en binnen landen in gebruik van mobiele telefonie. En leidt tot polarisatie op de arbeidsmarkt omdat bepaalde groepen hun werk zien veranderen, niet meekunnen of hun functie overbodig wordt. De Wereldbank richt zich in het rapport vooral op wat vroeger heetten 'derde wereld'-landen of 'ontwikkelingslanden'. Nederland kent een sterke positie in landenvergelijkend perspectief maar is geen voorloper. In de rijtjes van het rapport komt Nederland over het algemeen na de eerste tien. Ook hier heeft Nederland veel weg van Duitsland.

Als Nederland goed scoort op de kwaliteit van de rechtsstaat mag worden aangenomen dat de corruptie gering is. Is dat zo? Corruptie is heel moeilijk te meten. Transparency International meet desondanks sinds een aantal jaren (start in 1995) de corruptie met de '*Corruption index*'. Die is vooral gericht op corruptie bij de overheid. Emissieschandalen bij dieselauto's zijn niet meegeteld.



De index is onder meer gebaseerd op oordelen van experts en opinieonderzoek. Daarom wordt het ook wel de *'Corruption Perception Index'* genoemd. Over vier jaren (2012 tot en met 2015) gemeten scoort Denemarken als beste land met de minste corruptie in verge lijking met 175 andere landen in de wereld. Denemarken wordt gevolgd door het eveneens goed scorende Finland (op plaats twee), Zweden (op plaats drie), Nieuw-Zeeland (vier), Nederland op plaats vijf, Noorwegen (zes) en Zwitserland (zeven), Singapore (acht), Canada (negen), Duitsland (tien). De bekende landen staan weer haantje de voorste. Japan staat pas op plaats 20, Frankrijk op 25, Portugal op 29, Spanje op 36, Griekenland op 58, Turkije op 68, Bulgarije op 69, Marokko op 93, Niger 103, Kameroen 131, Oekraïne op 136, Nigeria 138, Zimbabwe 153, Jemen op 158, Libië 163, Somalië 168.

Corruptie komt in hele grote delen van de wereld nog voor. De *'failed states'* als Oekraïne, Soedan, Nigeria, Somalië springen er negatief uit. Ze zijn herkenbaar. Hun hoge plaats staat voor veel corruptie. Wie verbleven heeft in tal van Afrikaanse en Aziatische landen en in het bijzonder in *'failed states'* als Jemen en Somalië weet dat Nederland inderdaad een gaaf land is. Nederland is in de periode 2000 tot en met 2016 steeds bij de eerste tien te vinden. Ook nu weer scoort Nederland betrekkelijk goed. België zit in die periode constant voorbij plek tien.

Het Nederlandse volk is – zo gezien - niet zonder reden betrekkelijk optimistisch; het pessimisme is beperkt volgens studies van het Sociaal en Cultureel Planbureau (SCP). De natie behoort niet alleen tot de gelukkigste volken in de wereld, het is bovendien een *high-trust society*. De meeste burgers vertrouwen elkaar en gaan mede daarom makkelijk transacties met elkaar aan.

Er is links en rechts desondanks nog werk aan de winkel. Nederland is nooit af. Het Sociaal en Cultureel Planbureau constateerde in de periode 2010-2017 dat tal van burgers de volgende houding hebben: *'met mij gaat het goed maar met ons wat minder'*. Dat is wat overdreven, uitgaande van de internationale scores. Maar toch, waar denkt men aan? Dan wordt verwezen naar politiek niet of zwak verwerkte problemen zoals soevereiniteitsverlies aan de Europese Unie; de te gebrekkige ouderenzorg (handen aan het bed in verpleeghuizen); de migratiestroom; de pensioenen (verhoogde AOW-leeftijd) en de waarden en normen. Concrete problemen worden ook benoemd, zoals de veelplegers in grote steden die overlast veroorzaken (een zorgpunt voor oud-burgemeester Van der Laan); te omvangrijke straatintimidatie in grote steden (als Rotterdam); de groei van het aantal voedselbanken (van 121 in 2011 naar 162 begin 2016) als indicator voor de armoede bij personen die daar een beroep op doen; de zwakke inburgering en slechte integratie van sommigen met een migratie-achtergrond; het gevoel dat asielzoekers met een verblijfstatus voorgetrokken worden; de begin 2018 nog niet opgeloste problemen van veel Groningers die ernstige last ondervonden van aardbevingen als gevolg van de aardgaswinning. Denk aan scheuren in de muren en aan een gestut huis.

De conclusie van de *fact check* op de uitspraak van Mark Rutte over Nederland als gaaf land moet dus zijn: 'de uitspraak is volgens de gehanteerde indexen waar'. Er is natuurlijk wel een kanttekening te maken. Scoort Nederland in 2017/2018 hoog genoeg om de klimaatdoelen van Parijs te halen in 2025 of 2050? Op dit vlak scoort Nederland niet bij de beste 25 landen, klagen sommige politici. 'Na kijken bij de burens' blijft er werk aan de winkel.

1.4 Van land naar binnenland: een voorbeeld van benchmarking

Als er volop vergelijkingen plaatsvinden, hoe moeten we die dan in concreto zien zodat ook scherper benchmarking in beeld komt. Een voorbeeld is verhelderend.

Een universitair medewerker is al meer dan vijftien jaar betrokken bij het internationale onderwijsprogramma EuroMba. Dit programma is een initiatief van een hoogleraar, die in de jaren negentig van de vorige eeuw een internationaal samenwerkingsverband van universiteiten startte met betrokken docenten uit Duitsland, Frankrijk, Ierland, Finland, Nederland en andere landen. Het programma wierf deelnemers via internetmarketing en inzet van buitenlandse partners, die reclame maakten en studenten 'aandroegen'. Aanvankelijk waren de deelnemersaantallen betrekkelijk beperkt. Het programma liep niet echt goed en niet evident slecht. Toch schroefde een onderwijsdecaan de universitaire deelname aan het programma inhoudelijk en qua personele inzet aanzienlijk terug vanwege het argument 'te weinig studenten, teveel inzet, te weinig inkomsten.' Kortom, de kosten-batenbalans van het commerciële onderwijsprogramma was in zijn ogen negatief. Maar de stekker werd er niet helemaal uitgetrokken. Er bleef een 'policy residue' over, een beleidsrest.

Een van de auteurs van dit boek kent het programma en nam ook een keer als voorzitter deel aan een vergadering van de partners in Dublin. Je moet weten waarover je schrijft. Wat is de stand van zaken een aantal jaren later? Intussen is het programma sterk gegroeid. Enkele partneruniversiteiten trokken zich weliswaar terug (Finland, Ierland) en andere kwamen erbij (Spanje, Polen), maar belangrijker is dat een maximum gesteld is aan het aantal in geschreven studenten. Er bestaat intussen zelfs een wachtlijst. Het programma loopt dus meer dan goed. We vroegen aan de betreffende medewerker dr. Ronald Gerkens (niet de echte naam, schr.): om hoeveel studenten gaat het nu? Zijn antwoord: 'er hebben zich zoveel mensen ingeschreven dat we een stop moesten instellen'. Hoe kan dat? Zijn reactie: 'omdat het EuroMba voor komt in "benchmark rankings" en omdat het programma daarin jaar na jaar goed scoort, en steeds tweede of derde staat, dus bovenaan. Daar letten studenten op. Het is daarom dat de studenten toestromen.' Dat verklaart dus de doorbraak na 'sukkelen' in de beginjaren. Volhouden loont en rankings doen ertoe, althans bij Mba's.

Vergelijking en rankings van onderwijsprogramma's als Mba's zijn in het algemeen relevant voor deelnemende onderwijsorganisaties (Van Vught & Westerheijden, 2010). De leiding kijkt ernaar uit. Het doel is om met een aantrekkelijk *verdienmodel* te werken, dus om de kas te spekken. De redenering luidt: *hoe groter de reputatie volgens rankings, hoe méér studenten en hoe méér inkomsten.*

Onderwijsprogramma's zijn niet bedoeld als bezigheidstherapie voor boventallig personeel. Mba's lijken weliswaar belangrijk als publicitair uithangbord voor een instelling, maar ze zijn in zekere zin ook marginaal omdat ze begin jaren negentig (nog) niet te rekenen zijn tot de eerste kerntaak van een hoofdzakelijk door de staat gefinancierde universitaire onderwijsinstelling, die erkende universitaire diploma's uitreikt.

Rankings van niet-kerntaakactiviteiten roepen zo gezien (enige) twijfel op. Is het niet veel belangrijker dat een universitaire instelling door visitatie (en accreditatie) hoog scoort op academische programma's op het gebied van onder meer managementwetenschappen, recht,

cultuurwetenschap, natuurwetenschap en psychologie? Het is mogelijk. We zien dan ook dat er bredere universitaire rankings verschijnen (Van Vught & Ziegele, 2011). Sommigen spreken in dit verband van *'het juk van de lijstjes'*.

1.5 Het beste bestaat of toch niet

Maar hebben lijstjes uiteindelijk zin of toch niet? De twijfel over vergelijking en rangordes valt breder te trekken. De bekende Nederlandse socioloog Johan Goudsblom, die zijn hele volwassen leven lang al aforismen formuleert (Goudsblom, 1958, 1976, 1998), liet de volgende uitspraak optekenen: *'niets is het beste'* (Elsevier, 220613). Een mooi kort aforisme waarmee hij, onbedoeld wellicht, twijfel zaait over vergelijking en over rankings, ook in andere contexten dan onderwijs.

'Niets is het beste' is een meer dan dubbelzinnige uitspraak. Ten eerste wordt het 'niets' van de grote leegte hiermee bejubeld. Met het grote 'niets' zijn we het beste af, zo lijkt een interpretatie te kunnen luiden. Waarom zou er zoveel van hetzelfde moeten zijn, zo dat één daarvan als het beste wordt betiteld? Welke woestijn is als uitgestrekt onvruchtbaar gebied het grootste?

Het is ook mogelijk te zeggen dat *variëteit* van belang is. Zorg – metaforisch gesteld - niet zozeer voor een jacht op de teelt van de *beste* appel maar teel *uiteenlopende* soorten fruit, als peren, pruimen, nectarines, appels, etc.

De uitspraak van Goudsblom verwijst ten tweede ook naar 'dat het altijd weer beter kan': de eeuwig durende en nooit eindigende *wedloop* voor het betere. Wat vandaag het beste lijkt, is het morgen of over een jaren al niet meer. 'Niets is het beste'. Vraag het bij twijfel maar aan ondernemers.

Een derde interpretatie van 'niets is het beste' houdt in dat het maar helemaal de vraag is wat *in welk opzicht* het beste is. Wat in één opzicht het beste is of lijkt, hoeft dat in andere opzichten helemaal niet te zijn. Wat in het Mba-voorbeeld als een *financieel* aantrekkelijk programma oogt, is misschien *inhoudelijk* een onderwijsprogramma dat enkele jaren later al weer verouderd is. Dat vroege partners uit Finland en Ierland geen deel meer uitmaken van het organiserend universitair Mba-consortium en plaats maakten voor een Poolse en Spaanse instelling kan wellicht duiden op een strategiewijziging bij de partneruniversiteiten, en/of twijfel scheppen over de financiële voordelen, en/of twijfel doen rijzen over te weinig 'verbinding' in de gemeenschappelijke sturing, en/of duiden op een als gebrekkig gepercipieerde actualiteit van het programma voor deelnemers uit de betreffende landen Finland en Ierland. Dat zijn andere aspecten van kwaliteit dan het financiële of inhoudelijke. Het zijn natuurlijk slechts hypothesen maar ze belichten dat er meerdere 'waarheden' (kunnen) zijn, uitgaande van uiteenlopende kwaliteitsaspecten.

Met dit aforisme 'niets is het beste' wil gezegd zijn dat benchmarking in de vorm van vergelijking en ranking een intrigerend meerdimensionaal verschijnsel is, hoewel je dat op het eerste gezicht niet zou denken. Door middel van vergelijking zo maar streven naar 'het beste' roept tal van vragen op. Wat is uiteindelijk het beste?

1.6 Onderwerp en definitie

We willen alles meten. De wereld zit vol met hitlijsten en rangordes. Soms zijn rangordes onderdeel van een systeem of zelfs de kern ervan. In het profvoetbal is sprake van een stand van voetbalclubs op een ranglijst. In de volksmond heet dit 'het linker en rechter rijtje'. Het linker rijtje bevat de namen van clubs die tot de betere helft van een clubcompetitie horen. Wie op het laatst bovenaan eindigt, krijgt de titel van kampioen. En wie als laatste eindigt op de ranglijst degradeert. Benchmarking is echter meer dan louter vergelijken en rangordenen, zoals in het betaald voetbal, bij tennis, bij het kiezen van een vakantiebestemming of een verzekeringspolis.

Benchmarken heet in populair taalgebruik ook wel *spiegelen, spieken en sprankelen*. Spiegelen verwijst naar vergelijken en rangordenen, spieken naar zoeken van het beste en sprankelen naar het

daadwerkelijk goed of zelfs excellent functioneren. Wie benchmarking zo ziet, verwaarloost de verschillanalyse.

Benchmarking is een benadering met de volgende kenmerken en daarmee verbonden ambities.

Benchmarking is gericht op:

- vergelijken van prestaties van organisaties of specifieke andere eenheden zoals bij voorbeeld opleidingen, vakgroepen of faculteiten;
- het rangordnen van resultaat- of prestatiescores van eenheden als organisaties;
- het verklaren van verschillen tussen prestatiescores (wie is *waarom* de beste?);
- het zoeken van 'beste praktijken' (wat is nastreefbaar om beter te worden?);
- en het daadwerkelijk zoeken naar verbetering.

Zo gezien, past benchmarking in een streven naar 'goed bestuur' of '*good governance*'.

Benchmarking in het licht van managementmodellen

Is benchmarking in het licht van managementmodellen een onderscheidend fenomeen? Inderdaad.

Wie een zogenaamd '*Managementmodellenboek*' raadpleegt, dat niet specificeert naar de private en publieke sector, kan daarin zo maar zestig modellen tegen komen. Te noemen zijn de aanpak van activity-based costing (abc-analyse), de balanced scorecard-aanpak, het concept van business proces redesign, de concurrentieanalyse, de Deming-cirkel, kennismanagement, het denken in termen van kerncompetenties, de overhead value analysis, de swot-analyse, het denken in termen van teams en teamrollen, de lerende organisatie en het zeven S-model (Ten Have e.a., 1999). Daarin is steevast ook een plek gereserveerd voor benchmarking. Benchmarking kan worden gezien als een in beginsel continu proces van het meten van producten, diensten of processen afgezet tegen de beste praktijken van 'degelijke concurrenten' of organisaties die anderszins als referentie dienen (ook al zijn ze geen concurrent). Het gaat om vergelijken, rangordnen van kwaliteiten (bij voorbeeld prestaties, effecten), verklaren van verschillen, het daardoor onderkennen van de beste praktijken en de wil om te verbeteren.

Heeft benchmarking sindsdien enige evolutie ondergaan? Benchmarking was oorspronkelijk, in de jaren dertig van de vorige eeuw, nog niet zo uitgewerkt want het werd slechts gezien als een vorm van *bedrijfsvergelijkende statistiek*. Maar in de loop van de tijd heeft benchmarking in het denken een ontwikkeling doorgemaakt van een relatief eenvoudig concept dat werd toegepast in de private sector tot een breder concept dat werd gebruikt in zowel de private als de publieke sector (Hakvoort en Klaassen, 2004, 2013). Tegenwoordig zijn er ook meer data beschikbaar; het woord 'big data' valt zelfs. Er valt meer te vergelijken dan in de vorige eeuw toen dataopslag niet zo ontwikkeld was als na 2000.

Spitsen we ons blikveld toe op de publieke sector, wat toont ons dan een boek als '*Publiek management, 65 modellen*' van adviseurs van Berenschot? Uiteraard zien we hierin meer toespitsing op de overheidsorganisaties, getuige het denken in termen van onder meer:

- organisatiefuncties (PIOFAH: personeel en communicatie, informatisering/archiefzaken, organisatie incl. research & development en incl. juridische zaken, financiën, control & faciliteiten, automatisering, incl. administratie, huisvesting);
- taaktijdanalyse, werkdrukanalyse in verband met personeelsvraagstukken;
- competentie management, kennismanagement;
- kwaliteitsmodellen als het INK-model;
- maar toch ook contractmanagement en het Tilburgs bestuursmodel;
- diagnose-aanpak en doelboom-aanpak voor beleidsanalyse;
- kerntakenanalyse in verband met bezuinigen;
- project-, programma-, proces- en ketenmanagement;
- simulatie en scenario-analyse in het kader van toekomstgerichtheid;

- contingentiedenken (niet alles kan op elk moment, elke plaats, bij elke persoon);
- werk- en inkomenmatchingmodel;
- risicomangement.

We zien hier voorbeelden van modelmatige aanpakken terug die we ook al zagen in het *'Managementmodellenboek'*. Maar onder deze 65 modellen vinden we toch veel modellen die we voor de private sector niet tegenkomen, zoals de begrotingsaanpak VBTB, die verwijst naar prestatiegericht begroten bij de departementen. Ook in deze opsomming is er plaats voor de balanced scorecard-aanpak en voor benchmarking (Martens e.a., 2002). Benchmarking is blijkbaar ook een nuttige aanpak voor toepassing in publieke sferen.

Leren door benchmarking

Het is van belang op te merken dat benchmarking geen doel op zich is. Robert Behn (2003) noemt benchmarking een methode van prestatiemeting met als ultieme doel: het verbeteren van de prestaties van de organisatie. Specifiek plaatst hij benchmarking onder de noemer van 'de lerende organisatie'. Achter benchmarking gaat de pretentie schuil een antwoord te vinden op de drie vragen: wat doet mijn organisatie goed, wat doet mijn organisatie niet goed (genoeg) en wat moet mijn organisatie doen om te verbeteren wat niet goed gaat?

Benchmarking moet dus behulpzaam zijn bij 'leren', bij verbeteren.

Bijdrage aan innovatie niet vanzelfsprekend

Benchmarking heeft een plek verworven in overzichten van instrumenten via welke geprobeerd wordt tot verbeteringen bij de overheden te komen. Gaat men verder snuffelen en waagt men het te kijken in het *'Groot Innovatie Modellenboek'*, dan vindt men daarin veertig innovatiemodellen voor het versterken van ondernemerschap in het bedrijfsleven en bij de overheid. Dit boek bevat werkwijzen en perspectieven in verschillende rubrieken, zoals de rubriek strategie & innovatie; de rubriek kansen & klanten; de rubriek keuzen & investeringen; de rubriek samenwerking & partners; de rubriek projecten & resultaten; en tenslotte de rubriek cultuur & organisatie. Welke concrete werkwijzen worden aanbevolen? De rubrieken volgend komen we het volgende tegen: bij voorbeeld scenarioplanning, het groeimodel van Ansoff, trendanalyse, de innovatieradar, crowdsourcing, innovatie portfolio management, de innovatie scorecard, five degrees of co-creation involvement, alliantievorming, netwerken, de customer value map, de innovation climate scan, het challenge driven innovation model (Kwakman en Smeulders, 2013).

Hierin is evenwel *géén* plaats ingeruimd voor benchmarking. Begrijpelijk want benchmarking betreft een comparatieve onderzoeksmethode of zoals deze vroeger wel genoemd werd: *een bedrijfsvergelijking*. Achter benchmarking gaat blijkbaar *niet* primair de pretentie schuil om onmiddellijk bij te dragen aan product- of procesinnovaties terwijl in het voorgaande toch al wel gesproken werd over het nastreven van beste praktijken. Wat is dan wel de ambitie?

Benchmarking als onderzoeksmethode

We herhalen even. Deze studie gaat over benchmarking, een onderzoeksmethode op het vlak van resultaatsturing of 'performance management'. Benchmarking staat voor de 3 S'en:

- *Spiegelen* (vergelijken),
- *Spieken* (bij andere organisaties kijken) en
- *Sprankelen* (verbeteren).

Bij benchmarking gaat het om het meten van prestaties en eventueel andere kwaliteiten, het vergelijken van prestaties tussen organisaties en het verbeteren van prestaties (door te zoeken naar beste praktijken). We verwijzen voor een voorbeeld naar het rapport *'Rotterdam vergelijkenderwijs – Benchmarks voor de verbetering van de bedrijfsvoering'* uit 2002 (Doorduyn e.a., 2002). Het gemeentebestuur geeft daarin aan dat tal van 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).

Toegepast onderzoek

Het is mogelijk om vanuit een wetenschappelijke doelstelling een benchmarkonderzoek op te zetten, maar hier, in dit boek, gaat het *niet* over kennisontwikkeling maar over benchmarking als *toegepast* of *praktijkgericht* onderzoek ten behoeve van opdrachtgevers, zoals bij voorbeeld besturen van universiteiten of opleidingen, pensioenfondsen, ziekenhuizen, verpleeghuizen, centra voor geestelijke gezondheidszorg, gemeenten, provincies of waterschappen. Benchmarkstudies van internationale organisaties als de Europese Unie, de OECD, het Internationaal Monetair Fonds en de Wereldbank blijven buiten beeld (verg. Van Vught & Westerheijden, 2010).

1.7 Onderwerp van het boek

Benchmarking is het onderwerp van dit boek. Benchmarkonderzoek is gericht op:

- a) prestatievergelijking van een aantal hoofdzakelijk subnationale overheidsorganisaties,
- b) het rangordnen van prestaties op basis van scores,
- c) het proberen te begrijpen van scoreverschillen,
- d) het zoeken naar goede of excellente praktijken, en
- e) uiteindelijk op verbeteren.

Wat is dat voor een beweging? Hoe kunnen we naar prestatie meting kijken? Als een aantrekkelijke of in werkelijkheid toch beperkte aanpak?

Achter dergelijke praktijkgerichte benchmarking gaan hardnekkige ambities schuil die steeds maar weer gedeeld en herhaald worden. We noemen ze *mythes*. Zoals de mythe of bezwerende formule dat benchmark-onderzoekers ook in werkelijkheid steeds gericht zouden zijn op ‘best practices’ en die ook steeds weten te vinden en aan te wijzen. Is die mythe werkelijkheid?

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? En, loopt die prestatie meting uit op een naargeestig ‘verdingelijkt’ of misschien juist fraai ‘*scorebordmanagement*’ of ‘*dashboardmanagement*’? Is sprake van het rangordnen om uit te vinden waar ‘het gras het beste is’, waar iets minder, en waar het er ronduit slecht bijligt?

Daarmee is benchmarking geïntroduceerd. Op verbetering gerichte vergelijking is in onze ogen in principe nuttig maar kent ook valkuilen. Het is vermoedelijk een begrensde hulpmiddel of instrument, zoals ook bestek niet voor alles geschikt is. Wie soep wil eten, kan het beste een lepel nemen omdat een vork en mes daarvoor ongeschikt is. Zo is het ook met benchmarking. Dat zal het hele boek door blijken.

Dit boek gaat over feiten en ficties over het meten, vergelijken, rangordnen van overheidsprestaties en begrijpen van verschillen daartussen, met bijzondere aandacht voor gemeenten.

We gaan in deze beschouwing allereerst op zoek naar een aantal kenmerken van benchmarking, voor- en nadelen 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. Met deze aanpak hopen we aanzetten te geven voor verbetering van benchmarkpraktijken.

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.

1.8 Doel

Deze beschouwing gaat in feite over hoe we een idee uit *New Public Management*, benchmarking, moeten beoordelen en wat ervan terecht komt. We willen in essentie de kwaliteit van benchmarking beoordelen. Wat is dus de betekenis van het werken met hitlijsten en rangordes bij de overheid, het streven naar de top en het verbeteren.

Het meer concrete 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. Wat zijn de voor- en nadelen? Zijn de nadelen van en valkuilen en risico's bij benchmarking te ondervangen? Sommige mogelijk wel, andere minder of niet.

Dat benchmarking nuttig is, wordt in dit boek duidelijk. Blijken zal dat de verwachtingen over *'analysis by numbers'* (vrij naar Hood, 2007), zoals benchmarking ook wel eens wordt aangeduid, op basis van praktijkervaringen echter niet zelden te hooggespannen zijn. Kortom, wie succes wil hebben met benchmarking zal uit de goede vaatjes moeten blijven tappen.

1.9 Bronnen

Wat we hier weergeven, komt niet uit de lucht vallen. Deze analyse en beschouwing baseren we op uitgebreide studie van benchmarking, op empirisch onderzoek uit binnen- en buitenland (o.a. Korsten, 1983; Aardema en Korsten, 2005; Abma en Korsten, 2009; Abma, 2012; Abma, 2014) en eigen ervaringen als deelnemer aan een benchmark, opdrachtgever hiertoe, toezichthouder 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). Een gemeentelijke benchmark krijgt hier bijzondere aandacht en dat is *waarstaatjegemeente* (voorheen: 'Staat van de Gemeente', een rond 2005 onder auspiciën van de Vereniging van gemeentesecretarissen (VGS) ontwikkelde en beproefde benchmark van de kwaliteit van gemeentebesturen. Wat zijn de uitkomsten van deze benchmark geweest en zijn de valkuilen bij deze benchmark vermeden?

Deze benchmark ('Staat van de gemeente') bestond in feite uit een reeks van jarenlang toegepaste benchmarks met steeds 15-20 deelnemende gemeenten. Deze benchmark werd via pilots ontwikkeld en daarna moesten deelnemende gemeenten de opzet volgen. Er kon niet door opdrachtgevers aan de opzet en toepassing gesleuteld worden. De opgedane ervaringen leidden er overigens toe dat er wel evoluties plaatsvonden; het instrument onderging aanpassing (zie Abma en Korsten, 2009). Recenter ging de 'Staat' door het leven als 'waarstaatjegemeente.nl'.

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

2.1 Inleiding

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, bij voorbeeld door initiatieven te nemen of uit te lokken, openbaar te gaan aanbesteden en klantgericht te gaan werken (Korsten, 1988). Een gedecentraliseerde organisatiestructuur met productverantwoordelijke eenheden zou ruimte bieden voor resultaatgericht werken, binnen de context van een productbegroting, managementrapportages en een jaarproductrekening. Hiervoor is wel de term 'contractmanagement' gebruikt (Van Helden, 1998: 20). De organisatiecultuur en de houding van medewerkers moest bovendien meer extern gericht worden (Ter Bogt, 2006: 170). De overheid moest het functioneren dus 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. Hood, 1991, 1995; Korsten, 1994; Korsten en Noordegraaf, 1995; Kickert, 2000; Teunisse, De Kiewit en Boshove, 2005; Kolthoff, 2007).

De door NPM-geïnspireerde vernieuwingen kregen bij Nederlandse gemeenten en provincies vorm via het zogenaamde BBI-project - het project beleids- en beheersinstrumentarium – en contractmanagement (Van Helden, 1998; Aardema, 2002; Ter Bogt, 2006: 170). Privatisering was ook niet langer een vies woord. En ook het op afstand zetten van overheidsorganisaties in de vorm van zelfstandige bestuursorganen (zbo's), met meer eigen beheermogelijkheden en dus vrijheid van handelen, paste in dat perspectief (Van Thiel, 2002; Jurriëns en Sylvester, 2005). Tenslotte, ook benchmarking opgevat als bedrijfsvergelijking werd bejubeld.

Volgens Christopher Hood (1991, 1995) waren de twee belangrijkste uitgangspunten van de vernieuwingsbeweging NPM het verkleinen van de verschillen tussen private en publieke organisaties en een transitie van publieke verantwoording over processen naar verantwoording over resultaten of prestaties. NPM richt zich op inzet bij overheidsorganisaties van meer bedrijfsmatige instrumenten en managementstijlen.

Gemeenten als Groningen en Delft maakten in de voetsporen van Tilburg (bekend vanwege 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. Sommige universiteiten, als de Universiteit Twente, werden 'omgekat' tot een ondernemende universiteit die meer budget uit de markt kon en wilde halen (uitvoeriger in Korsten, 1995; Grit, 2000).

2.2 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; McLaughlin e.a., 2002; Denhardt e.a., 2002; Van Helden en Jansen, 2002; Ter Bogt, 2006; Meier & O'Toole jr, 2009; Blank, Felsö en De Groot, 2011; Bekkers, 2012: 104; Koppenjan, 2012). 'Onder invloed van NPM is in de publieke sector een sterke toename opgetreden in de aandacht voor meer decentralisatie, concurrentie/marktwerking, privatisering, sturing op 'output' (producten), efficiëntieverhoging, prestatiemeting en transparantie omtrent kosten en prestaties en aan de commerciële sector ontleende managementmethoden', schrijven Ter Bogt en Tillema (2010: 50). 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). Innovatie komt dus wel om de hoek kijken.

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. Steeds was de bedoeling om gewenste prestaties, dus normen te expliciteren. Een productbegroting geeft informatie over de drie W's: wat willen we, wat doen we daarvoor, wat mag dat kosten?

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 en veel processen zich niet laten vangen in kwantitatieve termen. En 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 doorgesloten bedrijfsmatigheid (ook Karsten en Van Veen, 1998; Kickert, 2000).

2.3 Verschil 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. Rainey c.s. (1976) noemen 25 verschillen tussen publieke en private organisaties. We noemen hier een beperkter aantal verschillen, die naar verwachting van invloed zijn op hoe publieke organisaties omgaan met benchmarking en benchmarkinformatie gebruiken om zich te verbeteren. In het algemeen maken deze verschillen de toepassing van 'best practices' uit benchmarkrapporten en 'het leren' moeilijker (cf Tillema, 2006).

1. *Geringe marktwerking*: In de private sector is expliciet sprake van marktwerking en in de publieke sector niet of nauwelijks. Burgers kunnen op de markt veelal kiezen tussen aanbieders terwijl burgers in de publieke sector niet kunnen kiezen; ze zijn veroordeeld om te wonen in één gemeente maar hebben de keuze om te verhuizen. In de publieke sector gelden bovendien tal van andere principes en waarden dan marktprincipes, zoals sociale rechtvaardigheid, humaniteit, duurzaamheid, vrijheid, legaliteit, veiligheid, democratische totstandkoming, toegankelijkheid, gelijke behandeling, openbaarheid, redelijkheid en proportionaliteit. In de publieke sector komt een variëteit aan waarden voor die de waardering bepalen van wat geproduceerd wordt (uitvoeriger in Korsten, 2010a; Hakvoort & Klaassen, 2010: 107). Evenals in de private sector komen in de publieke sector overigens wel onderhandelingen voor. Wat betekent deze geringe marktwerking voor benchmarking? Dat benchmarking een substituuat kan zijn voor publieke organisaties die marktinformatie missen (Tillema, 2006).
2. *Niet altijd uitsluiting*: Private bedrijven produceren goederen en diensten in een bepaalde kwantiteit tegen een bepaalde prijs. Wie de prijs niet betaalt, wordt in principe uitgesloten van een goed of dienst. In het kader van overheidsbeleid vindt er vaak discussie plaats over de vraag of een goed of dienst toch moet worden verstrekt als de prijs ervoor niet kan worden

opgebracht. Soms kan een goed of dienst tegen een inkomensafhankelijke aanpak worden verkregen (variërende eigen bijdrage). Soms vindt kwijtschelding van een schuld plaats of kan vrijstelling worden verkregen of kan een andere oplossing opportuun wordt geacht, zoals de verstrekking van een pas of kortingsfaciliteit voor bepaalde groepen of categorieën (Borst, Lako en De Vries, 2012: 16). Wat betekent dit voor benchmarking? Dit kenmerk van publieke organisaties kan een rol spelen bij de vereiste of wenselijke normering van producten, diensten of processen die in benchmarkonderzoek een rol spelen.

3. *Collectief of individueel?*: 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, huizenbezitter of energieverbruiker, en niet zozeer als individuele consument. In de private sector worden burgers juist als individuele klanten of consumenten gezien (Borst, De Vries en Lako, 2012: 16). Pogingen om door middel van kartel vorming iets af te spreken voor consumenten in een bepaalde sector, bij voorbeeld een minimumbenzineprijs, zijn hier en daar verboden. Wat betekent dit voor benchmarking? Bij benchmarking kan dit verschil tussen publieke en private organisaties ertoe leiden dat bij vergelijking van producten, diensten of processen van publieke organisaties er rekening gehouden moet worden met het specifieke gemeentelijke beleid dat achter bij voorbeeld rioleringen, afvalverwerking of sportvoorzieningen schuil gaat. Wie bij voorbeeld de staat van rioleringen vergelijkt, kan er niet bij elke organisatie vanzelfsprekend vanuit gaan dat de rioleringen van alle gemeenten even lang meegaan. Afgezien van kwaliteitsverschillen in buizen kan een rioleringssysteem in een bepaalde gemeente door verschillen in de fysieke structuur meer te lijden hebben dan in een andere. De afschrijftermijnen kunnen ook verschillen.
4. *Productkenmerken*: De heterogeniteit van producten is in de publieke sector groter dan veelal bij private organisaties het geval is. En er komen in de publieke sector meer moeilijk identificeerbare producten en processen voor. 'Unlike the production in private companies, the production of organizations in the public sector is often difficult to identify, and it is even less easy to completely express it in (financial) ratios' (Hakvoort & Klaassen, 2010: 107). Een gemeentelijke organisatie produceert naast de verstrekking van identiteitsbewijzen en paspoorten - voorbeelden van concrete producten -, ook slecht identificeerbare en niet-homogene producten, 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 problemen die alleen in beleidsnetwerken 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; Noordhoek, 2013).
5. *Processen*: Een reguliere 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 public relations, propaganda en convenanten). Soms zal een overheidsorganisatie ook symbolisch beleid formuleren terwijl een private organisatie dit minder gauw zal doen.
6. *Meervoudige doelstellingen*: 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 overheidsbestuur evenwel geld bij omdat iets maatschappelijk van belang wordt geacht. Consequenties? Die meervoudigheid maakt het moeilijker om benchmarkinformatie in publieke organisaties eenvoudig te benutten voor het opzetten en besluiten over verbeteracties (Tillema, 2006).

7. *Gemengde sturing*: Wie gemeentelijke programmabegrotingen bekeek en bekijkt, ziet dat sprake is van pogingen om de drie w's in te vullen: wat willen we bereiken, wat doen we daarvoor en wat mag het kosten? De vierde w ligt daarna op de loer: wat levert het op? Dat vierde punt lijkt een pre voor benchmarking in de publieke organisaties. Benchmarking levert immers prestatiegegevens. Dit aandachtspunt van wat het beleid oplevert, zien we bij voorbeeld ook terug in de begroting van de gemeente Amsterdam. Kunnen we spreken van een eenvoudige aanpak? Een bestuur zal tegenwoordig weliswaar vaak pogen te sturen op 'outputs' of op 'outcomes' (gewenste effecten) maar we weten inmiddels ook dat dit maar heel beperkt lukt, mede door de meervoudigheid van doelstellingen en het feit dat in politieke arena's soms tot symbolisch beleid wordt besloten, dat helemaal niet bedoeld is om uit te voeren. Van uitsluitend sturen op output blijkt in gemeenten in het algemeen echter geen sprake en van sturen op outcomes evenmin (Haselbekke, 1998; Bordewijk en Klaassen, 2000; Wassenaar en Verhagen, 2006; Bouckaert e.a., 2011; Abma, 2012).

8. *Beperkingen*: Overheden hebben meer dan private organisaties te maken met begrenzingen ten aanzien van hun doelstellingen en activiteiten als gevolg van wetten en andere regelingen en geografische binding. De wetten van hogere overheden gaan in het rechtsstelsel dat we kennen voor die van lagere. Een overheidsbestuur heeft dus geen onbegrensde keuze van alternatieven. Zo kan een gemeente uit Noord-Brabant niet verhuizen naar Flevoland of een andere gemeente in het buitenland opkopen. Het gevolg kan zijn dat een overheid ook bij de keuze van verbeteracties als gevolg van benchmarking minder vrijheid van handelen heeft dan een private organisatie (Tillema, 2006: 301).

9. *Politieke invloed*: Een onvermijdelijk te noemen verschil betreft de politieke invloed. Overheidsbesturen kennen een dagelijks bestuur en een algemeen bestuur. Een algemeen bestuur bestaat uit door kiezers gekozen representanten die op een kandidatenlijst stonden. Na verkiezingen volgt de coalitievorming. Die resulteert in een dagelijks bestuur dat, naast een benoemde bestuurder, meestal bestaat uit een aantal bestuurders die voortkomen uit de gekozenen. De benoemde bestuurder (burgemeester, commissaris des konings) en de overige dagelijks bestuurders (wethouders, gedeputeerden) moeten het vertrouwen hebben van het algemeen bestuur (gemeenteraad, provinciale staten). Deze leden van het dagelijks bestuur krijgen bovendien te maken met een actieve informatieplicht en openbaarheid van beraadslagingen in het algemeen bestuur. In private organisaties is geen sprake van gekozen bestuurders en speelt openbaarheid nauwelijks een rol. Maar ook hier geldt wel een vertrouwensregel. De rol van de politiek wordt in overheidsorganisaties ook duidelijk via verwijzingen naar 'de achterban', de wenselijkheid van het verkrijgen van steun voor voorstellen en het fenomeen lobby. Derden kunnen via lobby of andere participatievormen proberen invloed op bestuurders uit te oefenen. De besluitvorming in overheidscontexten is als gevolg van de electorale aspecten, openbaarheid en het belang van steun complexer dan in de private sfeer. Wat is het gevolg? Dat kan een benutting van uitkomsten van benchmarkonderzoek in overheidsorganisaties bemoeilijken in vergelijking met private organisaties (Tillema, 2006: 301).

In de (inter)nationale literatuur vinden we meer in het algemeen analyses van overeenkomsten en verschillen tussen publieke en private organisaties en management daarin (o.a. Allison, 1983, 1997; Mintzberg, 1996; Denhardt e.a., 2002; Rainey, 2003: 55-79; Boyne e.a., 2006; Pestman, 2007; Starling, 2011; Bouckaert e.a., 2011; Noordegraaf e.a., 2013).

We kunnen wel wijzen op verschillen tussen publieke en private organisaties maar daarmee is nog niet gezegd dat bedrijfsmatigheid en zakelijkheid helemaal uit de wereld van publieke organisaties te bannen zijn. Dat is ook wel gebleken. De NPM-wind ging niet onmiddellijk na deze discussie liggen. NPM kan zeker een impuls hebben gegeven die positief is geweest of nog is. Geldt dat ook voor benchmarking?

2.4 Benchmarking past bij een zakelijke overheid

Benchmarking heeft zich ontwikkeld van een hoofdzakelijk op doelmatigheid van productieprocessen gericht onderzoeksinstrument (economische rationaliteit) naar een breder toepasbaar instrument waarin ook aandacht aan andere waarden, passend in een perspectief van juridische en politieke rationaliteit, wordt besteed (Hakvoort & Klaassen, 2010: 108). Benchmarking mag dan – historisch gezien - ontleend zijn aan toepassing in de private sector, het lijkt erop dat benchmarking naadloos past in een meer bedrijfsmatige aanpak binnen de publieke sector, ook al is het zo dat veel van de criteria ter beoordeling van overheidsbeleid grotendeels verschillen van die uit de private sfeer.

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).

2.5 Het gras is elders groener

Benchmarking is dus een vanzelfsprekendheid geworden en geaccepteerd geraakt (Francis & Holloway, 2007: 171; Hakvoort & Klaassen, 2010). Maar waar hebben we het eigenlijk over bij benchmarking? Het wordt tijd om preciezer te worden dan tot nu toe in dit hoofdstuk. Benchmarking wordt op de website van Berenschot 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 vergelijkend 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. Anders gesteld: benchmarking moet voorkomen dat elke deelnemende organisatie steeds het wiel opnieuw probeert uit te vinden (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 is dat nog steeds 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 benchmarkings 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 gedijt benchmarking het beste? 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 Dronkers-lijst van schoolprestaties 2012; ook Dronkers, 2011; Visser, 2003; Pestman, 2007) 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 gemakkelijk 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.

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 gemeenten waar het grote belang van 'een lerende overheid' door het bestuur een en andermaal uitgesproken wordt en *minder* voorkomt in bestuurlijke probleemgemeenten, waar de sneuvelkans voor bestuurders groot is (Korsten en Schoenmaker, 2011).

Probleemgemeenten zijn bestuurlijk risicovolle gemeenten waar in een jaar of twaalf, drie perioden tussen raadsverkiezingen, vaker burgemeesters en/of wethouders sneuvelen en coalities uit elkaar vallen. Achter die status van bestuurlijk risicovolle gemeente 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 gemakkelijk daarvan afgeweken. In die lijn past dat veel beleidskwesties als individuele gevallen worden behandeld. Daardoor neemt de kans op cliëntelisme, bestuurlijke wispelturigheid en conflicten in de besluitvorming toe. Vaak is in probleemgemeenten ook sprake van een *rauwe politieke cultuur*. Afsplitsingen van politieke fracties bevorderen dat.

De conflict *aanjagende* krachten zijn in probleemgemeenten sterker dan de *temperende* krachten. De temperende krachten, zoals een krachtige verbindende rol van de burgemeester en informeel

overleg van fractieleiders om goed – lees beter - te vergaderen, zijn betrekkelijk zwak ontwikkeld of komen niet voldoende uit de verf.

Is er een directe indicator aan te treffen die verwijst naar de status van probleemgemeente? Meestal is dat een cultuuranalyse. Als zelfs de gemeenteraad (bijna), het college van b&w of de commissaris van de koning(in) ten einde raad is over het functioneren van een gemeentebestuur, wordt vaak opdracht gegeven tot een externe analyse die gaat onder de vlag van *cultuuranalyse*. Voorbeelden van probleemgemeenten waren Zundert, Leiden, Maasdriel, Den Helder, Delfzijl, Lingewaard, Bloemendaal. Actuele probleemgemeenten uit de jaren na 1998 zijn bij voorbeeld Losser, Stein en Steenbergen. In een aantal van deze gemeenten zijn ook cultuuranalyses uitgevoerd, zoals bij voorbeeld in Maasdriel, Stein, Steenbergen, Oude IJsselstreek, Bloemendaal.

Als probleemgemeenten al meededen met de integrale benchmark *waarstaatjegemeente* dan scoren ze op tal van indicatoren voor kwaliteit van het gemeentebestuur relatief lager dan tal van andere gemeenten (Abma en Korsten, 2009; Abma, 2014). In een ranglijst van deelnemende gemeenten met scores van burgers op bestuurlijk vertrouwen staan ze constant in de lagere regionen. Heel begrijpelijk.

Bestuurders van probleemgemeenten kunnen doorgaans niet of niet goed voldoen aan een eis uit de *'Code voor goed openbaar bestuur'*, namelijk het beschikken over zelfreinigend vermogen. Ze krijgen dit vermogen door middel van evaluaties en benchmarking niet goed geactiveerd. Ook na verrichte cultuuranalyses blijft het vaak nog aanmodderen, zoals raadsleden uit Stein en Steenbergen ons melden. Dat wordt mede veroorzaakt doordat bepaalde spelers die bijdragen aan de status van bestuurlijke probleemgemeente ook weer betrokken zijn bij verbeteracties en het daar dan opnieuw laten afweten (Korsten en Aardema, 2006; Korsten, 2010a; Korsten en Schoenmaker, 2011; Schoenmaker, 2011). Het enige middel om 'rotte appels' te verwijderen uit de mand zijn, althans voor wat de gemeenteraad betreft, nieuwe verkiezingen die leiden tot het wegstemmen van de probleemveroorzakers in de raad. Bestuurders van een politieke partij, welke dan ook, hebben zelf vaak onvoldoende vermogen om intern te saneren en ze zien vooral de rotte appels bij andere partijen of verwijzen naar de gehele cultuur.

2.6 Op bepaalde gemeentelijke terreinen

Gemeenten kennen sinds een aantal jaren tal van benchmarks. Fig. 2 geeft een indicatie hiervoor. De lijst is gebaseerd op wat er zoal via informatie bij de Vereniging van Nederlandse Gemeenten, internetsites en gemeentesecretarissen van gemeenten te achterhalen viel. Desalniettemin, de lijst is niet volledig. Er bestaat geen register van erkende benchmarks. In werkelijkheid bestaan er meer benchmarks. Zo komen we via googelen ook wel IT-benchmarking tegen, maar we kennen geen voorbeelden uit de gemeentelijke praktijk. Daarnaast moet opgemerkt worden dat we uitgegaan zijn van het woord 'benchmark' in de titel. In werkelijkheid is echter niet gegarandeerd dat het hier ook met zekerheid gaat om benchmarking.

Uit het overzicht blijkt een behoorlijke spreiding naar object of onderwerp. In vergelijking met de functiegebieden van de gemeentelijke begrotingen valt het op dat benchmarks niet of nauwelijks betrekking hebben op winkelvoorzieningen, ruimtelijke ordening, verkeer en vervoer.

Fig.2: Inventarisatie van benchmarking van gemeenten in de periode vanaf 2000, in willekeurige volgorde

Nr.	Onderwerp
1	Basisbenchmark Wet maatschappelijke ondersteuning (wmo)
2	Benchmark afvalinzameling
3	Benchmark afvalscheiding
4	Benchmark ambtelijk apparaat gemeenten
5	Benchmark begroting(scijfers)
6	Benchmark behandeling bezwaarschriften
7	Benchmark bouw- en woningtoezicht
8	Benchmark brandweer
9	Benchmark burgerparticipatie
10	Benchmark formatie afdeling veiligheid, vergunningen en handhaving
11	Benchmark gemeentelijk ondernemingsklimaat
12	Benchmark gemeentelijk ondernemingsklimaat middengemeenten
13	Benchmark grondbedrijven
14	Benchmark handhavingscapaciteit
15	Benchmark individuele verstrekkingen
16	Benchmark inspanningen gemeenten actualiseren bestemmingsplannen
17	Benchmark jeugd
18	Benchmark juridische functie
19	Benchmark kwaliteit gemeentelijke dienstverlening
20	Benchmark publiekszaken
21	Benchmark rioleringszorg
22	Benchmark schuldhulpverlening
23	Benchmark verbeter- en innovatieprogramma economische zaken
24	Benchmark wet voorziening gehandicapten
25	Benchmark work first
26	Benchmark algemene bijstandswet
27	Benchmark binnenstadsdistributie
28	Benchmark financiële functie
29	Benchmark wet werk en bijstand
30	Benchmark wet waardering onroerende zaken (WOZ)
31	Benchmark implementatie eerstelijnstoezicht Wet kinderopvang gemeenten 2005
32	Benchmark kwaliteitsverbetering openbare gezondheidszorg
33	Contractenbenchmark Blik op werk
34	Doe mee-onderzoek verbonden partijen
35	Exploitatieonderzoek zwembaden
36	Facilitaire kengetallen gemeenten
37	Haalbaarheid benchmarking leerlingenvervoer
38	Lokaal onderwijsbeleid: zoeken naar voorbeeldgemeenten en 'good practices'
39	Woonbelevingsonderzoek/ wonen en leefbaarheid
40	Stopwatch: doorlooptijden vergunningverlening
41	Staat van de Gemeente/waarstaatjemeente
42	Strategie promotie en acquisitie
43	Vergelijking gemeentelijke uitgaven
44	Verkenning bevordering arbeidsparticipatie

2.7 Mythen over benchmarking

Welke kwaliteiten worden nu aan benchmarking toegeschreven?

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 meer bedrijfsmatigheid in het functioneren van organisaties.

In dit boek wordt getracht over die mythen iets te zeggen aan de hand van onze ervaringen met benchmarking in het openbaar bestuur, in het bijzonder bij 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.

Welke mythen onderscheiden we hier? We noemen er enkele 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 houdt in: *Benchmarking leidt tot het opsporen van 'best practices' (goede, navolgbare praktijken) en bezinning daarop in benchmarkrapportages.* Benchmarking is dus meer dan het opmaken van een klassement, dan een rangorde van organisaties al naar gelang de geleverde prestaties.

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 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 (met als gevolg dat minder personeel nodig is, minder budget nodig is, een betere afstemming van werkzaamheden kan plaatsvinden wat leidt tot meer snelheid, enz.).

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



Benchmarking een verdienmodel?

3 Benchmarking: kenmerken, vormen, doelen, grenzen

Leidinggevend 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. Evalueren en leren is volgens de code een vereiste. En ook 'zelfreiniging' werd door minister Guusje ter Horst en staatssecretaris Ank Bijleveld wenselijk geacht. Benchmarking is een wenselijk streven, ook al staat zelden in een lokaal coalitieakkoord dat een benchmark op een bepaald terrein moet worden gestart (Addink, 2010; Korsten, 2011; Abma, 2012: 353 e.v.; Schutgens, 2014).

In dit hoofdstuk wordt nader ingegaan op kenmerken van benchmarking, vormen van benchmarking en de grenzen.

3.1 Benchmarking is geen beleidsevaluatie of monitoring

Heeft benchmarking veel of weinig weg van beleidsevaluatie en monitoring? Benchmarking is een vorm van vergelijkend cijfermatig beoordelen van wat organisaties deden en doen, maar het is *geen* beleidsevaluatie *vanuit* overheidsdoelstellingen. Benchmarking houdt over het algemeen geen rekening met de strategie en missie van een organisatie (De Vries en Lako, 2009: 75).

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 overheidsbeleid zelf of aan omstandigheden daarbuiten (Bressers en Hoogerwerf, 1995; Rossi, Lipsey & Freeman, 2004; Swanborn, 2007).

Beleidsevaluatie kan op het verleden gericht zijn, dus retrospectief, of meer prospectief zijn.

Benchmarking impliceert *zelden of nooit* het beoordelen van een bestuur op beleidsdoelstellingen. In een benchmarkrapport valt nooit te lezen welke doelstellingen aan de benchmarking deelnemende organisaties erop nahielden of nahouden en evenmin of de beleidsdoelstellingen wel of niet bereikt zijn. Benchmarking is bovendien *altijd* ex post. Benchmarking kan zeker toekomstgericht zijn maar is dat zelden, tenzij men het op zoek gaan naar 'best practices' vanuit de houding om te willen verbeteren als toekomstgerichtheid opvat.

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 proces 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 (doorvoer, verwerking) of de output (resultaten). Benchmarking vervult die functie 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 wordt (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).

Dat evaluatie geen benchmarking is, blijkt uit het volgende voorbeeld.

In 2007 maakte toenmalig minister voor Wonen, Wijken en Integratie Ella Vogelaar (PvdA) het Actieplan Krachtwijken bekend. Dat beoogde in tien jaar tijd markante verbetering te bereiken in veertig achterstandswijken op het gebied van wonen, werken, leren en opgroeien, integreren en veiligheid; deze zogenaamde Vogelaarwijken moesten van klachtwijken naar krachtwijken groeien. In een publicatie van het Sociaal en Cultureel Planbureau '*Werk aan de wijk*' valt te lezen hoe het de bewoners van de aandachtswijken sinds de start van de overheidsinterventie sociaaleconomisch en in ander opzicht is vergaan. Zijn er kansen voor sociale stijging ontstaan? In hoeverre kampten de wijken met selectieve migratie: lage inkomens die instromen en hogere inkomens die uitstromen? Het meest benieuwd waren de onderzoekers naar de effectiviteit van het krachtwijkenbeleid om de leefbaarheid, veiligheid en sociaaleconomische positie van de veertig aandachtswijken te verbeteren. Die werd geëvalueerd met een quasi-experimentele onderzoeksopzet. De onderzoekers toetsten afzonderlijk de effectiviteit van de wijkinterventies op het vlak van herstructurering, de verkoop van sociale huurwoningen en een pakket sociale maatregelen. Ze vergeleken de Vogelaarwijken met andere wijken in elke onderzochte gemeente. Wat bleek onder meer?

Het krachtwijkenbeleid heeft tussen 2008 en 2013 *geen* onderscheidend gunstig effect gehad op de sociale stijging, leefbaarheid en veiligheid in de onderzochte aandachtswijken. Wijken die er qua problematiek in 2007 het meest op leken, maakten vergelijkbare ontwikkelingen door. Wel waren er enkele positieve trends in de aandachtswijken, zoals dat de concentratie van mensen met lage inkomens sinds het midden van de jaren nul geleidelijk minder is geworden en dat de tevredenheid met de woonomgeving in de laatste tien jaar aanhoudend verbeterd is. Bewoners van aandachtswijken blijken ook optimistischer te zijn geworden over de ontwikkeling van hun wijk. De verkoop van sociale huurwoningen in stadswijken had in de periode 2007-2010 echter geen meetbare invloed op de leefbaarheid in de wijk (Permentier e.a., 2013).

Deze studie kreeg veel aandacht in de pers. De hoofdconclusie werd: geen gunstig effect meetbaar van het krachtwijkenbeleid. Er blijft dus werk aan de wijk nodig, zoals het Commentaar in De Volkskrant luidde (300713). De vraag is wiens taak het is. De commentator schrijft: 'Het is een primaire taak van gemeentebesturen en, waar veel sociale woningbouw staat, van woningbouwcorporaties. Het Rijk kan er bij behulpzaam zijn, maar dirigisme vanuit Den Haag helpt niet'.

Overigens is het beeld van de onderzoekers een algemeen beeld. Er zijn natuurlijk afwijkingen. Een wethouder van Enschede is er als de kippen bij en maakt dat al direct in een reactie duidelijk: 'De investeringen in de wijk Velve-Lindenhof hebben wel degelijk een flinke verbetering teweeg gebracht', aldus wethouder Hatenburg (NRC, 030813). Daarbij wordt verwezen naar een ander onderzoeksbureau en een ander rapport.

Wat opvalt aan dit beleid en het onderzoek is dat het krachtwijkenbeleid een zaak van langere adem is. Ten tweede, en daar gaat het hier om, de studie is *geen* benchmarkonderzoek omdat het doel van de SCP-onderzoekers weliswaar was om wijken te vergelijken maar *niet* om de beste wijk eruit te halen. Het ging niet om het opsporen van 'best practices' en dus niet om manieren te vinden waarop de corporatie(s) en het gemeentebestuur van een specifieke gemeente met een of meer aandachtswijken hun aanpak zo nodig zouden kunnen verbeteren. Het ging om een evaluatie van *rijks*beleid ten behoeve van de *rijksoverheid*. Dit onderzoek was om allerlei redenen nodig. Een belangrijke reden was en is dat de Algemene Rekenkamer oordeelt dat van dit beleid uiteraard een evaluatie nodig is. Het moet voor de Rekenkamer steeds duidelijk worden, zeker als het gaat om betrekkelijk omvangrijke budgetten, of sprake is van (on)zinnige beleidsinterventie(s).

Omdat benchmarking geen vorm van beleidsevaluatie vanuit overheidsdoelstellingen of van netwerkpartners is noch van interbestuurlijk gerichte monitoring, kan men de vraag stellen: zouden lokale rekenkamer(commissie)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 door rekenkamer(commissie)s uitgevoerde onderzoeken zijn er, uitgaande van de titels en korte beschrijvingen van de inhoud, vrijwel geen te

typeren als benchmark-onderzoek. Wie ook benchmarking verricht, lokale rekenkamer(commissie)s doorgaans 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 *waarstaatjegemeente* 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; Abma 2014: 32-33).

3.2 Omschrijving

Benchmarking impliceert, we herhalen het nog eens

- a) het vergelijken van organisaties of delen of aspecten daarvan (zoals producten, diensten of processen) in kwantitatief opzicht door middel van scores,
- b) het (rang)ordenen van uitkomsten op vergelijkingspunten,
- c) het proberen te begrijpen van verschillen,
- d) het zoeken naar 'best practices', en
- e) 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 (zie Abma en Korsten, 2009). Of denk aan het vergelijken van vijftien verpleeghuizen of verzorgingshuizen.

Dit is de *brede* (normatieve) opvatting van benchmarking.

Nico Nelissen en Peter 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.

Wij gaan hier uit van de gangbare (brede) opvatting van benchmarking.

Kwaliteitshandvest

Vergelijken zonder een geschikte norm, zonder dat bekend is wat een goede, voldoende of (on)aanvaardbare score is, is lastig. Vergelijking kan plaatsvinden op verschillende vergelijkingspunten. Een vergelijking is mogelijk ten opzichte van een standaard, zoals een op theoretische basis verworven *ijkpunt* of een door de praktijk geïnspireerd *streefcijfer*. Standaarden blijken te ontleenen aan jaarprogramma's, strategienota's of rapportages van inspecties. Is geen standaard te vinden dan biedt onderlinge vergelijking van organisaties op (clusters van) diensten, producten of processen wellicht een mogelijkheid om te zien wie het beter doet en wie slechter. Maar alleen vergelijken is voor ons niet voldoende om te spreken van benchmarking.

Nogal eens hebben gemeenten 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 stellen niet heel veel voor. 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, re-integratie naar een baan, opvang van daklozen, parkeerbeleid en coffeeshopbeleid. 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? Zeker wel.

3.3 Perspectief op meting

Vergelijken, en dan? Er zijn naast de genoemde brede en smalle opvatting nog twee andere perspectieven op prestatiemeting bij benchmarking mogelijk: een meer productgerichte benadering en een procesgerichte benadering.

In het *productgerichte perspectief* op benchmarking, ook wel het *aanbodgerichte model* genoemd, is sprake van een *vast stramen van stappen* in het meetproces. De opzet van het benchmarkonderzoek wordt opgelegd aan de deelnemers. Er bestaat weinig oog voor variëteit aan de kant van de deelnemende organisaties, bij voorbeeld ten aanzien van de opdracht of rapportage, want die variëteit kan de vergelijking en oordeelsvorming alleen maar bemoeilijken. De opzet is sterk gericht op *'take it or leave it'*. De onderzoeker is vooral geïnteresseerd in aantallen, zoals medische handelingen in ziekenhuizen, stoelbezetting in schouwburgen, toetsuitslagen van leerlingen en scholen (De Bruijn en Van Helden, 2007: 21).

In de *procesgerichte benadering* is meer sprake van gezamenlijke beeldvorming tussen deelnemende organisaties over de opzet van het benchmarkonderzoek en daarmee van overleg met de deelnemers aan de benchmarking. Opdrachtgevers en onderzoeker trachten het eens te worden over de opzet van het onderzoek, de te onderzoeken prestatievelden, de indicatoren, de interpretatie van uitkomsten en de rapportagevorm. Daarbij bestaat enig oog voor bijzondere omstandigheden aan de kant van het te onderzoeken object. Er wordt door de opdrachtnemer, de benchmarkonderzoeker, niet alleen gekoerst op een adequate opzet van het onderzoek maar ook op draagvlak, op het begrijpen van uitkomsten door professionals en eventueel bestuurders van deelnemende organisaties en op de (condities voor) benutting van bevindingen. Het gaat tenslotte om te 'leren' van de bevindingen.

De benchmarkonderzoeker zal in deze benadering niet alleen letten op aantallen (bij voorbeeld niet alleen op aantallen vonnissen maar op 'recht spreken'), niet alleen op toetsuitslagen maar ook op de kwaliteit van onderwijs, niet alleen op de stoelbezetting in schouwburgen maar ook op het theateraanbod in zijn geheel, op de variatie in het productassortiment en de oogmerken daarachter (De Bruijn en Van Helden, 2007: 22).

Deze verschillen tussen de twee benchmarkbenaderingen zien we terug in de prestatienormen die een rol spelen in het benchmarkonderzoek en in de scoringseisen. Vanuit elke benadering zijn eisen te stellen aan de opzet van een zinvol benchmarkonderzoek.

We gaan verderop niet steeds dit onderscheid aanbrengen, maar komen er nog wel op terug. We zullen zien dat nogal eens *teveel* vanuit een productgericht benadering gedacht wordt en te weinig vanuit een procesaanpak. Bij benchmarking van wat gemeentebesturen doen en laten, domineert na het jaar 2000 de productgerichte aanpak. Het zijn adviesbureaus die een benchmark ontwikkelden en in feite in handen hebben. Zij zoeken steeds bij volgende benchmarks gemeenten aan om deel te nemen. Maar in werkelijkheid komt het wel voor dat gemeenteambtenaren voorkeuren en wensen kunnen uiten. Of dat (n)ergens toe leidt, zal nog blijken.

3.4 Kenmerken van benchmarkonderzoek

We hebben nu de (wenselijke) kenmerken van (vooral productgerichte) benchmarking te pakken, zoals die in de literatuur benoemd worden.

Een benchmarkonderzoek is een *vorm van vergelijkend toegepast onderzoek*, waarover verslag gedaan wordt in de vorm van een rapport. Een benchmarkrapport is daarmee dus *meer* dan een geheel van tabellen, figuren of statistieken of slechts een mondelinge presentatie en uitwisseling. Immers, dan zou het risico bestaan van onbegrijpelijkheid voor niet-insiders.

Een benchmarkrapport resulteert in *een verslag* van onderzoek op basis van een vraagstelling (waarop is de onderzoeker uit), een werkwijze, resultaten en conclusies, die zelfstandig leesbaar zijn.

Het (toegepast) onderzoek moet in principe voldoen aan eisen van relevantie, geldigheid (meet ik wat ik wil meten?), betrouwbaarheid en herhaalbaarheid. De lezer moet ook kunnen begrijpen hoe de onderzoeker tot conclusies komt en welke informatiebronnen gebruikt zijn. Dat heeft tot gevolg dat een benchmark niet slechts kan bestaan uit een aantal 'statements'.

Een benchmark kent in het openbaar bestuur in principe *deelnemende organisaties* en een *bereidheidsverklaring* tot deelname of in geval van financiële implicaties en wensen zeker een *opdracht*. Of volstaan wordt met één uitgebracht rapport waarin de vergelijking wordt gemaakt tussen alle deelnemers aan het onderzoek of ook afzonderlijke rapporten per deelnemer en gebruiker van de bevindingen verschijnen is de vraag. Als gekozen wordt voor een rapport en specificatie van bevindingen naar afzonderlijke deelnemers en de 'best practices' aan de orde komen, is de kans op gebruik ('utilization' of benutting) van de bevindingen in de ambtelijke en bestuurlijke sfeer groter. Ontbreekt deze vertaalslag naar gebruikers, dan moet het overall-rapport door de ontvangende gemeente *bewerkt* worden als natraject en dus als vervolg op de benchmark. De kans dat dit gebeurt, is niet erg groot maar groter in geval van gevoelde urgentie en bestuurlijke actualiteit.

Het gaat bij benchmarking niet alleen om het kiezen van een aantal vergelijkbare organisaties die participeren in de benchmark, maar ook om het daadwerkelijk vergelijken van organisaties op het punt van onderzoek (het object van benchmarking), het rangordenen van organisaties (en hun producten, diensten of processen) naar goede en minder goede scores, het analyseren van scoreverschillen en dus het begrijpelijk maken hiervan, het formuleren van 'best practices' en het aannemelijk maken waarom een 'best practice' er uitspringt. Dat zou benchmarking kenmerken, volgens de normatieve literatuur. Dat betekent: zo zou het moeten gaan. Een benchmark is dus niet iets dat men uit de losse pols schudt.

Daarmee is een aantal kwaliteitscriteria verkregen voor beoordeling van de kwaliteit van benchmarking.

Er is in de voorwaardelijke sfeer nog iets toe te voegen. De benchmarkonderzoeker moet niet zoveel betrokkenheid hebben met elk van de opdrachtgevers en zoveel maatwerk leveren dat het benchmarkonderzoek er vanuit vergelijkend perspectief onder lijdt. Van de andere kant moet een opdrachtnemer wel *interactie* plegen met de opdrachtgevers omdat de opdrachtgevers moeten meegroeien met het onderzoek. Meegroeien betekent ook zicht krijgen op de inhoud van de benchmark. Bij vroege en continue betrokkenheid kan de kans op later gebruik van de bevindingen toenemen. Blijft betrokkenheid en interactie achterwege, dan is de kans groot dat de 'gebruikers' het onderzoeksrapport na verschijning links laten liggen en het onderzoek nooit de bestuurlijke top bereikt en nooit aanleiding geeft tot een nota met actiepunten.

3.5 Hanteren van normen of een ander referentiekader

Op een punt gaan we nader in: het uitgaan van normen ter beoordeling van een object van onderzoek. Vergelijken met of vanuit normen of standaarden is nodig? 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. Een volgende stap is om te komen van een norm of referentiekader naar daarmee sporende indicatoren.

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. In een enkel geval wordt een streefcijfer gehanteerd dat afkomstig is uit gemeentelijke kwaliteitshandvesten, zoals de termijn waarbinnen een brief van burgers moet zijn beantwoord (VNG, 2006). 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 is ook mogelijk dat er wel een norm of standaard bestaat, maar dat die niet vanuit de praktijk is ontstaan. Of deze is wel ontstaan maar niet bruikbaar en/of niet erkend (Poerstamper e.a., 2007: 51). Als normen of standaarden niet erkend worden, kan dat een reden zijn om later een benchmarkuitkomst te negeren of terzijde te leggen.

Het gaat dus om vergelijken, om het zoeken naar de plaats van de organisatie te midden van een veld van vergelijkbare organisaties, om het begrijpelijk maken van verschillen in scores 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 en worden gespiegeld aan een streefcijfer. Dus, wie de show steelt, wordt bekeken.

Zo'n analyse in het licht van streefcijfers is van nut omdat het mogelijk is dat het collectief van vergeleken organisaties allemaal boven de streefwaarden zit of allemaal eronder. Als alle organisaties onder de kritische waarden zitten, is een 'beste' organisatie in feite de minst slechte. Als alle organisaties al boven de kritische waarden zitten, en het dus meer dan 'voldoende' doen, is de beste organisatie de superorganisatie. Dan is er een luxeprobleem.

3.6 Verbeterruimte en streefcijfers

Een voorbeeld van benchmarkinformatie met een 'best practice' is te vinden in figuur 3. Daaruit blijkt dat de verzorgingsinstelling *'La vie en rose'* op directe cliëntgebonden tijd bovengemiddeld scoort (met 77,36), wat positief is, maar dat het nog beter kan want de organisatie met de beste praktijk zit nog duidelijk hoger (met de score 108,00). Er is dus vermoedelijk wat we noemen *verbeterruimte*. We houden een slag om de arm want wellicht zijn er omstandigheden die bemoeilijken om op korte termijn te verbeteren, zoals bij voorbeeld een fusie van organisaties, waarin *'La vie en rose'* betrokken is, of een verhuizing (wat voor een verzorgingshuis een ingrijpende zaak is).

Fig.3: Voorbeeld van 'best practice'-informatie over aan cliënten bestede zorgtijd in verzorgingshuizen

Tijdscategorie	Uw instelling 'La vie en rose'	Gemiddelde score	'Best practice'
Direct cliëntgebonden tijd	77,36	68,64	108,00
Indirect cliëntgebonden tijd	10,45	9,48	17,72
Niet-clieñtgebonden tijd	23,7	21,90	28,37
Totaal	111,55	100,02	154,09

Bron: Poerstamper e.a., 2007: 15; Benchmark verpleeg- en verzorgingshuizen 2004/2005.

Over het opstellen van streefcijfers en het onderscheid naar typen streefcijfers 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). We noemen het probleem van de ontbrekende of eenzijdige normen in de vorm van streefcijfers, het probleem van de perverse effecten, het probleem van rommeligheid in het type informatie, en het probleem van slijtage. We lichten die problemen kort toe.

Ten eerste blijkt het niet zo eenvoudig om al het overheidsbeleid in duidelijke producten te specificeren en te komen tot streefcijfers. Onderwijs is bij voorbeeld vanuit gemeentelijk perspectief niet in hapklare brokken 'output' of prestaties uit te drukken, hoewel journalisten van het dagblad Trouw kwamen met kwaliteitskaarten. Hetzelfde geldt voor leefbaarheid of welzijn, stellen wij in de voetsporen van Bordewijk en Klaassen (2000). Het is soms wel mogelijk dat er prestatiescores zijn op te stellen maar dat de gemeentelijke overheid niet (goed) aan de knoppen kan draaien om de scores desgewenst omhoog te werken. Wellicht is het nodig om eerst een debat te hebben over de bevoegdheden. In dit verband valt te wijzen op de gemeente Amsterdam, waar wethouders zich zijn gaan bemoeien met de locatie van coffeeshops op afstand van scholen en waar een discussie ontstond over de (on)wenselijkheid van 'zwarte', 'witte' en 'gemixte' scholen. Tellen is een en over bevoegdheden en sturingsmiddelen beschikken is twee.

Ten tweede kunnen bij benchmarking perverse effecten optreden (De Bruijn, 2001, 2003, 2007; Van Thiel, 2009). 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 terwijl er in feite slechts sprake was van één uitvoering.

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

Fig. 4: 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 onderzoeker.

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).

Ten vierde, prestatie meting is aan slijtage onderhevig als gevolg van leereffecten (Van Thiel, 2009: 21). Als organisaties hun werkwijzen zo verbeteren dat ze steeds om en nabij honderd procent gaan scoren, is het tijd voor bezinning. Een volgende verbeteringsoperatie is in dat geval immers uitgesloten. Een wijziging in de keuze van indicatoren dient zich dan aan (verg. Abma en Korsten, 2009).

Conclusie: er zijn in verband met benchmarking nieuwe begrippen geïntroduceerd, zoals het begrip 'verbeterruimte', het begrip 'pervers effect' en het begrip 'slijtage'. Enkele van die begrippen verwijzen naar praktijken die voorkomen of beperkt moeten worden.

3.7 Eisen aan benchmarkonderzoek

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

De Commissie-Van Rijn stelt in 2007 over prestatiemeting bij woningcorporaties in de vorm van benchmarking dat sprake moet zijn van prestatie maatstaven die relevant, informatief en 'veelzeggend' zijn en 'het prestatieveld voldoende afdekken'.

Sandra van Thiel (2009) stelt meer in het algemeen: het aantal indicatoren moet niet te klein zijn. 'Het idee dat een indicator in één maat kan uitdrukken hoe een organisatie presteert, is een misvatting', stelt Van Thiel (2009: 21). Het is onmogelijk om een organisatie in één opslag te beoordelen. Een meetsysteem moet de prestaties 'volledig' in beeld nemen (Van Thiel, 2009: 21). Aanvullend wordt wel als eis gesteld dat er een 'goede balans moet zijn tussen het aantal indicatoren en de wijze waarop deze worden gebruikt om organisaties te (be)sturen' (Van Thiel, 2009: 20).

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 eisen rond de kwaliteit van de meting. Maar zij heeft ook oog voor relevantie en dat raakt aan functionaliteit en aan de communicatie rond benchmarkonderzoek. Daarmee raakt zij aan legitimiteit. Zij formuleert de volgende kwaliteitseisen:

- omvattendheid (comprehensive),
- begrijpelijkheid (understandable),
- validiteit van de maten (geldigheid: meten ze wat ze moeten meten?),
- actualiteit van de cijfers,
- betrouwbaarheid (reliability; blijkt bij herhaling de zelfde uitkomst?),
- 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 prestatiemetingssysteem of elke 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 van het meetobject erkend worden, dus van wat in onderzoek is. En waarbij 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 start en 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 (erkenning van de meting).

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.

3.8 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 (zie ook KING, 2010). Daarbij is vrijwel altijd sprake van een kwantitatieve, cijfermatige vergelijking (KING, 2010). Een benchmarkrapport staat dus bol van de cijfers (scores) en soms ook van zgn. spinnenwebben (een score op meerdere dimensies). Het is dan ook te begrijpen dat aan de oorsprong van benchmarking in de eerste helft van de vorige eeuw 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.

Benchmarking is sterk kwantitatief gericht. 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 toen op omdat de visitatiecommissies 'best practices' opspoorde en vermeldde in een overallrapport van de beoordeelde opleidingen (bijvoorbeeld alle wo-opleidingen bestuurskunde). In bestuurskrachtonderzoeken werd echter slechts sporadisch gebruik gemaakt van benchmarks (Abma, 2012: 54; Schutgens, 2014).

3.9 Doelen van benchmarking

Waarom benchmarking?

1. Benchmarking staat in functie van transparantie en verantwoording naar controlerende organen en de samenleving. Veel benchmarking in de publieke sector is daarom openbaar.
2. Benchmarkonderzoek appelleert aan de nieuwsgierigheid. Velen willen weten: waar staat mijn organisatie te midden van andere op het vlak van effectiviteit of doelmatigheid of andere criteria?
3. Scores die goed uitkomen, kunnen een motivatie-effect hebben op organisatieleden. Ze kunnen het elan bevorderen.
4. 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 officieel vooral 'spiegelen en verbeteren' (Van Tits, 2009: 239; Van Dooren, Bouckaert & Halligan, 2010). Het gaat om het leren.

Een voorbeeld van een benchmark is een woonaantrekkelijkheidsscan van gemeenten. Deze scan werd en wordt uitgevoerd door Bureau Louter. We lichten er hier een deel van een tabel uit om te laten zien hoe er kwantitatief te vergelijken en rangordenen valt. De naam van de gemeente waar het hier om gaat laten we weg omdat die naam er niet toe doet. De benchmark, zoals hier

weergegeven, bevat een vergelijking met andere gemeenten in de vorm van plaatsing van deze gemeenten in de rangorde van alle gemeenten. Het is een benchmark zonder een norm of kwaliteitsstandaard en zonder aanduiding van een 'best practice'.

Fig. 5: Weergave van een deel van de woonaantrekkelijkheidsmap van een gemeente in de Randstad

Onderwerp	Plaats van individueel in lijst van 418: rangorde	Bij beste 100 gemeenten van Ned.	Plaats in rangorde van 29 kernsteden
Historisch decor	Plaats 28	Ja	Plaats 5
Aantrekkelijke bebouwing in omgeving	Plaats 185	Nee	Plaats 13
Aanwezigheid van binnenwater (plas)	40	Ja	3
Parken en plantsoenen	119	Nee	12
Bos en natuur	261	Nee	18
Recreatiegebieden	156	Nee	9
Rust en ruimte	415	Nee	26
Huisartsen	7	Ja	6
Tandartsen	103	Nee	15
Ziekenhuisvoorzieningen	13	Ja	7
Basisonderwijsvoorzieningen	84	Ja	5
Kinderdagverblijven	3	Ja	3
Buitenschoolse opvang	6	Ja	4
Speelgelegenheid jonge kinderen	243	Nee	11
Middelbaar onderwijs	10	Ja	6
Hoger onderwijs	27	Ja	10
Voorzieningen voor jongeren	239	Nee	13
Winkelvoorzieningen (dag boodschap)	31	Ja	16
Luxe winkels	149	Nee	28
Musea	3	Ja	3
Theaters/podia	41	Ja	16
Bioscopen	4	Ja	4
Cafés	31	Ja	16
Restaurantvoorzieningen	6	Ja	6
Sportvoorzieningen	10	Ja	5
Burenoverlast	399	Nee	16
Overlast jongeren	391	Nee	21
Geluid/stankoverlast	406	Nee	27
Verkeersveiligheid	48	Ja	1
Fysieke veiligheid (weinig misdaden)	362	Nee	6
Saamhorigheid	414	Nee	25
Werkgelegenheid	120	Nee	19
Economisch sterke regio?	191	Nee	19
Bereikbaarheid per trein	16	Ja	8
Parkeermogelijkheden	396	Nee	22
Gemeentelijke dienstverlening	59	Ja	21

Bron: Bureau Louter, internet 2013

De in deze figuur opgenomen stad heeft een tamelijk duidelijk profiel. Het gaat niet om een stad in 't Gooi. Van landelijkheid, rust en ruimte is niet nadrukkelijk sprake. Het is een per spoor goed bereikbare stad met vrij veel voorzieningen op tal van terreinen (basisonderwijs, sportvoorzieningen, huisartsen, tandartsen, ziekenhuis, kinderdagverblijf), maar qua winkelen op luxe gebied weinig voorstelt. Het is een studentenstad, een stad met redelijk wat werkgelegenheid die het niet moet hebben van de saamhorigheid. De verkeersveiligheid is redelijk op orde maar qua overlast van burger of jongeren scoort de stad niet erg positief.

3.10 Vormen van benchmarking

Vermeldenswaard is dat benchmarking geen eenheidsworst is. Er bestaan verschillende vormen van benchmarking. In de literatuur wordt daar melding van gemaakt en wordt gesteld dat het waarnemen van deze variëteit van belang is (Francis & Holloway, 2007: 174-176; Hakvoort & Klaassen, 2010: 113; Poerstamper e.a., 2007: 47-57).

Elke benchmarkstudie is te typeren op basis van een aantal onderscheidingen. Aan de literatuur ontleen we de volgende onderscheidingen in soorten benchmarkonderzoek.

Een eerste onderscheid van benchmarkvormen is dat in een benchmark vanuit een leerstrategie van de organisatie of een benchmark waarbij die strategie ontbreekt. Verder valt te onderscheiden naar 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 overallrapport voor alle deelnemende partnerorganisaties of een overallrapport plus een rapport voor elke deelnemende organisatie; een rapport dat wel of niet openbaar is.

Fig. 6: Aspecten van benchmarking die leiden tot vormen van benchmarking: onderscheiden dimensies

1 Benchmarking is het gevolg van een leerstrategie van organisaties	of	Benchmarking is geen gevolg van een leerstrategie van organisaties
2 Vrijwillige deelname door deelnemende organisaties aan benchmarking	of	Verplichte deelname aan benchmarking
3 Leiding van benchmarking door visitatiecommissie	of	Leiding door een onderzoeksbureau
4 Onafhankelijke benchmarking t.o.v. deelnemende partnerorganisaties	of	Coöperatieve deelname van partnerorganisaties
5 Externe vergelijking tussen organisaties	of	Interne vergelijking binnen een organisatie
6 Partiële benchmark /aspectmatig	of	Integrale benchmark
7 Resultaat- en procesanalyse	of	Resultaat- of procesanalyse
8 Cijfers over meerdere jaren: vergelijking in de tijd	of	Cijfers over slechts een jaar
9 Vermelding van 'best practices' in het rapport	of	Geen vermelding van 'best practices'
10 Verklaring van verschillen tussen scores	of	Geen verklaring van verschillen
11 Overallrapport en rapport per partnerorganisatie	of	Alleen een overall-rapport; geen rapport per partnerorganisatie
12 Openbaarheid van rapportage	of	Geen openbaar rapport

Er is volgens ons sprake van een *favoriet benchmarking-onderzoek*. Dat is een benchmarkonderzoek met de volgende kenmerken, namelijk:

- met een duidelijke probleemstelling waar opdrachtgevers achter staan;
- passend binnen de leerstrategie van een organisatie (de relevantie van het onderzoek wordt erkend);
- geleid door een onafhankelijke commissie of onderzoeker met 'branchekennis' en affiniteit tot de professionaliteit, die in de vergeleken organisaties aan de orde is;
- met een vergelijking van prestaties van vergelijkbare organisaties en een verschillenanalyse gericht op het begrijpen van het waarom van de hogere en lagere scores;
- met een verantwoorde onderzoeks aanpak (relevante en essentiële indicatoren, zoveel mogelijk actuele gegevens, geldige bevindingen, betrouwbare bevindingen, etc.);
- met een aanduiding van (relevante) 'goede praktijken';
- conclusiegericht in zijn bevindingen en gebruikgericht;
- met een openbaar overallrapport;
- en een openbaar rapport voor elke afzonderlijke deelnemende organisatie.

Een integrale benchmarking van organisaties, zoals van ziekenhuizen, is minstens zo uitdagend als een partiële vergelijking omdat deze alomvattend is en dus een beeld zou kunnen geven van de organisatie als totaal. Bovendien worden organisaties als totaal onderling vergeleken (Abma en Korsten, 2009).

3.11 Thuiszorginstellingen als voorbeeld

We gaan nu aan de hand van een voorbeeld kijken naar hoe het er met benchmarking aan toe gaat. Daarmee verkrijgen we aanvullend inzicht en kunnen we enkele grenzen of beperkingen van benchmarking illustreren. Dat het voorbeeld al wat ouder is, het dateert uit 1997, doet voor het doel van dit betoog niet ter zake.

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

Hoe kwamen de onderzoekers nu tot hun conclusies? Ze verzamelden kengetallen over productie, ingezette middelen, proceskenmerken en hun onderlinge verhoudingen. Er is onder meer bezien wat het aandeel uitvoerend en niet-uitvoerend personeel is, en wat de relatie is tussen het productieve en niet-productieve personeel. Zorguren van wijkverpleging en gezinsverzorging zijn in verband gebracht met het totaal aan personeel of met het budget. Niet-leidinggevenden zijn gerelateerd aan het totaal van zorgpersoneel. Dat levert verhoudingscijfers op. Voor *ieder* vergelijkbaar punt voor de zeven deelnemende instellingen is een 'best practice' aangeduid met *vet* in de figuur. Het kan een minimum- of een maximumkengetal zijn.

Wat is nu de meest doelmatige instelling? Verrassend, die bestaat in werkelijkheid niet. Wel bestaat de *virtuele* doelmatige instelling, opgebouwd uit alle verzamelde 'best practices'. De beste organisatie is dus een *constructie* vanuit onderzoek. De gevolgen hiervan zijn groot. Dat betekent namelijk dat een organisatie die zich wil verbeteren, niet ergens kan gaan kijken en vragen 'hoe deden jullie dat?' Maar vanuit de goede praktijk kan elke organisatie wel zelf gaan nadenken, hoe zichzelf te verbeteren. Kan men die operatie opbrengen?

Fig. 7: Kenmerken van het productieproces bij zeven onderzochte instellingen voor thuiszorg vlg. Elias e.a. (1997)

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

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

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

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

Wat weten we nu? Ten eerste kan men de vraag stellen of de variabelen goed gekozen zijn. Is de doelmatigheid zeer gediend met een gering percentage niet-uitvoerend personeel, bij wijze van spreken met een kleine directie? En is het inderdaad uit een oogpunt van doelmatigheid zonder meer gunstig dat de verhouding tussen facilitair en uitvoerend personeel zo minimaal mogelijk is? Is het meest minimaal, zoals bij instelling I, niet te minimaal? Als er weinig faciliteiten zijn, bestaat de mogelijkheid dat uitvoerend personeel te veel zelf moet doen en niet aan zijn eigenlijke taak toekomt, aldus Jos Blank (1998). Het is nu wel duidelijk: benchmarking impliceert normatieve keuzen van kengetallen. De les is: het is steeds nodig bij benchmarking om de keuze van variabelen te verantwoorden! Dat geschiedt in tal van op toepassing gerichte benchmarkrapporten niet, ook niet bij gemeentelijke. Gevolg: de lezer moet gissen naar de relevantie van de vergeleken indicatoren voor een product(cluster), dienst of proces. Dat de kern gepakt wordt, is lang niet zeker. De kans bestaat dat een gebruiker die een rapport leest alleen al om die reden een rapport links laat liggen.

Wat ook blijkt, is dat een beste praktijk zuiver vanuit één cijfer bezien nog niet veel zegt. Het gaat erom in een benchmark duidelijk te krijgen waarom de praktijk van organisatie I of II of welke dan ook in een bepaald opzicht eruit springt. Neem een ander voorbeeld. Stel de volwassen burgers van een gemeente K geven deze gemeentelijke organisatie voor dienstverlening een cijfer 7.8 op een schaal van 1-10 en andere organisaties scoren alle een 7.0 of minder maar wel 'voldoende'. Weten we dan waarom die gemeente K zo goed scoort? Nee, misschien functioneert de gemeentelijke organisatie inderdaad wel beter, maar in welk opzicht dan?

Een alternatieve verklaring voor de 7.8 houdt in dat de burgers in die gemeente in het algemeen milder in hun oordelen zijn – bij voorbeeld omdat de bevolking in de 'bible belt' woont (de geografische strook van Tholen, Bunschoten, Veenendaal en Zwartewaterland tot Rijssen) en over het algemeen erg christelijk is en positief in het leven staat - en de burgers de gemeente meer goedgezind zijn dan geldt voor burgers in andere gemeenten.

3.12 Grenzen van benchmarking

We waren ook uit op zicht op grenzen van benchmarking. We hadden er al een ontdekt, namelijk dat de beste organisatie een constructie is uit het geheel van factoren dat inwerkt op scores.

Welke grenzen of beperkingen zijn in dit voorbeeld van thuiszorg te vinden? We noemen er hier vier.

Verantwoording: Een eerste grens, zo men wil een dreigende beperking, van benchmarking is het achterwege blijven van de verantwoording van de variabelenkeuze. Dat kan tot gevolg hebben dat een instelling denkt positief te scoren op een kengetal en daarmee een 'best practice' 'in handen te hebben' terwijl dat in werkelijkheid 'fake' is omdat het kengetal niet wijst op wat men wil meten. De variabele is dan in feite niet geldig.

Relevante variabelen: Een tweede grens betreft de vraag of de variabele waarop scores vergeleken worden wel relevant is. In het veld wordt hoog scoren op zo'n variabele als goed beschouwd, maar waarom eigenlijk? Een voorbeeld. Het wordt ten tijde van het onderzoek als 'goed' beschouwd om veel variabelen personeel in dienst te hebben in de thuiszorg en weinig parttimers. Waarom is dat zo

goed? Variabel personeel beschikt wellicht over minder 'verzorgende kwaliteiten' dan de parttimers. Onderzoek toont dat parttimers juist de doelmatigheid, althans in ziekenhuizen, bevorderen (Blank, 1998: 433).

Oorzaken van verschillen: Een derde grens van benchmarking is dat de benchmarking in dit voorbeeld niets zegt over de oorzaak of achtergronden van verschillen tussen de scores van organisaties op een score, bij voorbeeld een kengetal. Of schaafeffecten een rol spelen blijft duister. De grootte van instellingen kan een relevante factor zijn omdat bij grote organisaties de overhead in verhouding tot kleine organisaties waarschijnlijk procentueel geringer is. De missie van de organisatie en de werkfilosofie wordt ook niet achter de cijfers zichtbaar.

De echte organisatie die het beste is: Een vierde grens impliceert dat de benchmarkonderzoeker niet, bij een gegeven productie, op zoek gaat naar welke organisaties 'van vlees en bloed' verschillen in de manier waarop ze de productie realiseren. Zou dat wel gebeuren dan komen er aanwijzingen naar voren of de ene organisatie dat resultaat anders bereikt dan de andere.

In de praktijk van benchmarking wordt overigens soms wel degelijk gezocht naar welke organisaties er op tal van variabelen kwalitatief gunstig uitspringen. Bekend is het beeld van 'het Tilburgs model' met contractmanagement, waarbij sprake was van outputsturing, daaraan gerelateerde contracten tussen de bestuurlijke en ambtelijke top en decentrale units, en managementrapportages als vorm van terugkoppeling over bereikte prestaties. Organisaties die veel benchmarks hebben uitgevoerd, zoals PwC, zoeken ook in de sfeer van de thuiszorg, de verzorgings- en verpleeghuizen, de gehandicaptenzorg en het beroepsonderwijs naar clusters van variabelen die maken dat een bepaalde organisatie of een top drie of vijf er gunstig uitspringt. Dit streven stelt wel eisen aan de opzet van de benchmarking. Met een benchmark van 15-20 deelnemende organisatie is dit doel niet te realiseren. Er zijn al gauw minstens 100 deelnemers nodig om via een kwantitatieve analyse inzichten in clustering van variabelen te verwerven. Anders gesteld, hoe meer bouwstenen nodig zijn om te bepalen wat een goede organisatie is, hoe groter het aantal aan een benchmark deelnemende organisaties moet zijn (Poerstamper e.a., 2007). Lukt het dan makkelijk? Zeker niet. Bij veel benchmarkonderzoek is het aantal deelnemende organisaties echter juist klein, denk aan 15-20, om te bereiken dat verschillen in scores (wellicht) te achterhalen zijn in bij voorbeeld groeps gesprekken waarin resultaten aan de orde komen. Maar of in die gesprekken dan diep gegraven kan worden en wordt, is een volgende kwestie.

3.13 Progressie in benchmarking: verpleeghuizen als domein

Met grenzen eindigen, vinden we niet gelukkig. We bieden graag perspectief. We gaan daarom nu in op verbetering van de benchmark-aanpak en kiezen als invalshoek doelmatigheid. Wat is qua doelmatigheid het beste verpleeghuis?

De aandacht voor doelmatigheid van dienstverlening is in de jaren negentig van de vorige eeuw en daarna toegenomen. Een grotere doelmatigheid kan voordelen hebben, zoals bij voorbeeld het ontstaan van kortere wachtlijsten voor klanten of patiënten. Het meten en spiegelen van instellingen - die eenzelfde product of dienst leveren - aan elkaar is daarbij een hulpmiddel gebleken. Dat vergelijken, rangordenen en verklaren van verschillen in prestaties heet benchmarken. Te onderscheiden is in zinloos en zinvol vergelijken. Van zinloos vergelijken hebben we geen voorbeelden gegeven. Wel gaven we aan dat benchmarking door *het opsporen van een virtuele, doelmatige organisatie niet genoeg is*. Immers, die *bestaat niet*. De kunst is: om juist bij eenzelfde hoge productie de verschillen te begrijpen en te verklaren. Die operatie levert gegevens op over soorten ondoelmatigheid en aanzetten (nog geen actieprogramma) over hoe de doelmatigheid te vergroten.

Het Sociaal en Cultureel Planbureau (SCP) denkt op dit punt met benchmarking progressie geboekt te hebben. Waarom? Het SCP scoort niet slechts de 'best practice' per kengetal voor een (virtuele)

organisatie. Het SCP probeert daadwerkelijk een ‘best practice’ van een bestaande instelling te vinden, een gunstig patroon van productie dus, en dit te verklaren. De onderzoekers spreken van de data-omhullingsmethode (Blank, 1998; Blank en Eggink, 1998). Hierbij worden instellingen met eenzelfde productie met elkaar vergeleken. In onderzoekstermen gesteld: de afhankelijke variabele wordt dus gelijk gehouden. Onderzoek met deze aanpak bij verpleeghuizen laat zien dat dezelfde productie met 30 procent minder middelen gerealiseerd kan worden. Er is blijkbaar inefficiëntie. Daarvan zijn drie vormen bekend: allocatieve, technische en schaal (in) efficiëntie.

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

Volgens de SCP-benadering kan met een regressie-analyse voor elk proceskenmerk worden nagegaan wat de bijdrage is aan de efficiëntie van de instelling. Wat beïnvloedt de technische efficiëntie van een verpleeghuis nadelig? Het SCP-onderzoek ‘Zuinig op zorg’ toont dat een hoog aandeel van leerlingverpleegkundigen, een hoog ziekteverzuim, veel in- en uitbesteed werk en grote afdelingen de technische efficiëntie op een negatieve manier beïnvloeden, dus verslechteren. Een positief effect op de doelmatigheid gaat uit van hooggekwalificeerd personeel en een hoge bezettingsgraad (Blank en Eggink, 1996).

Fig. 8: Efficiënties van verpleeghuizen vlg. SCP-onderzoek

Totale efficiëntie	0.70
Waarvan: allocatieve efficiëntie	0.85
Waarvan: schaalearciëntie	0.94
Waarvan: technische efficiëntie	0.88
Bijdrage kenmerken productieproces aan technische inefficiëntie	
Aandeel leerlingverpleegkundigen	-0.12
Kwaliteitsprijs gediplomeerde verpleegkundigen	0.22
Kwaliteitsprijs overig personeel	0.20
Eigen apotheek	0.01
Bezettingsgraad	0.30
Ziekteverzuim	-0.05
Uitbesteed werk	-0.05
Inbesteed werk	-0.03
Afdelingsgrootte	-0.04
bron: SCP, Zuinig op zorg, 1996; Blank, 1998.	

Daarmee komen we bij een volgende conclusie. Als het nodig is om bij benchmarking steeds zo te werk te gaan, wordt benchmarking werk van superspecialisten op het gebied van methoden en technieken van sociaal-wetenschappelijk onderzoek. Wat het SCP liet zien als fundamenteel benchmarkonderzoek, is te onderscheiden van ‘toegepast onderzoek’.

Onze studie gaat hoofdzakelijk over het laatste. Daarin komen geen hoogstaande kwantitatieve analyses voor. Vanuit dat gegeven zullen de meeste benchmarkonderzoeken moeten worden bekeken.

4 Voordelen en nadelen van benchmarking

Benchmarking past geheel in de 'Code voor goed openbaar bestuur' zoals die door het kabinet-Balkenende IV is vastgesteld want elke overheidsorganisatie heeft volgens die code de opdracht om te 'leren', om zich steeds weer tijdig te verbeteren. Benchmarking is zo gezien niet meer weg te denken. Het is een instrument dat een opstap biedt om van een organisatie 'een lerende organisatie' te maken. Dan is een volgende stap nabij. Het instrument benchmarking neemt aan belang toe. 'Public sector benchmarking is of increasing importance' schreven Van Helden & Tillema (2005: 337) al. Het aantal benchmarkonderzoeken is internationaal zelfs spectaculair toegenomen.

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

De groei in benchmarking zit vooral in de toepassing in bepaalde sectoren van (semi-)overheidshandelen en met name in organisaties met kenmerken van massaproductie in een gesmeerd lopende '*machinebureaucratie*'. Denk aan een bureau Burgerzaken van gemeenten (verstrekking paspoorten etc.) en aan verzorgings- en verpleeghuizen. Benchmarking van tamelijk unieke organisaties, die in de typologie van organisaties van Henry Mintzberg andere namen krijgen, zoals bureaus voor onderzoek en statistiek, inlichtingendiensten en architectenbureaus komt zelden voor. Overal waar *verbeterruimte* in de taakuitvoering aanwijsbaar is, daar kan benchmarking een richtingwijzer zijn waar je die verbetering moet zoeken. De zwakkere organisatie kan zich optrekken aan de betere.

In 2017 meldt de Nederlandse Zorgautoriteit (NZa), het belangrijkste bestuurs- en toezichtorgaan in de gezondheidszorg, dat minder goede verpleeghuizen eerst doelmatiger moeten gaan werken voor een kabinet na het initiatief van Carin Gaemers en Hugo Borst (het manifest), dat aansloeg in de Tweede Kamer, extra in deze ouderenzorg investeert. De leiding van verpleeghuizen moet minder budget uitgeven aan gebouwen en aan management en meer aan directe zorg (de zogenaamde 'handjes' en 'zorg'). Dat was de boodschap van de NZa. Volgens de NZa moeten de minder goed functionerende verzorgings- en verpleeghuizen een voorbeeld nemen aan de goede: die goede steken minder geld in overhead en zetten personeel doelmatiger in dan de slechtere. De Inspectie voor de Gezondheidszorg had in 2016 in een rapport over 150 verpleeghuizen al vastgesteld dat een derde blijvend intensief toezicht moest hebben. Bij elf verpleeghuizen was de situatie ronduit zorgwekkend. De NZa wil dat matig of slecht presterende verzorgingshuizen, die veel *verbeterruimte* kennen, in de leer gaan bij goed functionerende die evenveel budget ontvangen. De slechtere moeten zich optrekken aan de betere huizen met betere zorg. Gemiddeld gaat bij verpleeghuizen iets meer dan de helft van het toegekende budget naar de zorg, dertig procent gaan naar leidinggevenden en overhead en de rest naar verblijfkosten. Er blijken tussen de organisaties grote verschillen in deze kosten, aldus de NZa.

Benchmarking kan worden toegepast bij gemeenten of bij onderdelen daarvan, maar evenzeer in de hoek van onder meer de volgende (semi-)publieke organisaties: ziekenhuizen, verpleeghuizen, thuiszorginstellingen, huisartsenpraktijken, onderwijsorganisaties, politieorganisaties, woningcorporaties, afvalverwerking, pensioenfondsen en waterschappen. Hier is sprake van grote aantallen. De voor- en nadelen van benchmarking moeten we dan ook in uiteenlopende sectoren kunnen vinden.

4.1 Veronderstelde voordelen

Benchmarking is een veel toegepast instrument om via onderzoek gericht op vergelijking en rangordening van prestaties te komen tot verbetering (o.a. Poerstamper e.a., 2007). Vergelijking zegt nog niet alles, het gaat om de waarneming van *achterblijven* ten opzichte van een referentiemaat. In de wereld van verpleeg- en verzorgingshuizen bestaan er kwaliteitseisen voor 'voldoende personeel'. De mogelijkheid van verbetering valt te benutten als er 'betere' organisaties zijn.

Ook wie het al goed doet, kan mogelijk verder verbeteren, hoewel er op een bepaald moment sprake kan zijn van de *wet van de afnemende verbeteringsmogelijkheid*. Om verbetering van beleid of correctie van een beleid of een beleidspraktijk is het te doen want het sanerend feedbackeffect van marktwerking ontbreekt in de publieke sector.

De verbeteringsambitie rekenen we hier tot de kenmerken van benchmarking en daarom kunnen we verbetering sec geen voordeel noemen. Maar we rekenen de aard van bepaalde verbeteringen wel tot de potentiële voordelen. Het is zinnig om bij anderen over de schutting te kijken want dat voorkomt dat je zelf het wiel opnieuw probeert uit te vinden. Verderop gaan we daarop verder in.

We noemen hier nog eens enkele voordelen van prestatiemeting in de vorm van benchmarking. Daarbij betrekken we dat zowel sprake kan zijn van voordelen op het niveau van een individuele organisatie die deelnam aan een benchmark (microvoordelen) als op het niveau van een sector (mesovoordelen). Dit is een algemeen aanvaard inzicht. Bij benchmarking van corporaties komt niet alleen zicht op goede praktijken bij een individuele corporatie naar voren, als die er tenminste is, maar ontstaat een beeld van de branche. Dat geldt ook voor sectoren als verpleeghuizen en thuiszorgorganisaties.

Kunnen ook organisaties die niet deelnamen aan een benchmark profiteren van benchmarking? Dat is zeker niet uitgesloten, maar die mogelijkheid krijgt hier geen speciale vermelding. In de literatuur wordt dit item nooit aangeroerd.

We komen bij de volgende voordelen van benchmarking.

1 Het opstapargument: naar nadere analyse en verbeteringsdiscussie

Benchmarking impliceert dat een organisatie zich te weer stelt tegen dreigende en al te sterke naar binnen gerichtheid. Benchmarking activeert juist het zelfoplossend vermogen van deelnemende organisaties. Ze zoeken immers door middel van benchmarking naar een goede praktijk en naar verbetering van of in de aanwezige verbeterruimte.

Diverse inherente aspecten van benchmarking zijn al genoemd, zoals de aanzet tot transparante verantwoording en verbetering in wat een organisatie doet en voortbrengt (ook Van Dooren, Bouckaert & Halligan, 2010). Benchmarking kan feedback leveren over hoe beleid uitpakt(e) en daarmee - heel concreet - tal van bespreekpunten opleveren voor verbetering, bij voorbeeld in de zin van meer effectiviteit van beleid of meer doelmatigheid.

Een voorbeeld. Een vergelijking van mortaliteitscijfers van ziekenhuisafdelingen kan zichtbaar maken of een afdeling eruit springt. Na een eerste reactie van 'niets aan de hand' of 'verouderde cijfers' kan dan toch, na nader onderzoek, blijken dat op de afdeling sprake is van slechte organisatie en/of slechte samenwerking, waarmee een opstap naar verbetering gezet kan zijn en verantwoordelijken aangesproken worden. Benchmarking kan dus een richtingwijzer voor nader onderzoek zijn (De Bruijn en Van Helden, 2007).

2 Het preventieargument: benchmarking levert vroege waarschuwingen

We noemen een tweede voordeel: benchmarking kan vroege waarschuwingen ('early warnings') verschaffen. Dit is niet vanzelfsprekend zo, maar wel als een organisatie een goede selectie van wat hoe aan benchmarking wordt onderworpen, en bij de start van benchmarking dus niet afwacht tot een benchmarkaanbieder zich meldt.

Een voorbeeld. Een analyse en vergelijking van de terugloop in aantallen huishoudens en geboorten in wijken en dorpen ('krimp') kan een gemeentebestuur zicht geven op het aantal en typen in omvang bedreigde scholen in de gemeente of regio en op het moment waarop samenvoegings- en sluitingsvraagstukken zich voor schoolbesturen aandienen. Dreigt sterke terugloop, dan leidt dat weer tot discussie onder ouders van schoolgaande kinderen en anderen. Sommige ouders kunnen het

proces van schoolsluiting versnellen door kinderen al eerder van school te halen en elders te plaatsen, met mogelijk als gevolg animositeit onder de bevolking omdat het van school halen van leerlingen niet in dank wordt afgenomen door de ouders die proberen de school te redden. Dit doet zich hier en daar feitelijk ook voor. Kortom, benchmarking kan een 'early warning'-effect hebben dat aan een discussie over het voortbestaan van een school vooraf gaat. Die discussie betreft niet altijd 'verbetering', maar kan ook gaan over opvang of kanalisatie van 'Verelendung' (zie Rob, 2008; Dijkstal en Mans, 2009; Oevering, 2010; Reverda, 2011; Hospers en Reverda, 2012).

3 Het argument van de samenhang van prestaties

Een benchmark brengt prestaties in kaart maar daar blijft het niet bij. Een ideale benchmark geeft ook een beeld van de samenhang van prestaties. En dat is aantrekkelijk want het verschaft een benchmark toegevoegde waarde. Oordelen kunnen nu gestoeld worden op feiten.

Een voorbeeld. De Benchmark thuiszorg 2004 heeft laten zien dat een positief cliëntoordeel samengaat met een hoger percentage kosten voor leidinggevend, maar ook dat dit verband boven een bepaald kostenvoordeel niet meer bestaat. Een les hieruit is: het is nuttig dat er leidinggevend zijn; bezuinig niet te gretig en rigoureuus op leidinggevend want zo een bezuiniging kan de kwaliteit van medewerkers in de frontlinie aantasten.

Tevens bleek dat medewerkers die positief zijn over hun werk in organisaties werken waar leidinggevend duidelijk sturen en de bedrijfsvoering en daarmee de planning & control op orde hebben (Poerstampere a., 2007: 17). Uit een andere benchmark, namelijk de Benchmark verpleeg- en verzorgingshuizen kwam naar voren dat hoe kleiner de 'span of control' van leidinggevend uit het middenmanagement is, hoe positiever het cliëntoordeel in die huizen is. Teams moeten niet te groot zijn. Bij grote teams daalt het cliëntoordeel. Begrijpelijk want dan wordt de hulp anoniemer. Dan heeft een cliënt veel meer wisselende contacten en kan de cliënt 'de namen van verplegers en andere medewerkers niet (gemakkelijk) meer onthouden'. Les: hou de teams overzichtelijk en hou de afstand met cliënten beperkt. Het gevaar is dat bij grote teams de leiding weer gaat differentiëren. Voor je het weet komen er naast de teammanager clustercoördinatoren in actie en weet de cliënt niet meer wie voor wat verantwoordelijk, wie beslissingen neemt en wie voor wat aanspreekbaar is.

Dit is een interessante les want die lijkt in de praktijk van alle dag op tamelijk grote schaal genegeerd te worden.

4 Het resultaatargument: benchmarking leidt tot opschoning van organisaties

Benchmarking kan gericht zijn op tal van aspecten van producten, diensten of processen, zoals de financiën. Het gevolg van vergelijking tussen organisaties, rangordening en het zoeken naar goede praktijken kan zijn dat zicht ontstaat op al degenen die niet in het primair proces betrokken zijn en op de omvang en het type 'indirecten'. Te veel managers, te veel secretariaatsmedewerkers, teveel ondoelmatigheid? Een te grote facilitaire dienst? Een opschoning van bureaucratie ('debureaucratisering') behoort dan tot de mogelijkheden (De Bruijn en Van Helden, 2007; Huijben en Geurtsen, 2008).

5 Het hrm-argument: benchmarking motiveert en geeft energie

We komen bij een volgend potentieel voordeel. Er wordt wel gezegd dat de deelname aan benchmarkonderzoek en het rapporteren over bevindingen aan medewerkers van de deelnemende organisaties positieve energie geeft (Linker, 2009). De methode zou motiverend werken voor betrokken overheidspersoneel dat in benchmark een opstap ziet voor een ontwikkeling naar een 'lerende organisatie' (Francis & Holloway, 2007: 181). Zo gezien, raakt benchmarking aan personeelsmanagement (hrm of 'human resources management').

Er wordt overigens ook opgemerkt dat prestatie meting getuigt van een ongezonde meetlust en in het algemeen demotiverend werkt, zeker ook omdat het de professionaliteit van werknemers ontkent. Denk aan docenten van laag scorende scholen met veel kansarme kinderen. De kans is aanwezig dat dergelijke scholen nooit bovengemiddeld gaan scoren in een benchmark van

schoolprestaties hoe hard de docenten ook werken. Om deze reden houden directeuren hier en daar scores ook weg bij docenten. Of ze roepen om opname in de benchmark van een correctiefactor (Tonkens en Swierstra, 2008: 123). Zo kan de benchmarking zelf een object worden van politieke strijd en lobby.

6 Het argument van systeemdwang

Een ander potentieel voordeel van benchmarking verwijst naar een sneeuwbal effect waarvan niet helemaal zeker is of het een echt voordeel is. Wat wordt bedoeld?

Een specifieke discussie over benchmarking betreft de vraag of een organisatie zich in een bepaalde sector wel kan onttrekken aan deelname. Als een organisatie eenmaal een jaarverslag 'met prestatiecijfers' is gaan publiceren, of een document in het kader van onderwijsvisiteatie of in een strategische dialoog met een derde zoals een ministerie of zorgverzekeraar, dan kan het gevolg zijn dat prestaties gerankt gaan worden. Of men wil of niet. Als een onderzoeker besluit om Mba's in een rangorde te gaan plaatsen en beslist om die openbaar te maken, krijgen betrokken organisaties een belang bij de uitkomst en vervolgtkomsten. Dan kan het zaak zijn om gewoon mee te doen en ook op prestaties te gaan sturen die in metingen een rol spelen. Voor ziekenhuizen, scholen en universiteiten speelt dit. Maar deelname is toch vrijwillig? Dat is maar de vraag.

In bepaalde sectoren ontstaat druk om mee te doen. Universiteiten zijn jaren geleden al beland in een traject van kwaliteitszorg en prestatie meting. De in Europees verband door ministers van onderwijs opgestelde *Bologna-verklaring* versterkte dit (Douma, 2004; Dittrich, 2010; NVAO, 2010). Als in professionele sectoren dan toch prestatie meting plaatsvindt, kan de sector er een belang bij hebben om te voorkomen dat een 'doods' systeem ontstaat of 'aftands' afrekenen het gevolg is. Een perfectioneringslag komt dan binnen bereik. Een verbetering die ook op de werkvloer erkend wordt? Is een sector hiermee gediend? Universiteiten en hbo-instellingen hebben zelf een systeem van verantwoording geschapen omdat de besturen van mening waren dat de inspectie hoger onderwijs de beoordeling van prestaties toch niet aan zou kunnen. Er moest iets gebeuren en een alternatief diende zich niet aan.

Zo ontstaat een *haasje-over-effect*: men wil geen prestatie meting, wordt dan gedwongen mee te doen, tracht er vervolgens het beste van te maken en gaat tenslotte sturen op prestaties. Worden de voordelen van benchmarking zo bereikt? Niet uitgesloten is dat de prestatie cijfers inzichten verschaffen in onderwijssterkten en -zwakten die tot dan slechts intuïtief werden geïdentificeerd.

De Vries (2005: 7-8) meent overigens dat iets grondig mis is met dergelijke nieuwe verantwoordingspraktijken: geldverslindend, betekenisloos (omdat er volgens hem weinig gevolgen aan worden verbonden) terwijl onduidelijk is of de prestaties geproduceerd worden onder verantwoordelijkheid van degenen die de benchmarking laten verrichten.

7 Het zelfreinigingsargument

De noodzaak van zelfreinigend vermogen is opgenomen in de *'Code voor goed openbaar bestuur'*. Een potentieel voordeel van benchmarking kan zijn dat een sector zelfreinigend vermogen laat zien en daarmee zelfreflectie toont.

Een voorbeeld. Jaren terug zijn woningcorporaties als gevolg van de bruteringsoperatie door de centrale overheid op afstand gezet. De corporatiebesturen kregen meer vrijheid van handelen. Die vrijheid werd op allerlei manieren benut. Sommige corporaties gingen hun rol in de discussie over leefbaarheid van (pracht- of kracht-) wijken intensiveren en zo in vrijheid de organisaties inrichten op verbetering van de leefbaarheid, samen met andere partners. Er waren ook corporaties die volop investeerden in woningbouw en zelfs grote investeringsrisico's namen, wat leidde tot maatschappelijk debat over de vraag of corporaties wel deden waarvoor ze waren opgericht. Er dreigden ingrepen van de rijksoverheid, die later ook volgden (onder andere naar aanleiding van de 'Vestia'-zaak). Vanuit een koepelorganisatie (Aedes) of vanuit 'frontrunners' in de sector kon en kan dan de suggestie naar voren komen om te proberen nieuw centraal overheidshandelen af te wenden door zelfvisiteatie van

corporaties of benchmarking te starten (Ter Bogt, 2005; commissie-Van Rijn, 2007; Van Overmeeren, 2007). De redenering kan dan luiden: de sector kan beter zelf orde op zaken stellen, dan dat afgewacht wordt tot een kabinet overgaat tot regelgeving, waarvan nog maar de vraag is of de sector 'daarmee uit de voeten kan'.

8 Alleen voordelen; niemand is tegen

Na vermelding van deze voordelen, lijkt het antwoord op de volgende vraag duidelijk: wie is tegen benchmarking? Raden van toezicht van organisaties voor middelbaar beroepsonderwijs blijken zelden tegenstander van deelname aan een benchmark (Boers en Van Montfort, 2009; Douma, Steven en In 't Veld, 2012; Stevens, 2012). Gemeentebesturen zijn ook vrijwel nooit tegen. Waarom ook. Benchmarking past in het denken van politiek links tot politiek rechts, zo lezen we in de internationale literatuur (Triantafillou, 2007: 830). En waarom zou men tegen zijn als een *Code voor goed openbaar bestuur* (gepubliceerd door het kabinet-Balkenende IV) stelt dat overheden, alle overheden, steeds moeten proberen om zich te verbeteren. Dan grijp je alles aan en daar hoort benchmarking dan ook bij.

Benchmarking is een *hoerabegrip*; je kunt moeilijk op benchmarking tegen zijn. Benchmarking is misschien mede daarom ook wel 'vermaatschappelijk'. Het verlangen naar rangordes blijkt immers groot. Benchmarking prikkelt 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.) zelfaan 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 beste buurten (27 juni 2015), de veiligste gemeente, de innovatiefste regio, de beste ziekenhuizen (19 nov. 2016), de beste middelbare scholen en de beste universitaire opleidingen. Blijkbaar bestaat hier een behoefte aan. 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 of andere organisaties bovenaan staan en andere onderaan. Want alleen als je daarover veel achterhaalt, zijn in principe verbeteracties mogelijk die te betitelen zijn als '*benchlearning*' (cf Ten Tije e.a., 2010; Blank e.a., 2011) en het daadwerkelijk verbeteren als '*benchaction*' (Ten Tije e.a., 2010).

4.2 Veronderstelde nadelen

Er blijken ook wel enkele problematische kanten. Een kritiek is dat benchmarking niet ambities oproept maar verdrijft. Het grote dromen is als een wolk die overdrijft omdat benchmarking dwingt tot het realisme van de blik op de concurrenten of anderszins vergelijkbare organisaties (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 is teveel gericht zijn op prestaties en te weinig op gewenste effecten;
- het normalisatie-argument: benchmarking leidt tot uitbanning van innovatie, variëteit en pluriformiteit en juist een 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, die zou moeten verbeteren overschat wordt?;
- het argument van schijnsucces door strategisch gedrag rond normen (het goedkoop scoren van makkelijk te behalen resultaten);
- de groei van de bureaucratie door administratieve lasten rond deelname aan benchmarking;
- permanente veranderingen door juist veel benutting van benchmarkuitkomsten;
- het schuttersputjes-argument;

- schadelijkheid van prestatiemeting voor professionaliteit?

We lichten deze nadelen toe.

1 Het argument van de eindeloze rangordes

Een eerste aan benchmarking toegekend nadeel luidt: benchmarking zou leiden tot een eindeloze reeks 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 zwaarwegende 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 scoorde goed: het had korte doorlooptijden, weinig gegronde bezwaarschriften en lage kosten per klant. Amsterdam bleek de grote stijger. Het gevolg was dat de toenmalige 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. Zo kreeg elk zijn succesje: de top en de stijger.

Bij deze rangorde bleef het niet. Er ontstond het afgelopen decennium nationaal en internationaal een ware *rangorde-industrie* (Hood, Dixon & Beeston, 2008). Er kwamen lijstjes van bij voorbeeld de beste woongemeente, de beste universiteit of school, het meest doelmatige afvalbeheerder 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? De vraag 'welke universiteit is de beste van Nederland?' kan uitmonden in differentiaties: wie is de beste van de *oude* universiteiten, wie van de *brede* universiteiten, wie van de *kleine* universiteiten, wie van de universiteiten voor *afstandsonderwijs*, wie van de *internationaal gerichte* universiteiten?

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). Ook bij lilliputters is er een de grootste. Wie niet onder universiteiten de beste is, kan nagaan of een van de eigen onderzoeksprogramma's de beste is van de programma's. Wie niet de beste is in één jaar kan bezien of de meting niet meerjarig moet zijn. Wie niet als universiteit eruit springt, kan kijken of er misschien een faculteit is die glanst of een professor (de Nobelprijswinnaar 'in spe') of een voorziening ('onze collegezalen').

Benchmarking kan dus *pervers gedrag* oproepen. Wie niet de beste is in jaar x, kan de prestatiemeting voortzetten door over de jaren u, v, w en x te meten. De '*rangorde-industrie*' heeft meerdere manieren voortgebracht om slechte of matige scores onder het tapijt te vegen. Dergelijk pervers gedrag kan ingefluisterd worden door overambitieuze bestuurders of ontstaan als kerntakendiscussies opkomen of een bezuiniging dreigt. Het nadeel is echter te bedwingen.

2 Het reductie-argument: kwaliteit komt onvoldoende aan bod

We komen bij een tweede vaak aan benchmarking toegekend nadeel. Benchmarking werkt verhullend als de werkelijkheid teruggedbracht wordt tot *het meetbare*, tot het *kwantitatief* meetbare (Van Thiel, 2009). Niet alles is echter te beperken tot het kwantitatieve. Misschien produceert een politie-eenheid wel makkelijk te verzamelen verkeersboetes in plaats van het oplossen van inbraken.

Misschien benoemt een gemeente wel het aantal daklozen door ze te tellen als buurtgemiddelde maar wat zegt dat over de oorzaken.

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* die in het openbaar bestuur een rol kunnen spelen, zoals rechtvaardigheid of mededogen, laten zich bovendien niet of niet gemakkelijk kwantificeren. Denk aan de beoordeling van prestaties van de sociale werkvoorziening, die werk biedt aan personen met een lichamelijke of geestelijke beperking en probeert te komen tot reïntegratie. Een dergelijke organisatie kan niet optimaal concurreren met andere organisaties (Bekkers, 2012: 106).

Het *reductie-argument* – het teveel terug brengen van prestaties tot het meetbare – valt markant te illustreren aan de hand van een internationale benchmarkstudie naar prestaties van de rechterlijke macht, die werd verricht in opdracht van de Raad voor de Rechtspraak (Blank, Van der Ende e.a., 2004). Daarin is de rechterlijke macht van Nederland vergeleken met die in een tiental andere landen. In de studie stond de simpele vraag centraal ‘hoe wij presteren’. Het eerste antwoord op die vraag luidt: niet goed, althans in een periode van vier jaar. Nederland staat per uitgegeven euro in strafrechtzaken op de achtste plaats van elf landen, in civiele zaken op de negende plaats en in bestuursrechtelijke zaken op de elfde plaats. Polen scoort beter.

Is dit terecht? Had de benchmarkstudie beter niet kunnen worden verricht? Een nadere verkenning leert dat het prijs- en salarisoniveau in de analyse een grote rol speelt en veel invloed heeft op de bevindingen. Dan is het logisch dat het ‘duurdere’ Nederland slechter scoort dan het ‘goedkopere’ Polen. Poolse rechters hebben per euro het hoogste aantal afgehandelde zaken op het vlak van strafrecht en burgerlijk recht.

De Nederlandse productiviteit van rechters is echter ook in ander opzicht laag, namelijk als de blik valt op het aantal afgehandelde zaken per medewerker, dus als zowel rechters als het ondersteunend personeel meegeteld wordt. Hoe laat zich dit begrijpen? Daarvoor moet men kijken naar de rechtssystemen.

Het Nederlandse rechtssysteem kent *veel informele filtermechanismen*, meer dan het Poolse. Filteren betekent: zaken waarbij ‘partijen’ er via een onderlinge schikking uitkomen, komen niet meer voor de rechter. Het gevolg is dat de complexe zaken bij de rechter komen. Ontbreken dergelijke filters grotendeels en komen ook eenvoudige zaken bij de rechter dan gaat een bepaald land al snel hoger scoren op een prestatie-indicator. Is wel sprake van filters dan daalt de productiviteit van rechters, wat in het rechtssysteem ook de bedoeling is. Dat moet men weten en daarvoor moet men corrigeren, om goed benchmarkonderzoek te doen. Want informele filtermechanismen zijn niet slecht.

De lage plaats van Nederland in de prestatierangordes van de rechterlijke macht in elf landen kan dan ook in belangrijke mate verklaard worden uit het accent dat in Nederland valt op het informeel afdoen van geschillen. Jos Blank en zijn mede-onderzoekers relativeren de conclusies daarom ook. Zij pleiten voor een benchmarkonderzoek op een lager niveau dan het niveau van de totale rechterlijke macht. Je zou dit *de wisseling van macro naar meso en micro* kunnen noemen. Het is wenselijk om rechters te vergelijken die binnen *eenzelfde institutionele setting* opereren. Anders gesteld, het is onzinnig om landen te vergelijken die een verschillend rechtssysteem kennen want dan ben je appels met peren aan het vergelijken in plaats van appels met appels (zie ook Dijkstra en Toonen, 2004). Maar ook dan moet je bij benchmarking opletten.

Peter de Jong (2004) onderzocht het tijdsbeslag van rechterlijke procedures in eenzelfde bestel, het Nederlandse bestuursrecht, en bekeek in dit verband de rechtbanken van Almelo en Arnhem. De gemiddelde doorlooptijd bedroeg in Almelo 195 dagen en in Arnhem 581 dagen. In Arnhem duren

procedures bijna drie keer zo lang. Is Almelo dan 'best practice'? Je zou denken van wel want het ging de onderzoeker om 'tijdigheid' als benchmarkcriterium. Opnieuw is echter wantrouwen van de cijfers gewenst. De werklust verklaart het verschil. De rechtbank van Arnhem is qua personele omvang duidelijk groter dan die in Almelo, maar het aantal binnenkomende zaken is verhoudingsgewijs nog groter. De rechtbank in Arnhem heeft per medewerker meer te verstouwen maar werkt per euro desondanks meer weg. Zo gezien, is de rechtbank Arnhem 'best practice'. Conclusie: wijs niet zo maar de beste praktijk aan, let bij de vergelijking op de keuze van het prestatiecriterium en achterhaal wat schuil gaat achter verschillen. Wie een degelijke analyse verricht, kan dit nadeel dus aan de kant schoffelen.

3 Het 'outcome'-argument: effecten buiten schot

We komen bij een derde mogelijk nadeel. Worden beleidsprestaties of juist (eind)effecten van overheidsbeleid gemeten? Wie wegversmallingen in wegen aanlegt, zet wel prestaties neer, maar belangrijker is om te bezien of er minder verkeersslachtoffers vallen (effect).

Rond deze vraag stuiten we op een ander nadeel. Bij overheidsbestuur gaat het veelal niet alleen om prestaties (wel outputs genoemd) maar juist om gewenste effecten van beleid (ook outcomes geheten). Benchmarking meet vaak wel prestaties maar juist geen *gewenste effecten* en leidt *niet* tot vergelijking daarvan (Van Dooren, Bouckaert & Halligan, 2010: 176). Een voorbeeld. Het gaat niet om het aantal op straat zichtbare politiemensen (prestatie), 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).

De implicatie van dit nadeel is dus: het is geen inherent nadeel maar een nadeel dat men moet trachten te vermijden.

4 Het normalisatie-argument: uitbannen van verschil

Een aan benchmarkonderzoek toegedicht nadeel is dat het zou leiden tot normalisatie, tot het uitbannen van het bijzondere, tot fletsheid. Immers, het beste uit de sector is het richtpunt. Imitatie dreigt. Is er dan geen sprake van luiheid van denken? Ontbreekt het dan niet aan durf om bijzonder te zijn en zelf na te denken? Deze bezwaren gelden alleen als benchmarkonderzoek op grote schaal zou plaatsvinden en ook zou doorwerken in besluitvorming. We moeten nog zien of dat het geval is.

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 wordt 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: alsof netwerken niet bestaan

We komen bij een vijfde toegekend nadeel. Benchmarkonderzoekers wekken de indruk dat een te benchmarken groep organisaties zelfstandig aan alle knoppen zit om te komen tot invoering van een 'beste praktijk', dat sturing vanuit een cockpit plaatsvindt. Maar die indruk is onterecht. Deze leidt tot overschatting van de sturingsmogelijkheden van één actor. We lichten dit toe.

Benchmarking brengt over het algemeen *niet* netwerkproductiviteit in beeld terwijl gemeenten voor tal van taken, opgaven en middelen afhankelijk zijn van medeoverheden en andere actrices. Denk maar eens aan de bestrijding van droogte in De Peel. Gemeenten moeten samen met enkele ministeries, provincies, waterschappen, grondeigenaren en pachters een gezamenlijke strategie volgen om aan verdroging of teveel vernatting paal en perk te stellen. Gemeenten moeten dus met partners komen tot gemeenschappelijke beeldvorming en hun doelen en middelen verknopen, want geen enkele organisatie beschikt alleen over alle doorzettingsmacht om het gewenste zelfstandig af te dwingen (Cörvers, Glasbergen en Korsten, 2003; Kaashoek e.a., 2010; De Rynck, Voets & Van Dooren, 2010; Van Dooren, Bouckaert & Halligan, 2012; Noordhoek, 2013). Andervoorbeeld. Denk ook maar aan gemeenten die woningbouw plannen, en corporaties en projectontwikkelaars die bouwen. Ook deze organisaties hebben elkaar nodig.

Nog een stap verder, organisaties kunnen onderdeel zijn van een keten. Prestaties van de ene organisatie kunnen dan afhankelijk zijn van een andere organisatie. Dat kan een ongewenst effect hebben, namelijk dat ketenpartners zich indekken tegen elkaars falen (Van Thiel, 2009: 23).

Deze ontkenning van 'governance' is als een nadeel van benchmarking te beschouwen. Er moet in geval van aanwezigheid van versnipperde verantwoordelijkheden en middelen nagedacht worden over hoe dit nadeel te ondervangen.

6 Het argument van schijnsucces: te makkelijk scoren

Outputsturing, die benchmarking als hulpmiddel gebruikt, heeft als potentieel nadeel dat deze vernieuwing in de praktijk kan belemmeren en kan bijdragen aan verstarring (Commissie-Bovens, 2006: 28; Leeuw, 2008). Dit nadeel van verstarring kan naar voren komen doordat organisaties doelen en prestaties gaan selecteren die gemakkelijk (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 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 gemeentex het percentage opgeloste inbraken iets is gestegen maar het veiligheidsgevoel juist is gedaald (mede omdat bepaalde inbraken niet eens meer gemeld worden).

7 Het bureaucratie-argument: benchmarking vergt administratieve lasten

Benchmarking is niet gratis. Wil je een goede benchmark laten verrichten dan kan het niet 'op een koopje'. Deelnemen aan een benchmark impliceert het opstellen van een offerte en de selectie van een onderzoeksteam, bezinning op de opdracht, de selectie van nuttige indicatoren, gegevensverzameling, analyse, rapportage, het aanwijzen van de beste praktijken. 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; Tonkens, 2004; Frey & Osterloh, 2009; Noordegraaf en Sterrenburg, 2009; Bekkers, 2012: 106).

Tonkens (2004: 51) spreekt van een *bureaucratie-explosie*. Dat wordt ook gezien als een nadeel van benchmarking. 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). Daar staat tegenover: wie een lerende organisatie wil zijn, moet er wat voor over hebben.

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. Tot de verbeterruimte helemaal gevuld is. Overdreven? Misschien dat dit argument geldt bij de productie van raceauto's voor Hamilton, Verstappen en andere coureurs die deelnemen aan F1-autoraces maar niet voor alle taakgebieden van gemeenten of waterschappen.

Hier is ook 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 reserves over juist dat lerend effect. Immers, zo zegt de commissie Toekomst Lokaal Bestuur: benchmarks werken nivellerend en drijven (selectieve) ambitie uit (Commissie-Bovens, 2006: 28). Dat duidt niet direct op 'double loop learning', zeg maar fundamenteel leren.

Conclusie: dat benchmarking tot permanente verandering zou aanzetten, is incidenteel niet uitgesloten maar in de breedte gezien lijkt deze veronderstelling niet juist.

9 Het schuttersputje-argument

Als benchmarking op grote schaal plaatsvindt, werkt dat verkoking en fragmentatie in de hand. Aldus een toegekend nadeel. De redenering is als volgt.

Als sterk op benchmarking geleund wordt, zullen vooral die prestatiegebieden in beeld komen die overzichtelijk zijn, gerelateerd zijn aan routinematig uit te voeren taakgebieden en massaproductie kennen. Vooral wat te tellen is, komt in de kijker. 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 een minder hechte verankering bij 'eigenaren'. Het gevolg is dat organisaties minder bereid zouden zijn om zich in te spannen voor benchmarking op dat terrein. Benchmarking kan bevorderen dat iedereen zich terugtrekt in het eigen 'schuttersputje', meent de Rotterdamse hoogleraar Victor Bekkers (2012: 106).

10 Benadeling van professionaliteit

Er is nog een andere kritiek op benchmarking. Die kritiek houdt in dat prestatie meting 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 prestatie meting plaats, juist om de prestaties 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 vakmatige kwaliteit 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 muskusrattenpaar een nest jongen heeft om daarna pas te gaan vangen. Tel uit je winst? Hij benadrukt een en andermaal dat prestatie meting 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. Voor wie handig maar pervers wil opereren, gelden de volgende tips.

- 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 gelijkblijvende 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.

Benchmarking heeft door deze perverse effecten in feite *een morele schaduwzijde*.

4.3. Enkele conclusies

Dit hoofdstuk levert een boeiende oogst op. Benchmarking heeft de volgende toegekende of geproclameerde voordelen:

- a) het is een opstap voor nadere analyse van de positie in een rangorde (waarom staan we zo laag?) en voor verbetering;
- b) benchmarking kan bovendien vroege waarschuwingen ('early warning') opleveren over wat in een organisatie verbetering behoeft en welke kans zich in de vorm van een goede praktijk aandient; benchmarking maakt duidelijk of er verbeterruimte zit op een bepaald (beleids)terrein of taakgebied;
- c) benchmarking brengt prestaties samenhangend in beeld;
- d) benchmarking leidt in potentie tot opschoning in organisaties van zwakheden;
- e) benchmarking geeft zicht op wat niet en wat wel bereikt is en is daarmee in potentie een middel om personeel te motiveren;
- f) benchmarking brengt een organisatie in een sfeer van verbeteren; daarmee ophouden is eigenlijk niet goed mogelijk;
- g) het zelfreinigingsargument: de 'code van goed openbaar bestuur', vastgesteld door minister Ter Horst en staatssecretaris Bijleveld (kabinet-Balkenende IV), verwacht van elke organisatie dat die evalueert en lerend vermogen aan de dag legt.

Als het verantwoord gebeurt, lijkt er niets tegen benchmarken in sectoren en bij producten, diensten en processen die zich kwantitatief laten vangen.



Benchmarking geeft grote of kleine verschillen te zien?

Tot de geproclameerde nadelen van benchmarking behoren de volgende:

- a) het argument van de eindeloze rangordes: een gebenchmarkte organisatie is altijd wel ergens goed in; je kunt benchmarken tot die kwaliteit gevonden is, maar dat ontkent de waarde op andere terreinen;
- b) het reductie-argument: de werkelijkheid laat zich niet goed alleen in meetbare kwantiteiten vangen;

- c) het outcome-argument: prestaties (outputs) vergelijken en rangordenen miskent dat het meestal om de effecten (outcomes) gaat;
- d) het normalisatie-argument: alles wat gemeten is, gaat na benchmarking en benchlearning op elkaar lijken; variëteit verdwijnt vanwege het overnemen van goede praktijken;
- e) het monocentrisme-argument: benchmarks stellen een organisatie centraal alsof die altijd over alle taken, bevoegdheden en middelen beschikt terwijl in werkelijkheid een gemeente afhankelijk is van andere organisaties en dus het netwerk goed moet functioneren, niet alleen de organisatie (maar een rangorde van netwerken komen we nergens tegen);
- f) het argument van schijnsucces: benchmarking kan echte vernieuwing belemmeren omdat benchmarking van nature werkt in de richting van verbetering van het bestaande;
- g) het bureaucratie-argument: benchmarking vergt inzet en is niet gratis;
- h) het argument van het bezwaar van permanent veranderen: wie blijft benchmarken volstaat niet meer met 'behoud van het goede', maar wil steeds weer naar het betere;
- i) het schuttersputje-argument: benchmarking neigt ertoe zich te richten op producten, diensten en processen van afzonderlijke afdelingen (verkokerde organisaties) en blijft weg bij beleid dat dwars door organisaties loopt (door Belgen transversaal beleid genoemd);
- j) het argument van benadeling van professionaliteit: wie vertrouwen heeft in de professionaliteit van medewerkers hoeft niet steeds te benchmarken; de professionals zoeken als vanzelf naar verbetering door vergelijking met referentiekaders.

Bij deze overzichten van veronderstelde voor- en nadelen zijn enkele kanttekeningen te maken.

Ten eerste, benchmarking heeft op het eerste gezicht meer kwaliteit naarmate de onderzoekers erin slagen om de voordelen te benutten en de nadelen te vermijden. Aan het voorkomen van nadelen valt te werken.

Bepaalde nadelen, zoals het reductie-argument, zijn echter niet te ondervangen. Het is daarom verstandig om ook de pretentie niet te hebben om benchmarking breder in te zetten dan het instrument aan kan. Om het metaforisch te zeggen: een centimeter – een meetlint van anderhalve meter voor het meten van kleine afstanden - is geen geschikt middel om te gebruiken voor het opzetten van een afstandstabel tussen steden in Europa. Probeer het dan ook niet!

Ten tweede, bepaalde nadelen zijn zo evident dat ze direct de kwaliteit van benchmarking begrenzen. Benchmarking suggereert steeds dat er slechts sturing vanuit één cockpit mogelijk is. Dit monocentrisme-argument impliceert een sterke nadruk op de doorzettingsmacht van (slechts) één overheidsorganisatie. In werkelijkheid kan een gemeentebestuur echter maar weinig alleen voor elkaar krijgen. Een gemeente heeft op tal van terreinen juist te maken met afhankelijkheden in netwerken. Bij veel moet een gemeentebestuur de doelstellingen en middelen verknopen met andere 'partijen' waarvan men afhankelijk is omdat deze organisaties ook over taken, bevoegdheden en middelen beschikken. In werkelijkheid is een organisatie als een gemeente dus afhankelijk van tal van andere actoren, zoals bij woningbouw en zorg voor ouderen. Immers, de gemeente plant maar het zijn de corporaties en projectontwikkelaars die bouwen. Benchmarking houdt meestal geen rekening met netwerken.

Ten derde, benchmarking is ook als een drukmiddel te gebruiken. Wie lid is van een raad van toezicht kan aan een positiebepaling van een onderwijsorganisatie zien of er in vergelijking met andere organisaties in de eigen organisatie nog verbeterruimte moet zitten ('we zitten in de middenmoot, dus ...'). Dit argument geldt ook naar wethouders, leden van gemeenteraden en andere bestuurders. Zo gezien, is het een hulpmiddel dat bijdraagt aan 'checks and balances'.

5 Ervaringen met verschillende benchmarkingtrajecten: verdiepende impressies

5.1 Zijn academische bezwaren ook praktische bezwaren?

Treden voor- en nadelen in concrete gevallen zichtbaar op? Is bij voorbeeld het beeld van benchmarkonderzoek dat vertrouwen in ambtenaren ondergraaft en pervers gedrag uitlokt concreet voorstelbaar? En, is dit bezwaar 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 uit verschillende sectoren:

- uit de wereld van de geestelijke gezondheidszorg, de ggz (het regionaal instituut voor geestelijke gezondheidszorg, de Riagg),
- uit de wereld van het middelbaar beroepsonderwijs (mbo: ROC's),
- uit de universitaire wereld,
- uit de wereld van gemeenten, en
- uit de overige gebieden in 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 zijn min of meer integraal. 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).

Een aantal in de literatuur *geproclameerde nadelen* van benchmarking blijkt in de praktijk *klein of afwezig*. Onder meer zal blijken dat het genoemde bureaucratieargument en het schuttersputje-argument *niet* zijn waargenomen. Ze worden in de praktijk *niet* gevonden of benoemd.

- Dat benchmarking de administratieve lasten verhoogt (het bureaucratieargument), is iets dat in de praktijk reuze meevalt.
- Dat benchmarking leidt tot een schuttersputjeseffect (= benchmarking neigt ertoe zich te richten op producten, diensten en processen van afzonderlijke afdelingen en blijft weg bij beleid dat dwars door organisaties loopt) blijkt niet.
- Dat benchmarking variëteit vermindert, is zwaar overtrokken want dat kan alleen als benchmarkrapporten ook op grote schaal door professionals benut worden en leiden tot 'leren'. Een enorme, expliciete benutting is *niet* waargenomen in de periode 2000-2017.
- Dat benchmarkonderzoek tot pervers gedrag leidt, kan voor komen 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 & Halligan, 2010) blijkbaar zwaar overschat. Zo belangrijk zijn benchmarks ook weer niet. Ze zorgen doorgaans niet voor opschudding.

Deze argumenten zijn blijkbaar vooral *bedachte* argumenten die in de literatuur te vinden zijn. Dat de geproclameerde nadelen van geen of beperkte betekenis zijn, moet niet gesteld worden. Hier blijkt dat ze in de *hier onderzochte* gebieden niet zo opspelen.

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.

5.2 Benchmarking bij centra voor geestelijke gezondheidszorg

We komen bij het eerste voorbeeld. 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 ambities 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; integraal want de benchmark ging in op het algehele functioneren van de Riagg's; alle taakvelden, zoals hulpverlening aan ouderen in geestelijke nood, kwamen in de benchmarks 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 Riagg-beleid was geenszins sprake. De bespreking van het rapport werd in de vergadering van de raad van toezicht 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. Een benchmark is doorgaans niet voldoende om bestaande werkwijzen in een organisatie volledig ter discussie te stellen, laat staan op zijn kop te zetten. Het hele benchmarkdenken is gericht op het zoeken naar *verbeterruimte* in de *bestaande* werkwijzen tenzij de 'best practice' radicaal verschilt van de bestaande aanpak.

5.3 Benchmarking bij regionale opleidingscentra

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 van ons, dat gepaard ging met het in totaal bijwonen van grofweg 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 van het eigen ROC gaf doorgaans uitleg van wat er stond en hoe 'we' de uitkomst moesten begrijpen. Ook hier was weer nauwelijks sprake van tekst, van een verdiepende analyse en rapportage. Van een aanduiding van 'best practices' was bij onderzoekers 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 was er dus geen excellente benchmarkkwaliteit, maar toch was er tevredenheid alom. Alle leden van de raad van toezicht en het bestuur hadden altijd al met zoveel onderzoek van doen gehad dat ze niet verrast waren door de ene benchmark.

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 ten aanzien van een taakgebied als gevolg van die benchmarking was bij onze 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 begrijpelijk, relevant, actueel en interpreteerbaar zijn. Maar is dan ook sprake van een vertaalslag naar handelen? Weten bestuurders 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 het

betreffende regionaal opleidingscentrum (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 van scores is geen sprake. De bestuurder voegt toe:

'We weten natuurlijk wel waar de andere ROC's ongeveer staan. Op basis daarvan vraag je collega-bestuurders 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'.

5.4 Vergelijking van universitaire opleidingen

Een derde ervaring met benchmarking betreft de visitatie van universitair onderwijs in de vorige eeuw (Korsten, 2004b). Een van de eerste visitaties van bestuurskunde-opleidingen was zowel kwalitatief als kwantitatief gericht, een afwijking van het hoofdbeeld bij benchmarking dat wijst op een bij uitstek kwantitatieve gerichtheid. 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 daarvan, 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 in universitair beleid van de onderscheiden instellingen. Wie gaat immers navlooien 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 gemakkelijk 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 (bijna) iedereen een prijs. Benchmarking kan zo leiden tot een collectief 'hiep hiep hoera'-effect (Abma en Korsten, 2009; Dixon e.a., 2013).

5.5 Benchmarking van gemeentelijke organisaties en hun besturen

Een vierde ervaring met benchmarking betreft de analyse van wat eerst 'De Staat van de Gemeente' (Aardema en Korsten, 2005; Abma en Korsten, 2009) werd genoemd en naderhand [waarstaatjegemeente.nl](http://www.waarstaatjegemeente.nl) (Abma, 2014). 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 gemeentebesturen 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. Daarop werd een werkgroep opgetuigd om tot een instrument te komen. Deze stap leidde tot een instrument waarbij 'harde' gegevens vanuit het CBS en andere instanties werden gecombineerd met 'zachte' in de vorm van oordelen van burgers in verschillende rollen over hun gemeentebestuur. Die oordelen werden via tevredenheidsonderzoeken 'opgehaald' bij de inwoners. Al deze cijfers werden verdeeld over zes rollen die een burger aan kan nemen ten opzichte van zijn bestuur: als kiezer, als klant, als wijkbewoner, als belastingbetaler, als onderdaan en als partner. Een gemeente kon op basis van in totaal ongeveer zestig indicatoren, verdeeld over deze zes rollen, beoordeeld worden.

Na enkele pilots is dit instrument bij veel gemeenten toegepast. Abma en Korsten hebben in *'Gemeenten in rapportcijfers'* hiervan verslag gedaan. Waarstaatjegemeente werd in 2014 geëvalueerd (Abma, 2014). Het bleek dat gemeentelijke organisaties en hun besturen over het algemeen op de rollen van 'kiezer' en 'partner' het slechtste functioneerden. Dat impliceert dat burgers relatief kritisch zijn op het college van B&W en de gemeenteraad. Het vertrouwen daarin is in veel gemeenten, althans volgens 'De staat', gematigd negatief en in andere net voldoende. Ook de wijze waarop burgers betrokken werden bij de totstandkoming en uitvoering van beleid kan niet op veel waardering rekenen: burgers vinden dat ze te weinig gehoord worden.

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 één 'eiland'. De rijksoverheid spreekt dus niet altijd met één mond. Een effect dat de benchmarks *'De Staat van de Gemeente'* casu quo waarstaatjegemeente tot de politiek-bestuurlijke dood van tal van andere 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 voor de VGS uit het vuur.

Het instrument 'De Staat' heeft allerlei voordelen gehad. Opvallend is dat veel gemeentebesturen zich geen raad wisten met het gebrek aan vertrouwen dat nogal wat kiezers hadden in het college en de gemeenteraad.

In hoofdstuk 10 gaan we nader in op deze beide instrumenten.

5.6 Een benchmarksysteem 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 (van de Rob) was op grond daarvan positief over benchmarking maar dan wel in een kleine setting, dus een vergelijking van bij voorbeeld vijftien 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'?

Deze bevindingen waren een teleurstelling voor die medewerkers van het ministerie van Binnenlandse Zaken en Koninkrijksrelaties die dachten het 'grote' instrument in handen te hebben dat antwoorden op heel veel vragen kon geven. Maar in feite was de Raad voor het openbaar bestuur positief over benchmarking. Het Rob-advies had alleen de benchmarkboom goed gesnoeid, dus ontdaan van dode takken, dus van ambities die toch niet tot bloei konden komen.

5.7 Enkele conclusies

Benchmarking, waarvan de deelname vrijwillig is, is feitelijk op betrekkelijk grote schaal toegepast. 'De Staat van de Gemeente' is hier de treffende illustratie van.

De deelname aan benchmarking wordt niet altijd stevig overwogen. Soms is het nieuwsgierigheid die aanleiding is voor deelname of zijn het de lage deelnamekosten. Soms heeft een opdracht voor het starten van een benchmark, zoals bij de regionale instelling voor geestelijke gezondheidszorg (Riagg), iets van een verkenning, onder het motto '*laten we eens meedoen; deelname kan geen kwaad*'. Of heeft een opdracht iets van een rituele dans: '*we doen al jaren mee met zo'n benchmark*'. Men wil geen spelbreker zijn. Het visiteren van universitaire opleidingen is daarentegen stevig aangezet. De deelname weigeren is uitgesloten als gevolg van afspraken tussen instellingen, de minister en de onderwijsinspectie. Visiteren is hier opgenomen in een stelsel met afspraken.

De rapportages over 'De Staat van de Gemeente' zijn op het eerste gezicht positief te beoordelen. De benchmark is ontworpen, uitgetest en levert een begrijpelijke rapportage op. Uit andere analyses blijkt dat benchmarking niet steeds voor niet-insiders (goed) begrijpelijke rapportages biedt, niet steeds gericht is op een verklaring van verschillen tussen prestatiescores 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).

De lezer moet beseffen dat het doel achter benchmarking, namelijk verbeteren, bestuurlijk en ambtelijk niet altijd makkelijk te realiseren is, ook al is er *verbeterruimte* in de aanpak van een taak of het verlenen van een dienst of product aanwijsbaar. Het benutten van benchmarkrapporten vergt energie en de wil om te reflecteren en veranderen, wat een doorbreking van de status quo vergt, hoe klein of groot ook. Daar tegenover staat het verlangen naar 'behoud van het goede'.

Dat benchmarking tot verbeteringen leidt, is dus niet vanzelfsprekend. Een mythe, 'verbeteren door benchmarking lukt altijd', lijkt zo ondergraven te worden. Verderop komen we hierop terug en zullen we dit onderwerp uitdiepen.

Benchmarking heeft zeker voordelen. Wie geen voordelen ziet, start ook geen benchmarking. Een aantal voordelen is hier genoemd.

Zijn we de in de literatuur ook in het vorige hoofdstuk genoemde bezwaren tegen benchmarking tegengekomen?

- Het *normalisatie-argument* impliceerde dat organisaties door vergelijking en het overnemen van goede praktijken meer gelijk worden. Dit bezwaar tegen benchmarking van het verdwijnen van variëteit is *niet* in de genoemde praktijken (gemeenten, ROC's, Riagg's, universiteiten) vastgesteld. Daarvoor blijken de producten en processen in organisaties toch teveel contextgebonden of uniek.
- Een praktijk dat organisaties allemaal op de *beste* organisatie willen gaan lijken, dus een mechanische kijk op benchmarking, is evenmin waargenomen. Het beeld van wie de beste is of wat het beste is, is zeker niet altijd expliciet in een rapport opgenomen of is onscherp en teveel gehuld in 'mist'. Blijkbaar wijkt benchmarking in de publieke sector vanwege de aard van gemeten en vergeleken producten, processen en contexten toch teveel af van private organisaties, zoals bedrijven.
- Dat een benchmarkonderzoek een *bestuurlijke ambitieremmer* zou zijn, omdat de verbeterpotentie slechts gericht zou zijn op bestaande producten, processen of diensten en niet op radicale alternatieven, is echter ook niet gebleken.
- Dat benchmarking tot wel heel veel veranderingen in deelnemende organisaties zouden leiden, is in de praktijk van de benchmarking, zoals die in dit hoofdstuk aan de orde was, niet waargenomen. Voor zover er veranderingen optreden, gaat het om veranderingen in werkwijzen waarvan de hoofdstructuur blijft bestaan. Veranderingen zijn aanpassingen, *optimalisaties*. Over het algemeen is er dus géén sprake van 'double loop learning' (term in het werk van Argyris) als gevolg van benchmarking bij het Riagg of ROC's. Geen radicale vernieuwing dus. Zelden wordt het roer in deze sectoren als gevolg van benchmarking volledig omgegooid. Logisch eigenlijk, want benchmarkrapporten bevatten geen aanbevelingen hiervoor.
- Het *reductie-argument* houdt in dat de werkelijkheid in benchmarks verengd wordt tot meetbare kwantiteiten en de kwaliteiten van producten, diensten of processen niet in de volle werkelijkheid blijken. Bij benchmarking is die reductie ontegenzeggelijk al gauw aan de orde: waar hitlijsten en rangordes bestaan, staat het getal voorop. Wie zicht wil krijgen op het verhaal achter de getallen, op de kwaliteit van de personeelsformatie of van opleidingen van een regionaal opleidingscentrum of op de kwaliteit van een instelling voor geestelijke gezondheidszorg, moet dus aanvullende analyses plegen. Dat blijkt markant uit de eerste golf visitierapporten van universitaire opleidingen. De onafhankelijke visitatiecommissies hadden en hebben 'site visits' nodig om zich te vergewissen van bepaalde kwalitatieve aspecten van opleidingen, zoals van de sturing door opleidingsdekanen, de inhoud van cursussen, de kwaliteit van afstudeerscripties, de opstelling van docenten, de satisfactie van studenten.
- Van het optreden van *meer bureaucratie* als gevolg van benchmarking was in de onderzochte sectoren niet of nauwelijks sprake. De toename van administratieve lasten was bij ROC's en Riagg's heel beperkt. Dit punt wordt ook niet door betrokkenen bij benchmarking in de genoemde sectoren als kritiek geventileerd. Dat benchmarking de *verkokering* en *fragmentatie* binnen organisaties zou versterken, bleek hier niet.

Tal van nadelen die aan benchmarkonderzoek worden toegedicht, blijken dus in de benoemde sectoren onvindbaar of zwaar overtrokken, mede doordat benchmarking niet steeds is (geworden) wat we verwachtten. Deze stellingname heeft echter een beperking. Het aantal gevallen waarop deze stellingname is gebaseerd is nog te beperkt om te kunnen generaliseren. We zullen dus verder moeten gaan met onderzoeken. Op naar een volgend hoofdstuk.

Er blijken nog andere hoofdconclusies te trekken. Benchmarking heeft voordelen. Zoals dat naar voren komt hoe de organisatie zich verhoudt tot andere organisaties. Dat benchmarking zou werken als *'early warning'* voor wat later manifest aan de orde is in een sector of branche, als te waardenen iets of signaal voor een verkeerde weg, is niet gebleken.

Het confronteren van alle geproclameerde voor- en nadelen van benchmarking met de werkelijkheid van enkele sectoren waarin benchmarking is toegepast, laat zien dat de nadelen over het algemeen overtrokken zijn. Dat benchmarking het nadeel heeft dat het vooral op het kwantitatieve is gericht, is echter evident.

We moeten dus aan de hand van onze rapportages over enkele sectoren (Riagg, ROC, gemeenten) ook vaststellen dat de kwaliteit van benchmarking enigszins *overschat* wordt. De analyse van scoreverschillen in prestaties en het formuleren van navolgbare *'best practices'* is verre van vanzelfsprekend. Blijkbaar komt het bij een bestuur na afloop van een benchmarkonderzoek niet aanwaaien wat het aan consequenties kan of moet verbinden aan de uitkomsten.

Benchmarkrapporten dwingen in feite na verschijnen tot een aansluitend bezinningstraject in een deelnemende organisatie en die bezinning treedt niet altijd op of is sober. Zo gezien, valt er meer van benchmarking te maken en valt er meer uit benchmarking te halen.

6 Veronderstellingen achter benchmarking nader bezien

Laten we op zoek naar het echte verhaal over benchmarking nog eens verder graven. Achter benchmarking gaan uiteraard veronderstellingen schuil. We noemen er verderop een aantal. Helder (2001: 14) merkt naar aanleiding van ervaringen met benchmarking in de gemeenten Barneveld, Heusden, Noordoostpolder en Soest het volgende op:

‘Een benchmark in zijn pure vorm is vaak moeilijk te realiseren.’

Als dat zo is, en dat sluit wel aan op de bevindingen uit het vorige hoofdstukken, dan zullen bepaalde veronderstellingen die nog *wel* opgaan voor de private sector misschien *niet* opgaan voor de publieke sector. We benoemen dus niet alleen de assumpties maar zullen er ook kritische kanttekeningen bij maken om zo een realistisch beeld te krijgen over wat van benchmarking bij overheden te verwachten is en wat niet.

6.1 Veronderstellingen op een rij

1 Appels met appels vergelijken

Bij benchmarking worden organisaties, producten, processen of diensten vergeleken. De vraag is of het altijd lukt om vergelijkbare organisaties of eenheden te vergelijken? Over het algemeen lukt dat zeker omdat elke organisatie die wordt opgeroepen tot deelname aan een benchmark de impliciete of expliciete taak heeft zelf te bepalen of vergelijking met andere organisaties of eenheden realistisch is. Maar makkelijk is dat niet altijd. Neem vliegtuigmaatschappijen. Functioneert de ene maatschappij beter dan de andere? Ga er maar aanstaan om dan te vergelijken.

Het kan zin hebben de prijsstelling en doelmatigheid van enkele vliegtuigmaatschappijen, bijvoorbeeld een staatsmaatschappij en een private maatschappij, te onderzoeken maar dan moeten ze bij voorkeur over dezelfde vloot aan vliegtuigen beschikken en op vergelijkbare momenten naar zelfde bestemmingen vliegen. Een veronderstelling is dan dat het mogelijk is om organisaties op te sporen waartussen een vergelijking echt zinnig is. Bij benchmarking moet een onderzoeker immers ‘appels met (zelfde) appels vergelijken’ en niet appels met peren. 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 (voormalige) kleine en veelkernige plattelandsgemeenten Littenseradiel (29 kernen) of Súdwest-Fryslân (in 2017 met 69 kernen). Afwijkende factoren zoals gemeentelijke bestuurlijke ambities en speerpunten, omgevingsfactoren zoals de fysieke structuur (bodemkwaliteit, waterwegen) en het profiel van de gemeente (hoofdstad, winkelstad op hoog niveau of plattelandsgemeente?) spelen een rol bij een vergelijking. Ook de intensiteit van het verkeer, het type verkeer en de bodemstructuur verschillen.

Vliegtuigmaatschappijen met eenzelfde vloot die op zelfde dagen op zelfde bestemmingen vliegen, zijn heel moeilijk te vinden. Er wordt ten behoeve van benchmarking natuurlijk wel geprobeerd om echt vergelijkbare organisaties op te sporen, maar over het algemeen gebeurt dat toch op *indirecte* criteria, zoals een vergelijkbare gemeentegrootte qua aantallen inwoners. Soms zoeken vergelijkbare gemeenten elkaar ook op, zoals drugsgelateerde grensgemeenten als een vergelijking van dergelijke gemeenten gewenst is. Maar duidelijk zal zijn dat de benchmarkonderzoeker een handicap heeft als hij toch bij nader inzien niet erg vergelijkbare producten, diensten of processen aan het onderzoeken is.

Maar zelfs al is het bijna onmogelijk om een vergelijkbare organisatie op te sporen, dan moet er toch worden gebenchmarkt. De opdrachtgever wil dat. Zo lezen we in het verslag van de benchmark naar de formatie van de gemeente Súdwest-Fryslân de volgende constatering: ‘Door de bijzondere omvang van de gemeente Súdwest-Fryslân in combinatie met een exceptioneel groot aantal kernen, zijn op meerdere aspecten direct vergelijkbare gemeenten niet voorhanden. Mits wij in de interpretatie van de gegevens rekening houden met de verschillen zijn echter wel zinvolle vergelijkingen te maken.’ Daar kwam verder bij dat gemeenten die in beeld kwamen als partner ook

bereid moesten zijn om als vergelijkingspartner te fungeren. De vrijwilligheid in deelname kan het aantal potentiële partners behoorlijk reduceren.

Verder is het van belang dat de vergelijking bij voorkeur betrekking heeft op 'outcomes', dus op *gewenste en feitelijke effecten* en niet slechts op prestaties (Van Dooren, Bouckaert & Halligan, 2010: 176). Het gaat dus, om bij de metafoor van appels te blijven, om export en bijdragen aan welzijn, gezonde voeding en andere denkbare effecten van consumenten. Mocht het niet lukken om gewenste effecten in een vergelijkend onderzoek te onderzoeken dan moet een onderzoeker wel terugvallen op andere indicatoren, bij voorbeeld over prestaties. Denk dan aan de kwaliteit van appels (wel of niet beschadigd) en de smaak.

Uit de voorbeelden die in eerdere hoofdstukken van benchmarks zijn gegeven, blijkt overigens dat het lukt om vergelijkbare organisaties te vinden. Om een voorbeeld te geven, de vergelijking van de vijf Riagg 's was geen enkel probleem; vooraf niet en achteraf niet. De vergelijking bij 'De Staat van de Gemeente' was ook geen punt van twijfel. Deze eerste veronderstelling is bij nader inzien over het algemeen dus niet problematisch.

2 De opdrachtgeversrol

De tweede veronderstelling achter benchmarking betreft de keuze van partners in de vergelijking. In de benchmarkliteratuur (o.a. Camp, 1992; Keeley, 1996; Commissie-Van Rijn, 2007) worden processtappen beschreven. Die gaan uit van het gegeven dat de organisatie zelf kiest voor benchmarking, daarvan de omvang bepaalt, zelf zijn 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 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. De (eerder genoemde) *productgerichte* benadering domineert boven een meer *procesgerichte* aanpak van benchmarking. 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 gemeenten hebben dan op de inhoud van de benchmark geen invloed. Deze staan dus vast. Het Kwaliteitsinstituut Nederlandse Gemeenten (KING) heeft in het verlengde daarvan ook benchmarks gecertificeerd.

De rol van opdrachtgevers moet groter zijn, meent de commissie-Van Rijn, die actief was op het terrein van benchmarking van corporaties.

Bij meer diepgaande onderzoeken als bestuurskrachtmetingen is de opdrachtgevende rol van de gemeentebesturen groter dan bij benchmarks (Korsten, 2004b; Schillemans, 2009; Abma, 2012). Zij stellen ten behoeve van bestuurskrachtmetingen bestuurlijke opgavenprofielen vast, ze zijn betrokken bij de normenkaders, enzovoorts. Dit bevordert de bestuurlijke en politieke betrokkenheid. Iets wat juist een gemis kan zijn of is bij benchmarking, althans in een sterk aanbodgerichte benadering. Dat kan wellicht voor een deel verklaren waarom benchmarkresultaten sommige dagelijks besturen nog wel bereiken maar heel vaak de gemeenteraden niet (verg. ook Schutgens, 2014).

3 Commensuratie

We komen bij een derde veronderstelling. Die gaat over commensuratie. Wat wordt daarmee bedoeld?

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

onderzoekers, beleidsvoerders of journalisten over een dergelijke omvattende vergelijking te communiceren? Benchmarking heeft het voordeel dat in een vergelijkende analyse eenvoudige, compacte en gestandaardiseerde indicatoren die bij voorkeur weinig interpretatie behoeven, gebruikt worden. Ook in geval van pensioenstelsels.

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

Maar met het gebruik van indicatoren worden kwaliteiten gereduceerd tot kwantiteiten, tot 'behepbare' 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 wisselingen 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 de kostenstructuur bij politiekorpsen zien dat de paradox bestaat. Ze melden hier en daar dat een korps in een bepaald opzicht hoog scoort om een jaar later in vergelijking met organisaties op hetzelfde vlak betrekkelijk laag te scoren of omgekeerd. Het waarom kunnen ze niet steeds verklaren. Het gevolg van deze grilligheid kan zijn een potentiële benchmarkgebruiker de resultaten van een benchmark links laat liggen.

In werkelijkheid is commensuratie inderdaad een kritisch punt. Het reduceren van kwaliteiten tot kwantiteiten voert ons naar de kern van benchmarking omdat het daar altijd gaat om 'meten' en om 'kwantificeerbaarheid'. We stoten daarbij op de grens van benchmarking want benchmarking zou door de reductie van kwaliteit tot kwantiteit aan relevantie kunnen inboeten. Bij de benchmark 'De Staat van de Gemeente' speelt dit kritisch punt evenwel niet erg. Gemeentebambtenaren hebben in de oorspronkelijke opzet zelf gezocht naar zestig indicatoren ten aanzien van zes burgerrollen in de verhouding tot gemeenten.

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, zoals Duitsers en Vlamingen het zeggen.

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 ten aanzien 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 regionale organisaties in de geestelijke gezondheidszorg (Riagg's) of ziekenhuizen kan 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 gniffelen bij het vergelijken van de productie waarbij aantallen behandelingen van cliënten of patiënten telden. 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.

In het algemeen blijkt prestatiemeting beter mogelijk als het object van benchmarking minder complex is en er minder van politieke sturing sprake is. De toepasbaarheid van prestatiemeting verschilt daarom binnen de publieke sector (Borst, Lako en De Vries, 2012: 17-18). Dienstverlening laat zich makkelijker in benchmarking vatten dan een beleid gericht op complexe (ongestructureerde) problemen, zoals energietransitie.

5 Vergelijking van cijfers in meerjarenperspectief

Een volgende 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 in de praktijk lang niet altijd voldaan (Blank, Felsö en De Groot, 2011: 9). Soms komt evenwel herhaalde meting voor, zoals in geval van toepassing van 'de Staat van de Gemeente'.

6 Analyse van prestatieverschillen nodig

Waarom wordt Barcelona nationaal kampioen in de Spaanse profvoetbalcompetitie, Real Madrid tweede en wordt Gijon twaalfde? Als Gijon een volgend seizoen hoger wil eindigen, zal toch bekend moeten zijn wat er binnen de club moet gebeuren. Een groter budget zodat betere spelers aangetrokken kunnen worden en de goede spelers ook betere contracten krijgen? Een trainer die past bij de versterkte ambitie? Een stadion dat geschikt wordt voor het spelen in Europese competities voor het geval de club bij de eerste vier in de competitie gaat eindigen? Een alliantie met een fameuze buitenlandse club als Arsenal zodat deze Engelse club er tijdelijk spelers kan stallen die in het eigen elftal niet aan spelen toekomen?

Een zesde veronderstelling slaat op de verklaring voor verschillen op de ranglijst van gemeten prestaties tussen organisaties, bij voorbeeld gemeenten (Blank, Felsö en De Groot, 2011). Waarom scoort gemeente x hoger dan gemeente o tot en met z? Die analyse is nodig om te begrijpen waarom een gemeente in een rangorde aan de top staat en een andere onderaan bungelt en vervolgens om te achterhalen aan welke knoppen een overheidsbestuur moet proberen te draaien om tot verbeteringen te komen. Een benchmarkstudie naar gemeentelijk afvalbeheer toont dit nog eens aan. Doelmatigheidsverschillen tussen gemeenten op het vlak van afvalbeheer bleken veel meer afhankelijk te zijn van de stedelijkheidsgraad van gemeenten dan van de organisatie van de afvalophaal en de verwerking.

Dit is een belangrijke veronderstelling. Verderop zal blijken dat deze veronderstelling lang niet altijd uitkomt of gerealiseerd wordt. Het begrijpelijk maken van verschillen blijft dan achterwege. Blijkbaar kost het al genoeg moeite om te komen tot een vergelijking en rangorde.

7 Goede praktijken zijn transporteerbaar naar andere organisaties

Wat zijn 'best practices'? Om weer bij het profvoetbal te blijven. De beste praktijken kunnen misschien allemaal bij de kampioen gevonden worden. Dan beschikt de bestuurlijke leiding van de kampioen over het meeste budget, de meeste competenties en stabiele verhoudingen in de organisatie en het management van de club, een goede selectie met de beste spelers, het uitblijven van langdurige blessures bij spelers, een goede begeleidende staf, de beste trainer, een goede doorstroming van jonge talenten naar het eerste elftal, een succesvol transferbeleid om plekken op te vullen als een speler stopt met voetbal of weggekocht wordt, een excellent stadion, een sponsor met een gevulde portemonnee. En zo voorts. Echter, de kans dat alle indicatoren voor voetbalsucces

bij een club samenkomen is niet groot. In werkelijkheid kan nummer twee op de ranglijst er met enkele beste kwaliteiten ook uitspringen en nummer drie op ander vlak ook. Dan wordt de beste praktijk een mix van topscores van verschillende organisaties. Maar lukt het dan om die kunstmatige constructie ook in de werkelijkheid te realiseren? Dat is lang niet zeker. Wellicht kan het met de doorstroming van jonge spelers naar het eerste elftal alleen maar wat worden als een club ook in traditie heeft in deze en een organisatiestructuur en voorzieningen voor het opleiden van jongeren. Kortom, het beste op een indicator is niet zo maar over te planten van de ene op de andere organisatie.

Nu naar benchmarking. Een zevende veronderstelling betreft 'best practices' en het 'aanbrengen' van verbeteringen in producten van organisaties, diensten of processen. Stel dat het lukt om 'best practices' – in de internationale literatuur soms ook wel 'innovations' genoemd (Andrews, 2008: 191) - te traceren dan is een beste product-, dienstverlenings- of procespraktijk 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 en dat volop sprake is van onbegrensde 'transferability'. Alsof een hartchirurg zegt dat elk excellent kloppend hart van welke burger dan ook, uit Ivoorkust of Vietnam, is over te plaatsen naar een patiënt uit Luxemburg, van welke leeftijd dan ook, die een nieuw hart moet en zal krijgen. En dat een goede praktijk van zeg eens interactieve beleidsvorming of van loketdiensten uit Porto Alegre (Brazilië) zo maar overdraagbaar is naar Gdansk (Polen), Zürich (Zwitserland) of Vlagtwedde (Nederland).

In feite gaat achter het aanbrengen van verbeteringen door middel van een opgespoorde goede praktijk *mechanisch denken* schuil. Alsof je met een mecanodoos speelt, tal van elementjes al wel hebt maar nog een goed dopje, blokje of schroefje mist om een vliegtuigje of hijskraan te bouwen, of niet ziet hoe het vliegtuig af te bouwen met bestaande elementen. Voor niet-insiders: mecano is montagespeelgoed, in de jaren zestig voornamelijk gebruikt door jongetjes onder pakweg 12-16 jaar, bestaande uit honderden elementen, te bevestigen met schroefjes en moertjes. Wie mechanisch kijkt en denkt, beschouwt bouwen als spieken bij iemand die al een vliegtuigje maakte. Hij of zij ziet in benchmarking een benadering om elementen overal vandaan te halen en te gebruiken. Alsof de hele wereld uit mecano of uit lego bestaat en alles inpasbaar is. In werkelijkheid is dat niet het geval.

Een waarschuwing is op zijn plaats. In het vriendenboek voor organisatiesocioloog Cor Lammers '*Het organisatorisch labyrint*' lezen we: 'in wezen stelt de contingentietheorie dat er geen optimale organisatievorm bestaat; het aantrekkelijke van een bepaalde structuur is afhankelijk van de omgeving' (Van Doorn, Meurs en Mijs, 1993: 236; Aardema, 2000). Is sprake van dynamiek in de omgeving dan bezitten vooral organische organisatievormen een vermogen tot aanpassing. Mecanodozen zijn, bij wijze van spreken, niet meer geschikt als er naast vliegtuigjes halsoverkop ook raketten of drones gebouwd moeten worden.

Uit onderzoek is verder 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 (Korsten, 2005). Zo is het op meer gebieden. De context doet ertoe.

'Best practices' zijn dus niet zonder meer overdraagbaar vanwege dit verschil in context. Dat bleek in de praktijk ook uit een toekomstgerichte visitatie van Vlaamse grote steden in opdracht van de Vlaamse overheid waarbij een van ons sterk betrokken was (Korsten, 2012). De visitatiecommissieleden aarzelden steeds sterk om een gemeentebestuur een praktijk aan te raden. 'Beste praktijken' verschillen in organisaties naar tijd, plaats en context (Andrews, 2008: 171). De visitatiecommissieleden opereerden dan ook voorzichtiger. Ze suggereerden voorzichtig om een te gaan kijken in gemeente X of Y en te overleggen.

Andrews (2008) accentueert in *'Best Practices in Local Government'* het belang van de context als rem op de overname van 'best practices'. *'Context matters'*. Bij context kan het bijvoorbeeld gaan om het rechtssysteem per land, de sociaal-economische omstandigheden of de politieke cultuur in gemeenten (Andrews, 2008: 191). Zij stelt: het is al heel wat als je weet wat een goede praktijk is. 'Best practices' zijn dus 'geen panacee' met magische kracht, want veel hangt af van de vergelijkbaarheid van de context (Andrews, 2008: 191).

Wat te doen? Als onderzoeker van een benchmark moet je goed proberen de kenmerken van onderzochte diensten, producten of processen en achtergronden in beeld te krijgen. Een voorbeeld. De gemeente Venray heeft hoge kosten voor wegonderhoud maar het is dan ook een heel uitgestrekte gemeente met veel daarin woonachtige landbouwers die met hun grote en vaak brede landbouwvoertuigen gebruik maken van de vele, kilometerslange verharde wegen. Dat verklaart veel. Een dergelijke gemeente is goed vergelijkbaar met een andere Nederlandse gemeente met (bijna) even veel oppervlakte, landbouwverkeer en wegen. Of zo een gemeente bestaat?

Er is nog een andere kanttekening bij gebruik van 'best practices' te geven. 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 zich optrekken aan de 'best practice' dus geen vereiste dat blind gevolgd moet worden.

8 Verbetering altijd mogelijk

Achter benchmarking gaat ook de veronderstelling schuil dat prestaties van organisaties altijd te verbeteren zijn. Het kan altijd beter, zo luidt de redenering.

Daar staat dan tegenover dat bij benchmarking lang niet altijd gekeken wordt naar wat aan kenmerken van een te meten object beïnvloedbaar is en aangrijpingspunt voor verbetering kan zijn. Dat is nadelig want beleidvoerders willen juist beleid richten op het veranderbare (De Vries en Lako, 2009: 76). Metaforisch gesproken: kunstmest gooien op een rotsige bodem werkt niet groeibevorderend want er is niks om te laten groeien. Dat een praktijk altijd verbeterbaar is, is dus de vraag.

Een praktijk is ook moeilijk verbeterbaar als al sprake is van een kwalitatief goede praktijk. Benchmarking van 'paradijzen' geeft niet veel zicht op verbeterruimte.

9 Benchmarking leidt tot meer geloofwaardigheid van de onderzochte organisaties

De veronderstelling is dat benchmarking leidt tot meer geloofwaardigheid bij goede uitkomsten en tot minder geloofwaardigheid bij burgers en afnemers van diensten als de uitkomsten kritisch of ronduit negatief zijn (Borst, Lako en De Vries, 2012: 15). Wellicht is dit echter te instrumentalistisch gedacht. Immers, deze relatie leggen impliceert dat aan uitkomsten van benchmarking grote bekendheid wordt gegeven en burgers volop kennis nemen van deze uitkomsten. Dat zal lang niet altijd het geval zijn.

10 Benchmarking is belangenvrij

We komen bij een laatste hier te noemen veronderstelling. Het initiatief tot benchmarking wordt in het algemeen genomen om overheidsactiviteiten, zoals dienstverlening, eens te vergelijken met andere organisaties, te rangordenen, zich als overheid te verantwoorden en pogingen tot verbetering in gang te zetten. Dat is de officiële leer. In feite is dit te zien als een sprookjesverhaal zonder 'boze wolf' of 'heks'. Wat daarvan te denken?

Downs (1967: 80), Niskanen (1971a, b) en Dunleavy (1991) zijn vanuit een belangengerichte optiek op overheidsorganisaties ('public choice') sceptisch. Vanuit hun denken is benchmarking niet belangenvrij, maar te plaatsen in een perspectief van de vraag: kan een initiatief tot benchmarking

en een onderzoek met bevindingen het ambtelijk bureau of de eigen afdeling positief dienen? Kan benchmarking wellicht niet alleen een verbinding aangaan met eigen belangen van interne bureaus maar eventueel ook met categorieën ambtenaren of leidinggevenden? Genoemde auteurs stellen dat ambtenaren voorstander zijn van maatregelen die het eigen bureau laten groeien en floreren, en/of die het budget van het eigen bureau vergroten, en/of die het prestige van het eigen bureau doen groeien. Ook kan een rol spelen dat de positie van individuele ambtenaren versterkt wordt. Volgens deze belangengerichte wijze van kijken naar bureaucratieën strekt benchmarking ertoe om een organisatie of organisatiedeel te beschermen, te verdedigen of versterken (Borst, Lako en De Vries, 2012: 19). Zo doorgeredeneerd, zou benchmarking vooral daar gestimuleerd worden waar een dienst onder vuur ligt, om zo te proberen het beeld te corrigeren, of daar gestimuleerd worden waar een dienst toch al erg goed functioneert, om zo te proberen dat beeld ook extern zichtbaar te maken.

Deze verwachting is niet helemaal uitgesloten maar gaat wel erg ver. Het beeld lijkt zwaar overtrokken omdat bureaus en afdelingen in de huidige tijd niet een vaste eenheid zijn die slechts met 'één mond praat'. Veel werk wordt gedaan in projecten door programma- of projectteams die bestaan uit medewerkers van verschillende afdelingen (Hood & Dunsire, 1981; Goodsell, 1985; Korsten & Toonen, 1988; Wilson, 1989; Peters, 2001). Benchmarking is in de praktijk bovendien vaak aanbodgedreven. Het zijn onderzoeksbureaus die zich na stimulansen uit een sector of een coördinerende organisatie melden met een voorstel, met het verzoek aan een onderzoek om mee te doen. Daar staat tegenover dat het goed is geen al te naïeve kijk op de deelname aan benchmarking ten toon te spreiden. Vooralsnog houden we echter vast aan de veronderstelling dat de deelname aan benchmarking belangenvrij is. Dat is een betrekkelijk naïef standpunt. Gemeentebesturen doen nogal eens mee aan een benchmark om 'goede wil' te tonen, om te laten zien dat kwaliteitsstreven het bestuur niet vreemd is. Het komt zeker voor dat men meedoet vanuit de verwachting positief uit de bus te komen.

Het percentage ambtenaren dat positief is over prestatiemeting is in het algemeen hoog. Daaraan is toe te voegen dat veel ambtenaren hopen dat prestatiemeting '*bureaucracy bashing*' vermindert. De hoop is dat prestatiemeting het negatieve imago van ambtenaren corrigeert (Borst, Lako en De Vries, 2012: 29).

Resumé

We gaan er ondanks de geformuleerde kanttekeningen vanuit dat de genoemde veronderstellingen ten aanzien van benchmarking juist blijken te zijn:

1. appels zijn met appels te vergelijken;
2. de stap van kwaliteit naar kwantiteit met onzekerheidsreductie als gevolg, treedt op en bevordert begrip;
3. een organisatie die meedoet aan een benchmark is sterk betrokken bij de opdracht;
4. de kern van een benchmarkobject is door de keuze van (kwaliteits)indicatoren te vatten;
5. er is een meerjarenperspectief mogelijk in de datavergelijking;
6. een oorzakenanalyse van prestatieverschillen maakt deel uit van benchmarking (wat maakt dat deze organisatie aan de top staat op het vlak van X of Y of Z?);
7. goede praktijken worden in kaart gebracht en zullen overdraagbaar zijn naar andere organisaties;
8. prestaties zijn steeds te verbeteren;
9. positieve uitkomsten van benchmarking versterken de geloofwaardigheid van de onderzochte organisaties;
10. en het benchmarkproces wordt niet primair opgezet om een ambtelijk onderdeel in de vaart der volkeren op te stuwen of een specifieke categorie ambtenaren te dienen.

De werkelijkheid is echter niet zelden een andere. Dat zal blijken.

6.2 Bezinning op veronderstellingen

Reflectie op veronderstellingen (1): Organisaties kunnen niet altijd verbeteren

De veronderstelling is discutabel dat organisaties altijd kunnen verbeteren. Een politieorganisatie kan bijvoorbeeld geen hoger oplossingspercentage van misdrijven dan 100 procent halen (Blank, Felsö en De Groot, 2011; Hulst en De Groot, 2011; Bouckaert e.a., 2011; Peeters & Verschraegen, 2013). Ook is sprake van paradoxen. Wie meer misdrijven registreert, maakt de vijver groter. Als het aantal oplossingen van misdrijven dan iets groeit, kan het gevolg zijn dat het oplossingspercentage toch daalt ten opzichte van eerdere jaren (Wittebrood, 2002: 64).

Toegespitst op gemeentelijke benchmarking blijkt verder het volgende.

Reflectie (2): Niet alle beleid wordt onderworpen aan benchmarking

De meeste gemeentelijke taakgebieden en ambities en daarmee het meeste gemeentelijk beleid blijkt helemaal *niet* aan benchmarking onderworpen te worden omdat dit beleid zich blijkbaar niet kwantitatief laat vangen, andere waarden dan kwantiteiten een rol spelen en specifieke aanleidingen om benchmarking te starten blijkbaar ontbreken. Bureaus bieden benchmarking op bepaalde gebieden ook helemaal niet aan. Kent u een benchmark van lokaal armoedebeleid, van lokaal daklozenbeleid, van omgaan met illegalen, van ruimtelijke ordening in al zijn aspecten? 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, maar tal van zaken worden niet in kengetallen gevangen (Ter Bogt, 2005). 'Outcomes' zijn moeilijk meetbaar en daarom ziet men in begrotingen dat zicht op effecten hier en daar in (programma)begrotingen ontbreekt (De Vries en Lako, 2009: 74).

Reflectie (3): De kern vatten en verschillenanalyse

Voor zover benchmarking wel voorkomt, blijkt dat de kernvariabelen in benchmarkings om een prestatie te meten niet altijd gevonden worden, cijfers over een reeks jaren soms ontbreken, een oorzakenanalyse van prestatieverschillen tussen organisaties 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). Diverse veronderstellingen blijken bij nader inzien dus discutabel.

Reflectie (4): Beste praktijk soms niet te vinden

Een volgende waarneming bestaat hierin dat een benchmarkonderzoeker wel op zoek kan gaan naar de beste praktijken maar soms valt er geen beste praktijk te vinden.

Als twintig gemeenten op het gebied van klanttevredenheid over dienstverlening een 6.5 op een schaal van 1-10 scoren dan is het onmogelijk om een van deze gemeenten als de beste aan te duiden. De praktijk van benchmarking wijst ook uit dat de verschillen in scores soms heel klein zijn. Abma en Korsten (2009) en Abma (2014) hebben dit waargenomen toen ze alle scores uit 'De Staat van de Gemeente' casu quo waarstaatjegemeente op een rij zetten. Als, om bij het voorbeeld te blijven, de scores van klanttevredenheid bij gemeenten gemiddeld 6.5 zijn, maar de spreiding rond dit gemiddelde heel gering is en erop neerkomt dat de meeste gemeenten die geen 6.5 scoren een klanttevredenheid van 6.6 of 6.4 scoren, dan heeft het zoeken naar een beste praktijk geen zin. Dan is eerder sprake van een betrekkelijk *uniform* beeld. Dit verschijnsel kan optreden bij wat mechanisch medebewind genoemd wordt, dat wil zeggen een situatie waarbij gemeenten overal in Nederland op dezelfde wijze documenten verstekken op een manier die door de rijksoverheid is voorgeschreven.

Reflectie (5): Er valt ook te leren van slechte praktijken

Kan een organisatie door middel van benchmarking alleen maar leren van beste praktijken? Een organisatie kan leren van zowel goede als slechte praktijken die openbaar worden ('bad practices');

hoge score op een misère-index). Van slechte praktijken misschien nog wel meer dan van goede praktijken.

Ziekenhuizen zijn al jaren onderwerp van benchmarks. BDO (2012) verrichtte op basis van jaarverslagen van algemene ziekenhuizen een benchmark van de financiële performance, waaruit bleek dat middelgrote ziekenhuizen het over het algemeen 'goed doen'. Maar naast de jaarverslagen is er ook een andere werkelijkheid. Diverse ziekenhuizen kwamen in de periode 2000-2017 zwaar onder vuur te liggen. Denk bij voorbeeld aan het Orbisziekenhuis (Sittard-Geleen), het Catharinaziekenhuis (Eindhoven), het Ruwaard van Puttenziekenhuis (Spijkenisse), het Langeland ziekenhuis (Zoetermeer), het Slotervaartziekenhuis (Amsterdam) en het VUmc (Amsterdam). Daarbij is de aandacht onder meer gericht geweest op zwakten van en in het bestuur en in het toezichthoudend orgaan, de relatie tussen bestuur en het toezichtorgaan, en de relatie van het bestuur met de organisatie en de primaire processen die daarin plaatsvinden (ook Euwes, 2006; Goodijk, 2012).

Het VUmc lichten we er hier uit om te illustreren of er te leren valt. Het ziekenhuis werd in augustus 2012 door de Inspectie voor de Gezondheidszorg onder verscherpt toezicht geplaatst. Directe aanleiding was het onjuist verstrekken van informatie en het geven van een verkeerde voorstelling van zaken door de toenmalige raad van bestuur over ernstige samenwerkingsproblemen tussen medisch specialisten. Voorafgaand aan deze onder toezichtstelling hield de Inspectie voor de Gezondheidszorg (IGZ) al de vinger aan de pols bij het VUmc naar aanleiding van een melding bij de inspectie van het overlijden van een patiënt op de intensive care in maart 2011. Dat leidde eind 2011 tot een verbeterplan van de inspectie met veertien maatregelen. Een aantal maanden later, in mei 2012, vertelde de raad van bestuur aan deze inspectie dat de samenwerking tussen specialisten dankzij mediation inmiddels weer 'goed liep'. In juni 2012 werd toegevoegd dat een flinke cultuuromslag had plaatsgevonden. Geen hommeles meer in het hospitaal, het VUmc, was de gedachte die uitgedragen werd.

Echter, in augustus 2012 bleek in het radioprogramma Argos dat er nog wel degelijk sprake was van problemen. Toen traden conflicten tussen specialisten aan het daglicht die eerder schriftelijk aan het bestuur gemeld waren maar 'onder de pet gehouden waren'. Deze problemen vielen niet op te lossen via mediation of een paar ontslagen. De inspectie riep de raad van bestuur daarop op het matje en uit dat gesprek bleek dat de inspectie niet goed geïnformeerd was door de raad. De inspectie liet toen in het midden of dat al dan niet bewust was gebeurd. Hoe dit ook zij, de raad van bestuur had onvermogen laten zien. Het besef van de ernst van de zaak was bij de raad van bestuur te gering en het vertrouwen tussen bestuur en inspectie was geschonden. De raad had onvoldoende leiderschap getoond.

Na augustus 2012 stapten bestuursleden en meerdere toezichthouders op. Het nieuwe bestuur heeft daarop met de staf en andere medewerkers stappen gezet om de problemen op te lossen, en te komen tot een betere samenwerking en een 'veilige' cultuur in het ziekenhuis (lees: geen angstcultuur meer).

Twee rapporten zijn vervolgens van belang. In december 2012 verschijnt het Eindrapport van de Onderzoekcommissie Toezichtfunctie VUmc onder de titel '*Weer in contact komen*' (Grinwis e.a., 2012) en in juli 2013 volgt het rapport van de Inspectie voor de Gezondheidszorg '*Met de patiënt voor ogen*' (IGZ, 2013). Het eerstgenoemde rapport was bestemd voor bezinning binnen de raad van toezicht op het eigen handelen. Eisen die aan toezichthouders gesteld worden zijn volgens de rapporteurs van dat rapport: kennis van de sector en van het primair proces; gevoel voor zwakke signalen, dat wil zeggen voor het latente, voor wat zich nog in de onderstromen bevindt en dus voor wat nog niet manifest aan de oppervlakte is getreden; en vermogen om op te treden in geval van een vertrouwensbreuk en dreigende reputatieschade.

Een van de conclusies van Grinwis c.s. luidt letterlijk:

‘Vanaf 2007 tot mei 2012 kan niet worden gesproken van een adequate wijze van toezicht op het VUmc. De kwaliteit van het toezicht komt vanaf ca. 2007 geleidelijk onder druk te staan. Dat komt door een combinatie van factoren, zoals de deels opportunistische werkwijze met betrekking tot de samenstelling van de RvT als gevolg van de beoogde fusies waardoor de RvT onvoldoende grip had op de eigen samenstelling. Hoewel de toezichtstaak veel complexer werd, accepteerde de RvT deze gang van zaken kennelijk bewust. De RvT bestond vanaf eind 2008 tot eind 2011 uit 9 en enkele maanden zelfs uit 10 leden en hield toezicht op 5 instellingen. In de literatuur is bekend dat naarmate een raad van commissarissen groter is, de neiging van het bestuur om aan commissarissen (vertrouwelijke) informatie te onthouden groter zal zijn. Daar kwam nog bij dat door de invoering van de divisiestructuur (2007) het bestuur op grotere afstand van de operatie geraakte. De RvT organiseerde in deze periode onvoldoende een eigen informatievoorziening met de werkorganisatie om deze afstand te overbruggen’ (Grinwis e.a., 2012: 16). Toegevoegd wordt dat het bestuur de grip op de organisatie kwijt raakte op een steeds complexer worden bedrijf. Als dat het geval is, moet een raad van toezicht ingrijpen. De raad van toezicht trad echter niet op. De toezichthouders hebben onvoldoende de kwaliteit en continuïteit van het bestuur gewaarborgd (Grinwis e.a., 2012: 26).

Dit zijn ‘dodelijke formuleringen’, maar het IGZ-rapport van juli 2013 is een nog opmerkelijker rapport. Het handelt over het verscherpte toezicht van het IGZ op het VUmc in de periode augustus 2012 – maart 2013. De inspectie denkt, aldus het rapport uit juli 2013, dat andere ziekenhuizen ervan kunnen leren om soortgelijke situaties waarin patiëntveiligheid in het geding is te herkennen, te voorkomen of te herstellen. Volgens het rapport is sprake geweest van grote samenwerkingsproblemen tussen medisch specialisten en onjuist en onvoldoende handelen van de raad van bestuur van het VUmc. Het rapport bevat veel lesstof. Dit ‘leren’ valt ook anders te formuleren. Het rapport had net zo goed kunnen handelen over een ziekenhuis in elk van de andere provincies in Nederland. Alleen wat in het Amsterdamse ziekenhuis zichtbaar werd, was buitensporig.

Het rapport, dat zich laat lezen als een ware thriller met educatieve lading, is een vernietigend verslag van misstanden in het academisch ziekenhuis, het VUmc. Ook in andere ziekenhuizen is aan het licht gekomen dat artsen weliswaar goed opgeleid kunnen zijn en vakbekwaam maar dat ze lang niet altijd goed zijn in samenwerken. De website medischconflict.nl laat zien dat in de helft van de Nederlandse ziekenhuizen bepaalde medisch specialisten conflicten met elkaar hebben. In vier op elke tien ziekenhuizen is er sprake van conflicten tussen medisch specialisten en de raad van bestuur. Vaak zijn er in ziekenhuizen ook conflicten met patiënten en teamconflicten, waarbij verpleegkundigen overhoop liggen met artsen. Maar volgens het IGZ spande het VUmc de kroon. Dit ziekenhuis heeft zich bij voorbeeld ook jarenlang onttrokken aan landelijke registratieregelingen voor medische fouten en sterfgevallen. Er is een goed meldsysteem nodig bij medische calamiteiten, het is wenselijk geregeld functioneringsgesprekken te houden, ook met medisch specialisten. En er moet wel degelijk nog gewerkt worden aan een cultuur waarin patiëntveiligheid centraal staat. Het IGZ-rapport geeft nog tal van andere aanbevelingen.

Deze casus bevat een rijkdom aan informatie die niet uit een benchmark van ziekenhuizen zou zijn te halen. Op de betekenis of rijkdom van deze casus, uit educatief oogpunt gezien, is van invloed geweest dat meerdere elkaar opvolgende analyses zijn verricht. Deze rijke verzameling aan analyses gaat daarmee een reguliere comparatieve benchmark te boven. Wat we hier echter willen benadrukken is dat er van slechte praktijken te leren valt, zoals de inspectie zelf ook aangeeft.

6.3 Enkele conclusies

Dit hoofdstuk laat zien dat benchmarking op een aantal veronderstellingen berust. We noemen er enkele.

Veronderstelling 1: Een organisatie die meedoet aan een benchmark is sterk betrokken bij de opdracht, zodat ze ook kan meedenken over de keuze van de indicatoren, de analyse van goede praktijken, het achterhalen van verschillen tussen prestatiescores, een heldere rapportage. Deze

veronderstelling is gebaseerd op de aanname dat meedenken tot meer legitimiteit leidt en de kans op benutting later toeneemt.

Veronderstelling 2: Organisaties die vergeleken worden, moeten ook vergelijkbaar zijn. Vergelijk dus niet de eilandgemeente Texel met de grensgemeente Vaals, hoewel beide te maken hebben met toerisme.

Veronderstelling 3: De kern van een benchmarkobject is door de keuze van (kwaliteits)indicatoren te vatten. Er is geen sprake van marginale criteria.

Veronderstelling 4: Er is een meerjarenperspectief mogelijk in de datavergelijking. Daardoor worden toevalstreffers in het presteren uitgebannen.

Veronderstelling 5: Een oorzakenanalyse van prestatieverschillen maakt deel uit van benchmarking. Blijft die uit dan is de rangorde niet verklaarbaar en blijft onbekend waarom er zwakkere en betere praktijken zijn.

Veronderstelling 6: Goede praktijken zijn traceerbaar en zullen overdraagbaar zijn naar andere organisaties.

Veronderstelling 7: Prestaties zijn steeds te verbeteren. Er is steeds verbeterruimte.

Veronderstelling 8: Positieve uitkomsten van benchmarking versterken de geloofwaardigheid van de onderzochte organisatie.

Veronderstelling 9: Het benchmarkproces wordt niet primair opgezet om een ambtelijk onderdeel in de vaart der volkeren op te stuwen of een specifieke categorie ambtenaren te dienen.

Deze veronderstellingen zullen verderop nog aan de orde zijn. Hier zijn er al enkele kanttekeningen bij geplaatst. Ten eerste, veel benchmarkings schetsen een kwantitatief beeld van een onderzocht object (product, proces, dienst), geven een vergelijking tussen organisaties en een rangorde. Benchmarking opgevat als ranking is ontegenzeggelijk waarneembaar. Maar de benchmarkings schetsen niet de beste praktijk op een aantal variabelen in samenhang. Ze geven een constructie van beste scores, een optelsom van wat eruit springt. En ze geven geen analyse van prestatieverschillen (waarom staat die en die aan de top en die en die in de achterhoede?) en rapporteren daar dus ook niet over. Dat zijn tekortkomingen die tot gevolg hebben dat de lezers en gebruikers van benchmarks onthand zijn.

Eén bevinding verwijst naar de soms blijkende onvindbaarheid van beste praktijken. Dat een beste praktijk wel gezocht kan worden, maar soms eenvoudig niet te vinden is, komt als de scores in de praktijk heel dicht bij elkaar liggen. Qua beoordeling van diensten, producten of processen functioneren alle organisaties dan hetzij matig, hetzij voldoende, hetzij goed. Onderzoekers stellen dan dat de spreiding rond het gemiddelde gering is.

Ook is aangegeven dat er een te sterke focus op de beste praktijk kan liggen waardoor leren van slechte praktijken over het hoofd wordt gezien. 'Bad practices' kunnen ook interessant zijn om te analyseren; soms op zich al en soms vanwege de contrastwerking met goede praktijken.

We benadrukken hier tevens dat benchmarking die gericht is op 'best practices' een te mechanische kijk op verbeterprocessen kan opleveren. Alsof de context er niet toe doet. Een beste praktijk is niet zonder meer een module die je aan een fiets kunt hangen zoals je aan elke fiets een fietsbel monteert.

Past benchmarking in netwerkbestuur? In het openbaar bestuur kan een overheidsorganisatie gezien worden als een organisatie die over alle taken, bevoegdheden en middelen beschikt. Als dat het geval is, wordt gesproken van een organisatie met volledige doorzettingsmacht. Zo een organisatie is dan niet afhankelijk van andere organisaties. In die situatie wordt wel gesproken van monocentrisme, cockpitbestuur of hiërarchisch bestuur ('government'). Daar staat een organisatie tegenover die juist geen doorzettingsmacht heeft, maar afhankelijk is van andere (overheids)organisaties die ook beschikken over taken, bevoegdheden en/of middelen. Als dat het geval is, kan geen organisatie zo maar de wil aan een of meer andere organisaties opleggen want die

andere beschikken over remmacht of vetomacht. Dat is bij voorbeeld zo in de verhouding tussen een gemeente en een woningcorporatie. Het gemeentebestuur plant woningen en een corporatie bouwt ze. In dergelijke situaties van netwerkbestuur zal deliberatie vruchtbaar kunnen zijn. Partijen moeten zich verplaatsen in elkaars belangen, de probleemperceptie delen, elkaars kennis delen en fixaties opruimen die het elkaar verstaan en begrijpen belemmeren en van daaruit komen tot het verbinden van doelen en middelen. Benchmarking blijkt niet goed te passen in de context van netwerkbestuur ('governance'). Benchmarking meet niet de prestaties in beleidsnetwerken.

7 Uitkomsten van benchmarking: benchmarks centraal

Benchmarking kan gericht zijn op financieel-economische aspecten van overheidshandelen, waaronder doelmatigheid. Maar benchmarking is niet uitsluitend gericht op doelmatigheid. Dat blijkt uit dit hoofdstuk. We geven enkele voorbeelden van een dergelijk voor ons traceerbaar onderzoek gericht op gemeenten om te zien waarover ze handelen. Het gaat om de volgende domeinen: servicegerichtheid; burgerzaken; afvalbeheer; sportzaken; werk en inkomen; gemeentelijke overhead; duurzame mobiliteit. De betrokken onderzoeksorganisaties zijn Berenschot, SGBO/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).

7.1 Eerste voorbeeld: benchmark servicegerichtheid

Er is benchmarkonderzoek gedaan naar de kwaliteit van dienstverlening. Onderzoek volgens 'De Staat van de Gemeente' en de opvolger daarvan [waarstaatjegemeente.nl](http://www.waarstaatjegemeente.nl) laat zien dat gemeenten op basis van gegevens over het omgaan met burgers in verschillende rollen, burgers in de rol van klant over het algemeen voldoende of beter tegemoet treden (Abma en Korsten, 2009; Abma, 2014).

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 in termen van 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 is na te gaan wat gemeenten als opdrachtgevers hebben gedaan met de resultaten van de benchmark van de opdrachtnemer.

7.2 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 bedraagt volgens de benchmark burgerzaken in het laatste onderzoeksjaar 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 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.

7.3 Derde voorbeeld: benchmark afvalbeheer

We komen bij een derde benchmark. 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 procent 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 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 sterk *relatieveert*.

Dat gemeenten met meer dan 100.000 inwoners, 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 Jos Blank en Hans de Groot (2012).

7.4 Vierde voorbeeld: benchmark sport

Is de sportdeelname hoog? Wordt veel geld gespendeerd aan accommodatieonderhoud? Onderzoeksbureau SGB0 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 meer dere 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 SGB0 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(re) gemeenten (tussen 20.000-45.000 inwoners). Wat bleek? Vrijwel alle gemeenten gebruiken een sportnota als basis voor beleidsontwikkeling. 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 gemeentelijk 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 sportaccommodaties. De middelgrote gemeenten zetten juist weer in op subsidies aan sportverenigingen 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 deze gemeentebesturen meer investeren in het inhoudelijk sportbeleid dan de andere gemeentebesturen.

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 procent. De gymlokalen worden eveneens meer overdag (65 procent) gebruikt dan in de avonduren (46 procent) 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.

7.5 Vijfde voorbeeld: benchmark werk en inkomen

We komen bij een vijfde benchmark. 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. Voor 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 er ook wordt 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.

7.6 Zesde voorbeeld: benchmark overhead gemeenten

Hoe verhoudt de omvang van de overhead van de gemeentelijke organisatie 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 en 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).

7.7 Duurzame mobiliteit

In 2018 verscheen voor het eerst een nieuwe, tweejaarlijkse benchmark die de beleidsinspanningen van dertig steden op het gebied van duurzame mobiliteit in beeld brengt. Het onderzoek focust op de maatregelen die gemeentebesturen nemen om lopen, fietsen en openbaar vervoer te

stimuleren. Ook het beleid om te komen tot een schoner en efficiënter vervoer wordt onder de loep genomen. De benchmark is uitgevoerd door CE Delft, in opdracht van onder meer Natuur & Milieu, de Fietsersbond, Rover en Milieudefensie. Het onderzoek moet de gemeenten ertoe bewegen om na de gemeenteraadsverkiezingen van 21 maart 2018 'groene college-akkoorden' te sluiten. In de benchmark komen thema's aan de orde als: schoon en efficiënt autogebruik (milieuzones; laadpalen voor hybride auto's; parkeertarieven); het openbaar vervoer (schone bussen?; bereikbaarheid ov); de fiets (ov-fietsen; fietspaden; bewaakte stallingen?); wandelen (wandelpaden?; aandeel van weg met max. snelheid van 15 of 30 km). Er is ook aandacht voor beleidseffecten ('policy outcomes'), zoals de effecten van het verkeer op de omgeving (in termen van verschillende soorten emissies, geluid, verkeersdoden en gewonden) en de keuze van de vervoerswijze (aandeel van lopen, fietsen, en gebruik van de trein, bus, tram, metro in het verkeer).

Fig. 9: Positie van Enschede in de rangorde van steden op het vlak van duurzame mobiliteit (2018)

	Selectie van indicatoren/ clusters	Rangschikking (plaats tussen 30 gemeenten) 2018	Eenheid	Score
Effect verkeer op omgeving	CO2-emissie	15	456.4	6.3
	NOx-emissie	16	1.176.9	6.0
	PM2,5-emissie	14	49.3	6.4
	Geluid	7	13%	7.4
	Verkeersdoden	25	0.3.	4.1
	Ernstig gewonden in verkeer	13	2.0	6.7
	Fietsongevallen	18	63.3	6.4
	Totaal effect omgeving	16	5.2
Vervoerswijze	Aandeel fiets	8	36%	6.9
	Aandeel lopen	24	17%	4.5
	Aandeel km. trein	25	9%	4.3
	Aandeel km. bus/tram/metro	12	3%	5.7
	Aandeel niet-auto	16	57%	5.5
Wandelen	Aandeel 15 en 39 km-wegen	2	70%	8.1
Fietsen	OV-fietsen	23	0.5	4.8
	Fietspaden	15	127	5.9
	Relatieve bereikbaarheid per fiets	10	34%	6.7
	Totaal fietsen	16	5.9
Openbaar vervoer	Schone bussen	19	5.1	5.1
	Relatieve bereikbaarheid per OV	4	79%	7.8
Schoon en efficiënt autogebruik	Laadpalen	29	68	3.7
	Stadslogistiek	1	8	8.0
	Milieuzones	13	4	4.0
	Parkeertarieven	6	7.0
	Totaal schoon en efficiënt autogebruik	11	...	5.7
Totaal	Voorzieningen en beleid	14	5.9

In de eerste editie van de benchmark blijkt dat Leeuwarden, Nijmegen en Maastricht het slechtst scoren op het gebied van CO2-uitstoot door het wegverkeer. Bij geluidsbelasting door verkeer bungelen dichtbebouwde steden als Haarlem, Amsterdam en Rotterdam onderaan. Met de fietsveiligheid is het veruit het slechtst gesteld in Zwolle, Rotterdam en toch ook Groningen. Dat blijkt uit registratie van het aantal fietsers dat binnen de bebouwde kom bij een verkeersongeluk om het leven kwam (gedeeld door het totale aantal inwoners). Een opmerkelijke uitkomst omdat Zwolle in een recent ANWB-onderzoek naar voren kwam als voorbeeldgemeente op het gebied van door fietsers ervaren veiligheid. Wellicht wordt de slechte score beïnvloed door het grote aantal Zwollenaren dat geneigd is voor binnenstedelijk vervoer de fiets te pakken (na studentenstad Groningen het meest). In Rotterdam, Almere en Heerlen wordt relatief het minste gefietst.

7.8 Algemeen beeld

Welk algemeen beeld komt nu naar voren uit al die benchmarking van gemeenten?

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 behoort ook tot de mogelijkheden.

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).

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.

7.9 Kanttekening

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

Verschillen zichtbaar: 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 verschil verklarende 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 vals beeld. Maar ook blijkt dat de organisatiewijze van afvalbeheer toch een beetje uitmaakt.

Betrokkenheid organiseren: 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. Dat bevordert de kennisname van resultaten. Bovendien neemt de betrokkenheid toe, wat later van belang kan zijn om de bevindingen uit de benchmark 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.

8 Uitkomsten van partiële benchmarking: gemeenten centraal

In vorige hoofdstukken zijn diverse benchmarks de revue gepasseerd. Maar wat is het beeld van benchmarking als we uitgaan van gemeenten, dus als per gemeente bezien wordt wat de opstelling van het bestuur en ambtelijk leidinggevenden is? Deze vraag laat zich uitsplitsen naar vragen als de volgende: Tonen gemeentelijke ambtenaren een actieve of passieve opstelling ten opzichte van benchmarkinformatie? Houden leidinggevende ambtenaren van een dienst of afdeling benchmarkrapporten voor zich of worden ze in een deliberatieve setting breder besproken? Is de bespreking eenmalig of niet? Klimt iemand of een werkgroep in de PC om een lijst met overwegingen en/of aanbevelingen toe te voegen over wat te doen staat, afgestemd op de eigen gemeente? Worden benchmarkrapporten doorgestuurd naar de burgemeester en andere collegeleden? Zijn wethouders op hun beurt eigenlijk ontvankelijk voor benchmarkinformatie? Zijn er verschillen tussen gemeenten in het omgaan met benchmarks?

Dit soort vragen worden bezien vanuit één benchmark, die van bouw- en woningtoezicht. We komen zo weer een stapje dichterbij de ontmystificering van benchmarking en dus bij het ware gezicht van benchmarking.

8.1 Benchmark bouw- en woningtoezicht

We vertrekken vanuit de benchmark bouw- en woningtoezicht (BWT). Deze BWT is in 1999 van start gegaan. Om een indruk van aantallen te geven, in 2009 waren er 27 deelnemende gemeenten. Cumulatief is het aantal inmiddels ver over de honderd gekomen. Rotterdam is een voorbeeld van een gemeente die sinds 2001 deelnam. Door het blijven deelnemen, ontstaat een meerjarig beeld, een mogelijkheid tot vergelijking in de tijd (o.a. Hoogwout, 2007).

Het bouw- en woningtoezicht wordt in de benchmark vanuit drie invalshoeken bezien: de invalshoek van de aanvrager (denk aan klantgemak, waardering van de dienstverlening), die van het bestuur (beleidseffectiviteit, kwaliteit van bouwtoezicht), en die van de organisatie (denk aan doelmatigheid).

De benchmark had en heeft tot doel om overheidsbesturen en betrokken ambtenaren te informeren over prestaties op het vlak van vergunningverlening en bouwtoezicht. Vanaf 2010 werd de benchmark aan gemeenten aangeboden in combinatie met een doorlichting, een kwaliteitsaudit in jargon, op de vergunningdossiers en een klanttevredenheidsonderzoek. Daardoor kunnen tal van prestaties niet alleen kwantitatief maar ook kwalitatief in beeld komen.

De benchmark BWT werd en wordt in rondes gehouden. Voorafgaand aan een benchmarkronde komen representanten van deelnemende gemeenten bijeen om de vragenlijst te bespreken en gezamenlijk vast te stellen welke aspecten en prestatie-indicatoren worden vergeleken en hoe. Achteraf vindt minimaal één evaluatiesessie plaats. In het verleden zijn ook thematische werkgroepen opgezet om bepaalde vragen verder uit te diepen. Daaruit zijn ook projecten voortgekomen.

Deelnemers ontvangen een geïndividualiseerde rapportage met resultaten, conclusies en aanbevelingen.

Zijn er concrete resultaten geboekt? Een aantal gemeenten heeft de BWT gebruikt om succesvol claims voor extra formatie over bouw- en woningtoezicht in de eigen organisatie te onderbouwen. De bevindingen zijn door deelnemende gemeenten ook benut om bij het opstellen, prioriteren en evalueren van eigen verbeteractiviteiten en voor de informatievoorziening.

De benchmarkactiviteiten zijn daarnaast de opstap geweest naar *sectorale professionalisering*. In 2003 ontstond de Vereniging BWT Nederland, is een kennisbank opgericht (BWTinfo.nl), volgde het ontstaan van een vakblad en volgden een groot aantal collectieve projecten, zoals de opstelling van

een collectief toetsings- en toezichtprotocol en een gegevensstandaard voor het ontsluiten van welstands informatie (IMWE). Dat is volgens ons indrukwekkend.

8.2 Enkele gevalstudies

We wezen al op tactisch en ander gebruik van BWT-benchmarkrapporten. Er is ook nader onderzoek gedaan naar gebruik van rapporten met uitkomsten van BWT's ('the uses of benchmark information') in de – hier geanonimiseerde – vier gemeenten B, E, Em en G. Het gaat bij deze gemeenten om middelgrote en grote gemeenten, kortweg om steden (Noordhoek, 2013). In dat onderzoek wordt aandacht besteed aan de disseminatie van benchmarkinformatie, de organisatie van de bespreking van benchmarkuitkomsten of het uitblijven daarvan, aan verbetervoorstellen, aan het gedeelde beeld van geïnterviewde respondenten over gebruik en ondergebruik van een benchmark en aan de zinvolheid van het doorsturen van rapportages naar bestuurders ('shared understandings'). Opvallend is dat Noordhoek spreekt over het onderkennen van *leerepisodes*. Eens per jaar een benchmarkrapport en dan eens per jaar een serieuze leerepisode, dat is eigenlijk te weinig, te dun (Noordhoek, 2013: 111). Meer episodes en meer vormen van gebruik en meer deliberaties zijn eigenlijk gewenst, volgens deze onderzoeker en auteur.

Eerste gemeente

Hoe moeten we ons het omgaan met deze benchmark in een gemeente voorstellen? Onderzoek in de gemeente B, een bestaande gemeente in Brabant, laat zien dat de benchmark op het vlak van bouw- en woningtoezicht voor de gemeente achterstanden toont, zoals op het vlak van handhaving. Elders oogt het gras inderdaad in een aantal opzichten groener. Maar de gemeente is in andere opzichten ook weer voorloper, zoals in de sfeer van de persoonlijke contacten met klanten. Dat doet de gemeente B in het toezicht goed.

Is er daarmee voldoende reden voor een bespreking van het benchmarkrapport door het afdelingshoofd met de leden van het managementteam? Inderdaad. Die bespreking volgt ook. Er wordt vervolgens een werkgroep ingesteld en er volgen enkele verbetervoorstellen. Dat oogt positief; niemand is tegen. Er zijn – in de woorden van onderzoekster Marike Noordhoek – dan ook meerdere *leerepisodes* traceerbaar.

Maar, opvallend, het rapport verlaat de afdeling niet. Het college van burgemeester en wethouders neemt er geen kennis van en ook de gemeenteraad niet. De afdeling stuurt het rapport namelijk *niet* naar de bestuurlijk verantwoordelijken. Het rapport wordt door de leidinggevende ambtenaren niet geschikt geacht voor bespreking door collegeleden en raadsleden. Daarbij speelt een rol of ambtenaren van mening zijn dat bestuurders kunnen omgaan met een benchmarkrapport. De leidinggevende op het gebied van bouw- en woningtoezicht kijkt daar *negatief* tegenaan. Wat moeten collegeleden met een vergelijkend rapport met hoofdzakelijk figuren en tabellen...? De inschatting is: *'daar kunnen "ze" niet echt mee omgaan.'* De gedachte komt niet op om een leeswijzer toe te voegen of een interpretatienota. Concluderend, hier zien we dus dat eigen opvattingen van ambtenaren over bestuurders en hoe de bestuurlijke arena 'functioneert' een rol spelen in de manier van omgaan met een benchmarkrapport.

Er is aan informanten uit de gemeente B, ook wel 'respondenten' genoemd, gevraagd hoe ze over het gebruik van de benchmarkuitkomsten denken. Tal van respondenten menen dat er – ondanks het zich voordoen van leerepisodes - op tactisch en operationeel niveau *weinig* met de benchmarkinformatie is gebeurd. *'Little has been done with the benchmark information'*, schrijft Noordhoek (2013: 107) dan ook.

'Weinig' moet worden begrepen als 'te weinig'. De bespreking van het benchmarkrapport had ambtelijk verbreed kunnen worden en had dus niet hoeven blijven 'hangen' in het managementteam of een werkgroep. Een respondent stelt dat het managementteam meer tijd en energie had kunnen

en moeten steken in het uitwerken van wat de figuren (tabellen, grafieken) betekenen en in de verspreiding van die informatie. Een andere respondent merkt op:

'De benchmark is niet onderdeel van onze gewone werkzaamheden geworden.'

Deze situatie is aan te duiden als 'onderbenutting' of 'ondergebruik' van de rapportinhoud. Deze onderbenutting verklaart volgens Noordhoek (2013) de gepercipieerde beperktheid van het benchmarkgebruik. In haar woorden: *'the perceived limited use of the benchmark information'* (Noordhoek, 2013: 107).

Achter de onderbenutting van benchmarkinformatie gaat nog meer schuil dat het vermelden waard is. Ten eerste de niet excellente toekomstgerichtheid van het gemeentebestuur. Deze houding wordt als een barrière voor benutting gezien. Ten tweede valt op wat de inschatting van ambtenaren is over hoe bestuurders omgaan 'met dit soort rapporten'. Over dit laatste iets meer.

Het niet toesturen van een benchmarkrapport heeft te maken met wat genoemd wordt 'impliciete kennis' en vaardigheden bij ambtenaren ('tacit knowledge'). Als afdelingshoofden de indruk hebben dat collegeleden geen interesse hebben in benchmarkrapporten en/of ze niet lezen en/of ze wel bekijken maar 'er niks mee willen doen' speelt dat een grote rol in het besluit om wel of niet te attenderen op een benchmarkrapport. Kort gesteld:

'Als ze geen interesse hebben of niet goed kunnen omgaan met de rapportages, dan sturen we ook niks.'

Want, zo redeneren deze ambtenaren, niks is zo erg als colleges van burgemeester en wethouders die iets doen met uitkomsten waaraan ambtenaren *niet* hebben gedacht en waarmee ambtenaren geen rekening hebben gehouden. Je moet je als ambtenaar zo weinig mogelijk laten verrassen door bestuurders. Dat zijn immers passanten. Het is of je midden in de Britse tv-serie *'Yes, minister'* zit of een recenter evenknie daarvan *'Yes, prime minister'*.

Er is in deze gemeente B dus zowel gebruik als ondergebruik (lees onderbenutting) van BWT-benchmarkinformatie. Wat is dan het hoofdbeeld? Dan wordt de wereld van het 'understatement' betreden. Ambtenaren kunnen 'naar buiten' klagen over gering gebruik van rapporten maar eveneens met een stalen gezicht stellen dat er natuurlijk van voldoende gebruik sprake is, daaraan toevoegend: *'we mogen het niet beleven dat ze die rapporten allemaal echt gaan lezen en bespreken'*. Immers, *'voor je het weet, gaan ze ons opzadelen met zaken waaraan wij geen behoefte hebben.'*

Dissertaties ter verkrijging van de doctor-titel zijn niet het geschikte voertuig voor 'understatements' maar Marike Noordhoek zegt in haar dissertatie toch:

'The municipality of B..... claims a wide variety of uses of benchmark information' (2013: 108).

Er wordt wel vooruit gedacht op basis van de benchmark, wat een 'pluspunt' is, maar respondenten zijn ook van mening dat op tactisch en operationeel niveau meer gebruik van het benchmarkrapport en van de uitkomsten mogelijk was geweest (een minpunt). De proceskant van benchmarking in 'the benchmark network' had versterkt kunnen worden. Maar dat is een beeld van ambtenaren.

Wat is het beeld van en bij bestuurders, zoals wethouders? Bestuurders zijn over het algemeen niet begaan met kwesties als het gebruik van benchmarkrapporten. Zij streven politiek-bestuurlijke rationaliteit na (Ter Bogt, 2004a,b). Daar is het hen om te doen. Een rapport heeft voor hen pas betekenis als het hen helpt bij hun politiek-bestuurlijk functioneren en als burgers ermee gediend zijn. Een wethouder wil meerderheidsvorming in de raad bewerkstelligen rond zaken uit het coalitieakkoord. Of hij of zij wil zichzelf en de eigen partij profileren, of een ondernemer verder helpen, of wil steun verwerven, compromissen sluiten, overleven, veenbranden in de coalitie blussen, vertrouwen verkrijgen en behouden. Dat zijn allemaal zaken die te maken hebben met

politieke rationaliteit. Veel wethouders denken: *met rapporten lezen, samenvatten en uitleggen, win je geen verkiezingen*. Maar als een rapport leidt tot een pluim voor het gemeentebestuur en het betreft de eigen portefeuille dan kun je er in een persinterview wel weer wat mee. Conclusie: een benchmarkrapport *zonder* hoogst noodzakelijke beleidsvoorstellen voor nieuw of aangepast beleid is voor wethouders, raadsleden en burgers betrekkelijk betekenisloos.

Tweede gemeente

Ook een analyse van wat de benchmark bouw- en woningtoezicht betekent voor de gemeentelijke organisatie van de volgende casus, de gemeente E, laat zien dat er wel verbetermanagement plaatsvond:

‘Based on the benchmark results some new actions have been proposed and embedded’
(Noordhoek, 2013: 109).

De verbeteringen richtten zich op gebleken zwakke plekken.

Voorts blijkt uit interviews met ambtenaren dat enkele keren benchmarkinformatie in het college van b&w besproken is met als gevolg allerlei ‘gedoe’. Wethouders accepteerden de uitkomsten niet. Als dan ook nog de gevolgen van een reorganisatie in de organisatie spelen, denk aan herplaatsing van personeel, dan ligt animositeit in ambtelijke kring op de loer (Noordhoek, 2013: 135). Een benchmarkrapport kan samen met andere ‘hete’ informatie niet alleen een kwaliteitstraject agenderen, wat als positief gezien kan worden, maar juist ook leiden tot een giftige cocktail (‘gedoe’).

Conclusie: het toesturen van benchmarkrapporten aan politieke bestuurders is uitgaande van de BWT-rapportage niet vanzelfsprekend probleemloos, *ook* volgens de tweede hier genoemde casus.

Derde gemeente

Een andere case study, van de gemeente Em, leert eveneens dat het nog niet zo eenvoudig is om te leren van BWT-benchmarkrapporten. Een rapport in een ambtelijk team agenderen, is het punt niet. Dat gebeurt in deze gemeentelijke organisatie nog wel. Het rapport blijkt besproken te zijn, maar de afdeling straalt, zo blijkt, een reactieve houding uit ten opzichte van het benchmarkrapport (Noordhoek, 2013: 135). Wat houdt dat reactieve precies in?

In deze casus dringt zich het besef op dat een verandering in gang zetten tijd kost. Wie volgens een benchmark teveel indirect personeel heeft, kan dat boventallig deel niet morgen ontslaan, zeker niet als betrokkenen ook nog het gevoel hebben dat er sprake is van werkdruk en er bovendien nog genoeg werk te doen is. Dat betekent dat er in organisaties remwegen moeten zijn, korte en lange afslagroutes, en korte en lange invoegstroken. Soms ontbreekt een invoegstrook om op te rijden waar je die als bestuurder wel had willen hebben (Noordhoek, 2013: 135-136).

Over mogelijkheden tot het agenderen bestaat een bijzondere theorie. Als, in termen van de bestuurskundige John Kingdon, er geen momentum is, als er geen gelijktijdige click is tussen drie dingen - te weten een probleem, een oplossing die past op het probleem en een tijdsmoment dat urgentie uitstraalt om tot een oplossing te komen -, dan moet je als bestuurder wachten tot er wel een in politiek opzicht geschikt moment, een opening komt om iets te agenderen. Er moet zich dus een invoegstrook aandient. In termen van Kingdon (2003): er moet een ‘*policy window*’ komen, een opening.

Vierde gemeente

Tenslotte de gemeente G. Deze grote gemeente kende ‘limited learning episodes of benchmarking’, volgens Noordhoek (2013: 111). Voor deze gemeente blijkt de benchmark wel van enige betekenis maar feitelijk van beperkte betekenis. We geven daarvoor de volgende argumenten.

Ten eerste, de gemeentelijke dienst geeft de voorkeur aan eigen analyses boven de BWT-benchmark. ‘Suggestions for improvement come from their own analysis’, aldus Noordhoek (2013: 111). Is het dat alleen? Nee.

Het tweede argument houdt in dat het afdelingshoofd de interpretatie van het BWT-rapport naar zich toe heeft getrokken en 'niet heeft gedeeld' met anderen (Noordhoek, 2013: 137). Ten derde, opvallend is dat in de BWT-benchmarking een strategisch of tactisch verdedigend wapen wordt gezien tegen de burgemeester, de rest van het dagelijks bestuur (wethouders) en de gemeentesecretaris, vooral in tijden van bezuinigingen. Benchmarking helpt, volgens deze casus, blijkbaar bij het streven om budget te behouden vanuit het argument 'we willen verbeteren en dat kost geld'. De benutting van een benchmark beperkt houden, is dus een weloverwogen interne actie. Het vierde argument ter illustratie van het beperkte gebruik luidt: er is veel meer inhoudelijke benutting mogelijk. Een respondent merkt op:

'Er is weinig gedaan met de benchmark. Werk voor de teamleiders'.

Het BWT-rapport is niet gebruikt als stuurinstrument voor jaarplannen of andere afdelingsitems (Noordhoek, 2013: 138).

8.3 Enkele aanvullingen

Het overzien van de vier gemeenten leidt tot de volgende vier aanvullende inzichten voor wie een benchmarkrapport benut wil zien.

Het kiezen van een geschikt moment voor discussie: Het is zinnig te denken in termen van leerepisodes. Sommige momenten zijn beter geschikt dan andere om bezinning op gang te brengen op de uitkomsten van een benchmarkrapport. Kies bij voorkeur een moment voor publicatie en het opstarten van een discussie over uitkomsten als drie dingen samenkomen: een probleem, een oplossing (of alternatieve oplossingen) en politieke relevantie. Dus als een college en meerdere politieke fracties zich ontvankelijk tonen voor bezinning, er een gevoel is dat 'er wat moet gebeuren' en er al gedachten rondgaan over wat een oplossing kan zijn.

Voor aandacht in een ambtelijke organisatie is deze driedeling zeker niet genoeg. Dan komt het volgende in beeld.

De leerstijl van leidinggevend: Leidinggevende personen doen er bij gebruik of non-gebruik van benchmarkrapporten toe. Een positieve en actieve leergerichtheid is met name van groot belang. In de twee gemeenten Ben E legt het afdelingshoofd in de sector van het bouw- en woningtoezicht en handhaving (BWT) een *actieve (leer)stijl* aan de dag wat ook leidde tot daadwerkelijke verbeteringen in de BWT-sector (Noordhoek, 2013: 121).

Een *defensieve stijl* ten aanzien van prestatie meting en leren bevordert de totstandkoming van verbeteringen doorgaans *niet* (2013: 127). Logisch, een defensieve opstelling is niet alleen verbonden met 'afwachten' maar ook met scepsis over het bereiken van de veronderstelde voordelen van prestatie meting. Sceptici zien 'het glas eerder half leeg dan half vol'. Een defensieve opstelling kan ook ingegeven zijn door vrees harder te moeten werken en met complexiteiten te maken te krijgen. Verbeteren impliceert immers doorgaans veranderen en dat kost energie en gaat soms gepaard met spanning.

Een communicatie- en interactiestrategie: Wie 'leren' in een organisatie op gang wil brengen en wil continueren, moet nog een overweging in het oog houden. Zorg voor actieve communicatie en interactie tussen betrokkenen rond het onderwerp van benchmarking. Dit is een open deur die toch echt nog verder open gezet moet worden. Betrokkenen moeten geactiveerd worden. En ze moeten deelnemen aan een platform en de deelname aan discussies als 'welgemeend' ervaren. Noordhoek accentueert nogal het belang van bezinning in platforms en netwerken (2013: 128-130). Mensen bij elkaar brengen voor een bezinning op een rapport ligt natuurlijk voor de hand, maar soms moet het wel even georganiseerd worden. Het initiatief tot organiseren blijft soms achterwege.

Maar is het steeds zo dat de denklijn kan zijn: rapport – bezinning - verbeteren. Dat is te eenvoudig gedacht. Er is ook nog zo iets als tactisch gebruik.

Het tactisch gebruik: Ambtenaren geven aan soms *geen* positief belang te zien bij het zoeken van de communicatie met bestuurders over een benchmarkrapport. Dan vrezen ze misbruik van het rapport of incompetentie om de inhoud op waarde te schatten. Ambtenaren houden soms een rapport ook vast voor als een bezuiniging dreigt in hun eigen sector. Dan verzamelen ze materiaal om zich te verdedigen 'voor als het zo ver is dat men onze afdeling wil uitkleden'. Er speelt zich dus soms een tactisch spel af rond benchmarking.

Het komt ook voor dat niet ambtenaren maar bestuurders op hun beurt dit tactisch spel spelen. Dan willen deze bestuurders een rapport van tafel hebben omdat bepaalde bevindingen hen niet aanstaan. Ze vrezen wellicht een politiek conflict. Door het rapport intern te kritiseren en niet te erkennen, kunnen ze er wellicht onderuit komen om in het kader van de actieve informatieplicht het rapport naar de gemeenteraad te sturen. Misschien wordt de oppositie gevreesd of vreest men een enorm gezichtsverlies voor het totale lokaal bestuur.

Stel dat een rapport uitwijst dat een gemeente in een benchmark met 35 andere gemeenten als laagste uit de bus komt wat de kwaliteit van het riolenstelsel betreft. Stel, veel riolen zijn niet goed meer. De situatie is in bepaalde straten zelfs desastreus. Dan kent een rapport over dit onderwerp met deze bevindingen een enorme politieke geladenheid. En dienen zich direct vragen aan in de geest van 'wat is het college van plan te gaan doen?' Het is in die situatie mogelijk dat een bestuur tijd wil kopen voor bezinning.

9 Uitkomsten van integrale benchmarking: de beste woongemeenten

Er zijn in tegenstelling tot de opstelling van partiële benchmarks, zoals over bouw- en woningtoezicht, ook pogingen ondernomen tot bredere benchmarks. Dat gaat het om een vergelijking waarin tal van kwaliteiten van alle gemeenten, met uitzondering van de bestuurlijke kwaliteit, aan bod komen. Daarover hier meer. Diverse vergelijkingen komen aan bod, zoals de *'Atlas voor gemeenten'*, *'De beste woongemeente'*, de *'Sociale staat van de gemeente'*, de *'Staat van stadsbuurten'* en de *'Staat van de dorpen'*. Zijn dit statistische overzichten met enige analyse of commentaar of zijn het ook echte benchmarks, in de vorm van een vergelijking, rangordening, aanduiding van beste praktijken en analyse van het waarom van een voorsprong of achterstand? De *'Atlas van sociale lasten'* van COELO blijft hier om pragmatische redenen buiten beschouwing (Allers, Hoeben e.a., 2013; Janzen en Hoeben, 2013). Er zijn immers grenzen aan wat je kunt laten zien.

We benutten de discussie over de vraag of sprake is van (integrale) benchmarking tevens om te laten zien wat voor soort gegevens geproduceerd worden en wat voor tabellen mogelijk zijn en er aan uitkomsten te geven zijn. De ultieme vraag is natuurlijk wat een vergelijking en op zijn best een integrale benchmarking oplevert.

9.1 Atlas voor gemeenten

Vergelijkingen van gemeenten zijn er vele maar daarmee is nog niet vanzelfsprekend sprake van benchmarking. Immers, voor benchmarking is kenmerkend het vergelijken van organisaties op prestaties vanuit een norm of ander prestatiekader, rangordening van organisatie qua prestaties van best naar minder best, het opsporen van 'best practices' en het geven van een analyse achter de cijfers zodat begrijpelijk wordt wie waarom de beste organisatie is (op prestaties, zoals producten, diensten, processen) en hoe een 'best practice' eruit ziet en begrepen moet worden.

Wat is de 'Atlas van gemeenten' voor een vergelijking? De eerste atlas die wij onder ogen kregen is de atlas over 2003, waarin de destijds vijftig grootste gemeenten van ons land vergeleken worden op *veertig aspecten*, zoals bij voorbeeld op een aantrekkelijkheidsindex, een sociaal-economische index, vergrijzing, opleiding van de bevolking, armoede, aantal allochtonen, huizenprijzen, kantoorareaal, aantallen starters, groei van de werkgelegenheid, omvang van de werkloosheid, participatie van vrouwen, onderwijskwaliteit, nabijheid van natuur, bijstand, misdrijven, vernielingen, culinaire kwaliteit, aanbod van winkels, diversiteit van het winkelaanbod, bereikbaarheid. Deze vergelijking is heel nuttig maar *geen echte benchmarking* omdat er in dat jaar *geen* sprake is van rangordening van gemeenten, noch van een aanduiding van beste praktijken, noch van een verklarende analyse van overeenkomsten en verschillen tussen gemeenten. De atlas bevat tevens tal van indicatoren die *niet* te beschouwen zijn als prestaties van het gemeentebestuur en de eigen organisatie en niet van daarmee verbonden andere 'partners' in beleidsnetwerken.

Deze atlas is wel te beschouwen als een voorloper op andere analyses, zoals van de atlas voor gemeenten van latere jaren (zie o.a. Rigo/Atlas voor gemeenten, 2012) en de beste woongemeenten. De atlas is na 2003 overigens nog elk jaar verschenen. Steeds stond een thema centraal (www.atlasvoorgemeenten.nl). Een richtinggevende aanduiding: in 2017 was het thema geluk, in 2016 was dat de meerwaarde van water, in 2015 erfgoed, in 2014 economie & arbeidsmarkt, in 2013 groen & water, in 2012 gemeentegrootte, in 2011 cultuur en in 2010 demografische krimp. Deze atlassen zijn een knap stuk werk. Inmiddels is wel sprake van rangordeningen.

De rangorde qua grootste woonaantrekkelijkheid van 50 grote gemeenten is in 2016 als volgt:

1. Amsterdam (meest aantrekkelijk);
2. Utrecht;
3. Amstelveen;
4. Haarlem;
5. Den Bosch;

6. Den Haag;
7. Haarlemmermeer;
8. Nijmegen;
9. Leiden;
10. Leidschendam-Voorburg
11.

Onderaan de rangorde:

Plaats 48 Nissewaard;

Plaats 49 Sittard- Geleen;

Plaats 50 Emmen.

Voorbeeld: positie van Rotterdam in Atlas voor gemeenten

We hebben eens gekeken hoe de gemeente Rotterdam omgaat met de atlassen. Rotterdam is hier een voorbeeld. De 'Atlas voor gemeenten' is vanuit recentere atlassen bekeken een jaarlijks terugkerend ruimtelijk-economisch onderzoek. Twee lijstjes staan in de Atlas centraal: de woonaantrekkelijkheidsindex en de sociaal-economische index. Beide zijn samengestelde indices gebaseerd op hoe Rotterdam het doet in vergelijking met andere grote steden in Nederland op een aantal indicatoren.

Rotterdam is in de woonaantrekkelijkheidsindex gezakt van een 15^e plaats in 2011 naar een 18^e plaats in 2012 en een 17^e plaats in 2016. De woonaantrekkelijkheidsindex is samengesteld op basis van een econometrische analyse, niet op basis van de beleving van mensen. De index is opgebouwd uit meerdere factoren die een verschillende zwaarte zijn toegekend. Het zwaarst weegt de bereikbaarheid van banen. Daarop scoort Rotterdam een derde plaats. Rotterdam staat echter helemaal onderaan in de lijst als het gaat om veiligheid, een thema waaraan overigens door het bestuur na 2002 hard gewerkt is.

Door de oplopende werkloosheid is Rotterdam in de sociaal-economische index een plaats gezakt, van 39 naar 40. Hier speelt het relatief grote aantal laagopgeleiden de stad parten.

Het aantal hoogopgeleiden neemt toe in Rotterdam. In de ranglijst staat Rotterdam hiermee op de 23e plek. Economen zien het aandeel hoogopgeleiden in een stad als een belangrijke voorwaarde voor economische veerkracht en vernieuwing. Veel Rotterdams beleid is daarom gericht op het verhogen van dat aandeel.

Fig. 10: Plaats van de gemeente Rotterdam in de 'Atlas voor gemeenten 2012'

De positie van Rotterdam op de verschillende lijstjes	Rotterdam			Gemeente bovenaan	Gemeente onderaan
	2012	2011	2010		
<i>Woonantrekkelijkheidsindex**</i>	18	15	12	Amsterdam	Emmen
Bereikbaarheid van banen, Per auto en OV	3	4	3	Amstelveen	Emmen
Cultureel aanbod	19	19	14	Amsterdam	Spijkensisse
Veiligheid	50	49	49	Amstelveen	Rotterdam
Percentage koopwoningen	49	49	49	Ede	Amsterdam
Nabijheid natuurgebieden	24	24	24	Zwolle	Meerdere
Culinair aanbod	17	18	14	Maastricht	Vlaardingen
Aanwezigheid universiteit	3	3	3	Amsterdam	Meerdere
Percentage vooroorlogse woningen	7	7	7	Haarlem	Lelystad, Almere
<i>Sociaal economische index **</i>	40	39	41	Amstelveen	Heerlen
Bijstand	50	50	50	Haarlemmermeer	Rotterdam
Werkloosheid	50	49	49	Ede	Rotterdam
Arbeidsongeschiktheid	10	10	11	Breda	Heerlen
Armoede	50	50	50	Haarlemmermeer	Rotterdam

Percentage laagopgeleiden	39	39	39	Groningen	Venlo
Participatie vrouwen	42	43	36	Alphen a/d Rijn	Heerlen
Banen (per inwoner)	12	12	13	Haarlemmermeer	Spijkenisse
Percentage groeisectoren	13	13	12	Amstelveen	Velsen
Overige indices					
Creatieve klasse	22	27	23	Utrecht	Emmen
Vergrijzing	33	31	31	Leidschendam-Voorburg	Almere
Vergroening	11	11	11	Groningen	Leidschendam-Voorburg
Voetbalindex	1			Rotterdam	Meerdere gemeenten
** De woonaantrekkelijkheidsindex en de sociaaleconomische index zijn samengesteld uit de indicatoren die er onder staan					
Toelichting: Verreweg de meeste indicatoren zijn relatief, dat wil zeggen, ze zijn uitgedrukt als het aantal allochtonen, startende ondernemers enzovoorts ten opzichte van de totale bevolking of beroepsbevolking. Het (absolute) aantal allochtonen, werkloze jongeren maar ook aantal hoogopgeleiden is in Rotterdam veel groter dan in Emmen of Bergen op Zoom. Het gaat bij de Atlas om rangorde. De verschillen tussen de nummers 1 t/m 50 kunnen variëren dat wil zeggen ze kunnen heel klein of tamelijk groot zijn.					
Bron: Gemeente Rotterdam (www.rotterdam.nl/ atlas voor gemeenten 2012, 2016)					

De Atlas biedt een interessante vergelijking. We hebben ook volop waardering voor deze atlassen. Er is ook, zoals in figuur 10 te zien is, rangordening mogelijk maar omdat tal van cijfers niet betrekking hebben op bestuurlijke prestaties en geen 'best practices' worden aangegeven en het waarom niet geduid wordt, zijn we geneigd hier niet te spreken van benchmarking in strikte zin. Dus op naar de volgende vergelijking.

9.2 Atlas van beste woongemeenten

We maken nu een sprong van grote steden naar alle gemeenten, en naar een vergelijking met een rangordening maar nog steeds *zonder* 'best practices'.

Nederland telt begin 2013 408 gemeenten en 1 jan. 2017 388. Deze gemeenten zijn alle te beoordelen op kwaliteiten die een indicatie zijn voor de kwaliteit van wonen, werken en recreëren. Deze analyse is onder de vlag 'de beste woongemeente' de wereld in geslingerd. Om achter genoemde kwaliteiten te komen zijn door onderzoekers van Bureau Louter (2013) met Elsevier-journalisten 101 indicatoren opgesteld, die zijn gebundeld tot twintig onderwerpen binnen zeven thema's (Deijkers en Van Leeuwen, 2013). De thema's zijn:

- basisvoorzieningen, waarbij aan de orde is het zogenaamde jeugdpakket aan voorzieningen voor kinderen en jongeren; winkels voor dagelijkse boodschappen; het zorgpakket, zoals de nabijheid van de dichtstbijzijnde huisarts, apotheek en fysiotherapeut; sport- en fitnessvoorzieningen, waarbij indicatoren zijn ontwikkeld over de nabijheid van en verscheidenheid aan sportvoorzieningen waaronder fitnesscentra en de gemiddelde afstand tot het dichtstbijzijnde zwembad;
- plusvoorzieningen: waarbij het betreft de rubriek historisch decor, zoals de nabijheid van rijksmonumenten; de rubriek cultuur: die betrekking heeft op nabijheid en verscheidenheid van musea, bioscopen, theaters en andere ruimten voor podiumkunsten; de rubriek die verwijst naar aanwezigheid van en verscheidenheid aan horeca en luxe winkels; de rubriek zorg en hoogwaardig onderwijs, waarbij de aanwezigheid van een ziekenhuis, hogeschool en universiteit wordt meegenomen;
- arbeidsmarkt: waarbij verwezen wordt naar vitaliteit van de regionale economie en de werkgelegenheidssituatie;
- groen en blauw; dit thema kent drie rubrieken met indicatoren: allereerst recreatie, dat wil zeggen de nabijheid van en verscheidenheid van dagrecreatie en verblijfsrecreatie; water – plassen, meren, rivieren; en tenslotte groen en natuur, waarbij het gaat om positieve en negatieve landschapskenmerken en beoordeling – lees beleving - daarvan door inwoners;
- rust en ruimte: dit thema verwijst naar de lage of hoge bevolkingsdichtheid, de grootte van woningen qua aantal kamers, het gemiddeld perceeloppervlak van woningen en appartementen, en het wel of ontbreken van files;

- harmonieus leefklimaat kent als thema de rubrieken saamhorigheid van de bevolking; de rubriek die weinig overlast indiceert; de rubriek weinig misdaad; en de rubrieken verkeersveiligheid en samenstelling van de bevolking;
- het thema bereikbaarheid dat verwijst naar de aanwezigheid van openbaar vervoer (is er een treinstation?; hoeveel mensen reizen per trein?; aantal haltes voor openbaar vervoer in de buurt) en vervoer per auto.

Voor alle gegevens wordt verwezen naar Elsevier (29 juni 2013) en Bureau Louter (2013). Ze zijn door ons overgenomen, bewerkt of anders geordend. Figuur 11 laat zien wat de rubrieken zijn waarop een beoordeling per (beste) gemeente plaatsvond en welke de 'verzamelindicatoren' zijn die per rubriek zijn meegenomen. Deze verzamelindicatoren of rubrieksindicatoren zijn een samenvatting van delen van in totaal 101 criteria waarop elke gemeente beoordeeld is.

In deze figuur zijn ook bevindingen opgenomen, dat wil zeggen de scores van de twee gemeenten die op plaats een en twee zijn geëindigd van de rangorde van alle gemeenten. En enkele andere gemeenten zijn opgenomen. Bij wijze van contrast (verschil) is dat nr. 121 in de rangorde geworden (Utrecht) alsmede nr. 288 (Boekel) en nr. 407 (Pekela). De keuze had kunnen vallen op een andere stad dan Utrecht of een andere kleine gemeente dan Boekel. Maar nr. 407 is bewust opgenomen als gemeente die op de voorlaatste plaats in de rangorde staat.

Wie zit qua uitkomsten in de top van de rangorde van gemeenten en hoort daarmee tot de beste gemeenten? De Gelderse gemeente Rozendaal bezet als 'een klein Arcadië' de hoogste plaats op nummer 1 in de rangorde en de gemeente Naarden volgt op plek 2. Dat zijn dus de beste gemeenten op het gebied van wonen, werken en recreëren, kortweg de beste woongemeenten van Nederland in 2013.

Fig.11: Selectie van gemeenten uit 'De beste gemeenten van 2013' naar rangorde en afzonderlijke scores

Rubrieken	Indicatoren	Rozendaal	Naarden	Bunschoten	Utrecht	Boekel	Pekela
Inwonertal		1501	17261	20.316	321916	10062	12804
Positie op ranglijst met scores		1	2	39	121	288	407
Basisvoorzieningen	Jeugdpakket	++	++	++	++	-	-
	Winkels dagelijks voorzieningen	-	+	++	++	--	-
	Zorgpakket	0	++	++	++	--	0
	Sport en fitness	+	++	++	++	--	-
Plusvoorzieningen	Historisch decor	+	++	-	++	--	--
	Cultuur	++	++	0	++	-	--
	Horeca en luxe winkels	+	+	+	++	-	--
	Hoogw zorg en onderwijs	++	+	0	++	--	-
Arbeidsmarkt	Werkgelegenheid	0	++	++	++	0	-
	Economie	-	++	+	++	0	--
Groene en blauwe kwaliteit	Recreatiegebieden	++	+	0	-	-	-
	Water	0	++	+	+	-	-
	Groen en natuur	++	++	-	0	-	--
Rust en ruimte	Rust en ruimte	++	0	-	--	+	++
Harmonieus leefklimaat	Saamhorigheid bevolking	++	0	++	--	+	-
	Weinig overlast	++	0	+	--	+	0
	Weinig misdaad	-	0	+	--	++	--
	Verkeersveiligheid	++	+	0	--	++	0
	Samenstelling bevolking	++	++	0	--	0	--
Bereikbaarheid	Bereikbaarheid per ov en auto	+	++	0	++	--	-
Woonlasten	Woonlasten 2013 (€)	879	863	508	731	739	742
Bron: Elsevier 29 juni 2013/Bureau Louter							
Legenda: ++ = zeer goed ; + = goed; 0 = indifferent; - = matig; -- = zeer matig							

Opvallend is dat de gemeente Rozendaal, waartoe het zweefvliegveld van Terlet net niet hoort, met om en nabij 1500 inwoners tot de 'superkleine' gemeenten behoort. De gemeente ligt vlakbij Arnhem. Rozendaal wordt ook wel beschouwd als Nederlands rijkste gemeente, dat wil zeggen als de

welvarendste gemeente van het land. Wikipedia spreekt van een *'miljonairsgemeente'*, mede vanwege de WOZ-waarde van de woningen. Woningeigenaren met de hoogste hypotheekschuld van heel Nederland wonen in 2017 in het Gelderse Rozendaal. Het autobezit per 1000 inwoners is er ook tamelijk hoog. Rozendaal laat qua rijkdom gerenommeerde rijken-enclaves als Bloemendaal, Blaricum, Laren en Wassenaar achter zich. Als duurste woongemeente qua gemiddelde prijs van een koopwoning staat Blaricum met Bloemendaal, niet verrassend, ook in de top.

Rozendaal wordt vanwege het kleine aantal inwoners bestuurd, voor zover nodig, door een heel klein bestuur en ambtelijk apparaat, waardoor het nodig is om veel (delen van de) taakuitvoering in te kopen bij andere gemeentelijke organisaties zoals de gemeente Arnhem, of over te laten aan samenwerkingsverbanden en een ingehuurd adviesbureau. De ambtenaren hebben er over het algemeen brede taken. Zo is, bij voorbeeld, het hoofd burgerzaken en bestuurszaken tevens raadsgriffier.

Het college bestaat uit de burgemeester en twee wethouders; de gemeenteraad telt negen leden; niet meer en niet minder. In de raad zijn helemaal geen landelijk opererende politieke partijen vertegenwoordigd maar slechts enkele lokale lijsten. Bij verkiezingen voor de Tweede Kamer stemmen de meeste Rozendaalse kiezers overigens 'rechts' in het politieke spectrum. Kortom, een in allerlei opzichten heel bijzondere gemeente die er in slaagt om een sluitende begroting te krijgen en niet onder verscherpt toezicht staat.

Affiche met de top-tien

Tot de top-10 van beste woongemeenten volgens de Elsevier-indicatoren behoren op basis van de in de figuur genoemde indicatoren naast Rozendaal verder de gemeenten Haren (plaats 3), Blaricum (plaats 4), Heemstede (5), Laren (6), Bloemendaal (7), Hattem (8), Voorschoten (9) en Muiden (10). Uitgaande van de lijst van 101 indicatoren behoort daarmee *geen* grote stad tot de top tien. Gemeenten die vaker 'in de wandelgangen', op de golfbaan of polo, en bij 'de knallende champagnekurken' van kasteelheren en -dames positief genoemd worden en ook goed scoren op de indicatoren zijn daarnaast onder meer Wassenaar (plaats 11), Bunnik (12), Amstelveen (13), Waalre (14), Oegstgeest (15), Teylingen (16), Bussum (20), Goirle (25), Vught (30). Diverse van deze gemeenten bevinden zich in de schaduw van een grote (re) stad.

Gemeenten als Rozendaal en Bloemendaal scoren, wat niet verrast, ook hoog op het (gemiddelde) autobezit per 1000 inwoners. Het autobezit is in rangorde van veel naar minder voor een aantal gemeenten, waarvoor we dit enkele jaren geleden nagingen als volgt: Rozendaal, Bloemendaal, Nijmegen, Naarden, Goirle, Veendam. Veendam scoort dus van deze gemeenten het laagst op autobezit dat overigens verre van gering is. De vergelijking was tot deze groep gemeenten beperkt. Rozendaal, Naarden en Bloemendaal kenden voor het uitbreken van de banken- en woningmarktcrisis ook tamelijk weinig werklozen per 1000 inwoners. Deze gegevens passen in het beeld dat Rozendaal niet alleen de beste woongemeente is maar ook een rijke gemeente met betrekkelijk weinig arbeidsmarktproblematiek.

Haren

De gemeente Haren, bestaande uit Haren en vijf andere kernen waaronder Paterswolde en Noordlaren, neemt plaats 3 in. Wat is dat voor een gemeente? We gingen er kijken. Haren is een tamelijk groene gemeente, gelegen nabij Groningen. Veel inwoners pendelen naar hun werk; velen zijn werkzaam in de gemeente Groningen. Ze kunnen als ze willen makkelijk per fiets naar het werk. Het percentage hoger opgeleiden is er hoog. De gemeenteraad is betrekkelijk sterk gepolitiseerd, aldus enkele gemeentebestuurders en provinciebestuurders die het politiek bestuur van de gemeente Haren jaren gevolgd hebben.

Het bestuur van deze gemeente was recent niet alleen in moeilijkheden als gevolg van het facebook-feestje dat leidde tot rellen (zie het rapport-Cohen daarover) maar ook omdat de gemeentesecretaris in aanwezigheid van de toenmalige burgemeester (Rob Bats) en een van de wethouders aan ons verklaarde dat de ambtelijke organisatie kwantitatief en kwalitatief niet meer voldoende bij machte was om zelf de wettelijke en zelf gekozen bestuurlijke en maatschappelijke opgaven adequaat, dus

volgens de beginselen uit de *'Code voor goed openbaar bestuur'*, ten uitvoer te brengen. Dat is mede terug te voeren op de omgeving. Haren en Groningen vormen een agglomeratie – een aaneengesloten gebied – en inwoners denken in termen van voorzieningen van de stad en stedelijke ambities, een stad waarin men echter bestuurlijk niet wil opgaan. Dat ging schuren en blijft schuren. Veel inwoners van Haren willen niet bij opheffing van de gemeente Haren ineens in niet meer dan een wijkje van Groningen wonen.

Haren zou misschien, hiervan uitgaande, nog wel zelfstandig kunnen bestaan, ware het niet dat een ander punt rond 2011-2014 speelde. Tal van politici van deze gemeente toonden heel veel ambities in verhouding tot wat de ambtelijke organisatie aan kon en kan. Een bezinning op ambities en op de ambtelijke organisatie was gewenst. Die discussie liep in 2010-2017 dan ook in Haren.

Wat betekent dit voor de interpretatie van de hoge plaats van Haren in de ranglijst van beste woongemeenten? De bevindingen illustreren dat de hoge plaats van deze gemeente op de indicatoren grotendeels een gevolg is van de ruimtelijk-fysieke omstandigheden waarin de gemeente en haar inwoners verkeren (denk aan ruime woonlanen, groen, rust, historisch decor) en van maatschappelijke handelingen vanuit het bedrijfsleven en andere actoren (winkel- en zorgvoorzieningen).

Het is niet het recente handelen van het gemeentebestuur alleen of voornamelijk geweest dat de gemeente deze plaats in de rangorde bezorgde. Haren is net als andere gemeenten opgenomen in tal van beleidsnetwerken van 'partners' die van elkaar afhankelijk zijn en elkaars steun nodig hebben. Daarmee is niet gesteld dat Haren ook zonder gemeentebestuur de hoge plaats zou hebben bereikt. Maar we willen duidelijk maken dat de plaats niet volledig is terug te voeren op de kwaliteiten van het gemeentebestuur en de ambtelijke organisatie.

Voor deze vergelijking van 'beste woongemeenten' geldt daarmee gedeeltelijk hetzelfde als voor de 'Atlas voor gemeenten'; de vergelijking betreft niet voornamelijk of uitsluitend prestaties van het gemeentebestuur. Het heeft iets van vergelijking van landen. Alsof je zegt 'Nederland is het beste land om te wonen' en dan wijst op de Waddenzee, het IJsselmeer, de Afsluitdijk, de inpoldering, Naarden. Niet onjuist, maar dat we de Waddenzee als erfgoed bezitten ligt niet alleen aan het handelen van een recente regering.

De scores in 't Gooi

Nog even terug naar de ranglijst. Veel gemeenten uit 't Gooi zitten ook in de 'top 10' van beste woongemeenten, zoals Naarden, Blaricum, Laren en Muiden. De gemeentebesturen van Blaricum, Eemnes en Laren vormen samen de BEL-gemeenten. De besturen van deze drie gemeenten hebben enkele jaren terug tot een ambtelijke fusie besloten. Die fusie houdt in dat de drie ambtelijke apparaten zijn samengevoegd tot één organisatie die voor elk gemeentebestuur afzonderlijk werkt. De afzonderlijke gemeenten en gemeentebesturen bestaan nog maar de ene ambtelijke organisatie werkt voor elke gemeente apart tenzij ad hoc is afgesproken dat er een gemeenschappelijk plan wordt gemaakt voor twee van de drie of alle drie de gemeenten. Deze formule wordt wel het SETA-concept genoemd: *'Samen En Toch Apart'* (Korsten, 2002; Korsten, Becker en Van Kraaij, 2006). Is het concept in wat ook wel de BEL-formule wordt genoemd de opmaat naar beste woongemeente geweest? De woonkwaliteit van deze drie gemeenten uit 't Gooi blijkt niet of nauwelijks af te hangen van de BEL-formule want deze gemeenten scoren al jaren goed op lijsten van 'goed wonen' zoals in de figuur aangeduid, vóór de BEL-organisatie enkele jaren terug in uitvoering kwam (meer bij: Ter Braak en Von Grumbkow, 2008; Van de Laar, 2010; Herweijer en Fraanje, 2011).

Brede topkwaliteit?

Is de top van woongemeenten ook echt top, met andere woorden is een 'topgemeente' uit de 'top-10' ook in alle onderzochte opzichten goed voor op het affiche? Dat blijkt niet zo te zijn. Wie precies kijkt naar de scores van de gemeente Naarden, bij voorbeeld, ziet dat deze gemeente ten tijde van

het onderzoek goed scoort op basis- en plusvoorzieningen – denk aan aanwezigheid of nabijheid van onderwijs-, winkel- en zorgvoorzieningen -, op ‘historisch decor’, op werkgelegenheid en economie maar er *niet* bovengemiddeld uitspringt op het vlak van ‘harmonieus leefklimaat’. Wat hier gesteld is, geldt bij voorbeeld ook voor Laren (6), Bloemendaal (7) of Huizen (50).

Deze bevinding sluit aan op het beeld dat *‘de perfecte gemeente’* niet bestaat (Korsten, 2010a). Geen enkele gemeente van de lijst van meer dan vierhonderd op 101 indicatoren gescoorde gemeenten scoort ++ op alle indicatoren en in alle thematische rubrieken. Het is een onvermijdelijk cliché: elke gemeente heeft wel een eigen profiel met enige zwakten.

Typen gemeenten

Wie nader kijkt naar de kwaliteiten van alle gemeenten ziet dat er ondanks unieke kenmerken van ligging, bevolking en aan- of afwezigheid van voorzieningen uiteenlopende typen gemeenten bestaan. Enkele voorbeelden zijn te noemen: groene gemeenten, grootstedelijke gemeenten, ‘gelijkmatig scorende’ gemeenten.

Groene gemeenten als Leusden (37), Wierden (41), Dalfsen (131), Hof van Twente (75), Berkelland (171), Bronckhorst (179), Lochem (87), Raalte (162), Littenseradiel (197), Tubbergen (230) en Schermer (234) scoren over het algemeen positief op de indicatoren voor rust en ruimte, saamhorigheid van de bevolking, en hebben weinig overlast en scoren lage misdaadcijfers. Maar ze scoren veel minder positief op basis- en plusvoorzieningen en op de indicatoren voor de arbeidsmarktsituatie.

Een gemeente die ook hoog scoort en ook op ‘water’ een plus scoort, maar het juist niet van ‘rust en ruimte’ en van ‘groen’ moet hebben (twee keer een min), is het christelijke Bunschoten (plaats 39), waar 82 procent van de kiezers in 2010 bij de gemeenteraadsverkiezingen op een van drie christelijke partijen stemde. De cijfers waren vier jaar later vergelijkbaar.

Het tweekernige Bunschoten, waartoe ook de vissersplaats Spakenburg behoort, ligt nabij Huizen en Nijkerk. Bunschoten scoorde ook bijzonder goed op *‘De Staat van de Gemeente’*, de benchmark die verderop aan bod komt (Abma en Korsten, 2009).

Een gemeentetype met een tegengesteld profiel ten opzichte van de ‘groene’ gemeenten is de grootstedelijke gemeente Utrecht (plaats 121 in de rangorde). Utrecht scoort juist hoog op basis- en plusvoorzieningen, bereikbaarheid en arbeidsmarkt, maar daarentegen niet (of minder) op rust en ruimte, saamhorigheid, verkeersveiligheid, overlast en misdaad. In deze categorie van gemeenten zitten ook de gemeenten Amsterdam (93), Nijmegen (239) en Maastricht (322).

Er zijn ook gemeenten met een betrekkelijk evenwichtig ‘woongemeenteprofiel’: op de ene indicator ‘zeer goed’ en op een andere ‘goed’ of ‘indifferent’; maar op vrijwel geen enkele indicator de score ‘matig’. De Gelderse gemeente Nijkerk (34) behoort hiertoe: scoort ‘redelijk’ op basisvoorzieningen, werkgelegenheid en verkeersveiligheid maar is verder ‘niet ‘uitgesproken’; bijna geen minnen. Een vergelijkbare gemeente is de in Overijssel gelegen fraai scorende gemeente Rijssen-Holten (33), de nabij Tilburg gelegen Brabantse gemeente Oisterwijk (38) en de in de provincie Utrecht gelegen gemeente Wijk bij Duurstede (90). Veel van deze gemeenten vallen binnen het hoogste kwart van gemeenten, en behoren dus tot de ‘beste’ honderd in de rangorde.

De kelder van het klassement

Gemeenten die tot de laagste in rangorde van woongemeenten behoren, rekenen we tot ‘de kelder’, daarmee een term gebruikend uit het voetbaljargon, een term die verwijst naar voetbalclubs die onderaan de ranglijst bungelen. Gemeenten die op bijna alle indicatoren ‘matig’ of ‘zeer matig’ scoren, zijn de Zeeuwse gemeenten Terneuzen (400), Reimerswaal (405), Tholen (408) en de Groningse gemeente Pekela (407). De scores voor werkgelegenheid en economie zijn negatief in de gemeente Tholen, waarbij het opvalt dat alle gemeenten in de provincie Zeeland negatief scoren op werkgelegenheid. De Oost-Groningse gemeente Pekela scoort ook negatief op werkgelegenheid (veel

werklozen), evenals de gemeenten uit de omgeving zoals Veendam, Stadskanaal, Vlagtwedde, Oldambt, Menterwolde.

Provinciaal beeld van woongemeenten

We hebben ook uitgezocht in welke provincie hoeveel gemeenten tot de beste 100 gemeenten behoren (rangorde 1 t/m 100) en hoeveel in de rangorde de veel lagere plaats 300 of lager bezetten. Uit de figuur blijkt dat in absolute zin veel gemeenten uit Utrecht bij de beste 100 gemeenten horen uit de ranglijst 'beste gemeenten in 2013': 16 gemeenten op een totaal van 26 gemeenten in de provincie Utrecht. Ook Noord-Holland, waartoe in 't Gooi gelegen gemeenten als Laren, Blaricum en Bussum behoren, scoort goed met 24 gemeenten bij de top-100 op een totaal van 54 gemeenten die in deze provincie gelegen zijn.

Fig. 12: De beste 100 gemeenten uit het onderzoek 'beste gemeenten in 2013' en gemeenten uit de kelder van de ranglijst (plaats 300 en hoger)

Provincie	Behorend tot rangorde 0-100	In rangorde plaats 300 of meer (> dan 400)	Totaal aantal gemeenten
Drenthe	1	5	12
Flevoland	0	3	6
Friesland	3	11	27
Gelderland	16	14	56
Groningen	1	13	23
Limburg	0	13	33
Noord-Brabant	14	11	67
Noord-Holland	24	11	54
Overijssel	7	2	25
Utrecht	16	1	26
Zeeland	0	7	13
Zuid-Holland	15	16	67

Bron: Gegevens ontleend aan Elsevier, 290613

Wat is het geheim van die gemeenten in de provincies Utrecht en Noord-Holland? We gingen op zoek naar de scores op een specifieke, onderscheidende indicator en dat bleek 'werkgelegenheid' te zijn. Het blijkt dat alle 26 gemeenten uit Utrecht 'goed' (een +) of 'zeer goed' (++) op werkgelegenheid scoren. Heel veel gemeenten scoren een ++. De gemeenten uit de Randstedelijke provincies Noord- en Zuid-Holland scoren op deze indicator ook veel beter dan de meeste gemeenten gelegen in andere provincies. In Noord-Holland scoren 43 van de 54 gemeenten een + of ++ op werkgelegenheid en in Zuid-Holland 40 van de 67 gemeenten. Daarmee zetten deze Randstedelijke gemeenten een forse stap richting top-100. De Randstad is in het perspectief van geheel Nederland in feite een *banenmachine*.

Grote steden als woongemeente

Grote steden scoren niet erg goed in de rangorde van beste woongemeenten. Weliswaar beschikken bijna alle stedelijke gemeenten met meer dan 100.000 inwoners over ruime basis- en plusvoorzieningen op tal van gebieden, zoals onderwijs, horeca, recreatie en zorg, en is sprake van veel werkgelegenheid maar op harmonieus leefklimaat scoren ze doorgaans 'matig' of 'zeer matig'. Het gevolg is dat slechts enkele stedelijke gemeenten in de top-100 voorkomen en wel in de volgende volgorde: Amersfoort (plaats 52), Zwolle (58), Amsterdam (93), Apeldoorn (98) en Haarlem (99).

Fig. 13: Plaats van steden boven 100.000 inw. op de rangorde in 'beste gemeenten van 2013'

Gemeente, meer dan 100.000 inw.	Plaats in rangorde alle gem	Gemeente	Plaats in rangorde
1 Amersfoort	52	15 Breda	252
2 Zwolle	58	16 Arnhem	258
3 Amsterdam	93	17 Enschede	280
4 Apeldoorn	98	18 Dordrecht	292
5 Haarlem	99	19 Tilburg	295
6 Groningen	105	20 Eindhoven	302
7 Utrecht	121	21 Zoetermeer	304
8 Haarlemmermeer	132	22 Den Haag	308
9 Den Bosch	136	23 Zaanstad	313
10 Ede	141	24 Maastricht	322
11 Leiden	151	25 Venlo	332
12 Westland	200	26 Rotterdam	374
13 Nijmegen	239	27 Emmen	398
14 Almere	245		

Bron: Gegevens ontleend aan Elsevier, 290613

Kanttekening: maakbaarheid gering?

Van de gemeenten uit de top-50 zijn er diverse de afgelopen jaren in opspraak geweest door bestuurlijke incidenten of de noodzaak van bestuurlijke bypasses, zoals bij voorbeeld de langduriger aanstelling (dan gemiddeld) van een waarnemend burgemeester. Denk aan de hoog scorende Drentse gemeente Tynaarlo (32), een buurgemeente van het Groningse Haren, die bestuurlijke crises kende, evenals Wassenaar. De gemeente Muiden had een groot financieel probleem, wat een van de heftige discussiepunten was die een fusie met Naarden, Bussum en Weesp in de jaren nul van deze eeuw in de weg stond. Toch staan beide gemeenten, Tynaarlo en Muiden, hoog in de ranglijst van 'beste woongemeenten'. 'Beste woongemeenten' betekent kennelijk nog niet 'best bestuurde gemeenten'.

De vergelijking van gemeenten maakt een bepaald aspect van benchmarking duidelijk, namelijk dat bepaalde kwaliteiten van gemeenten een gegeven zijn en niet eenvoudig te veranderen zijn ('geen grote maakbaarheid') door inspanningen van overheidsbesturen, bedrijven en andere actoren in de samenleving. Een gemeente die geen historisch erfgoed heeft, kan niet goed scoren op de indicator 'historisch decor'; vroeger niet, nu niet, voorlopig niet. Tenzij een gemeente over gaat tot de bouw van een bijzonder museum. Een gemeente zonder universiteit kan er morgen geen starten. Er is toestemming nodig om studies te mogen beginnen en de rijksoverheid zal een doelmatigheidstoets toepassen (komen er niet teveel studies van een bepaalde soort met kleine studentenaantallen?).

Wie over veel bossen, grote woonkavels en ruiterspaden beschikt, zoals Wassenaar, scoort juist goed op de rubriek 'groen en natuur', en 'rust en ruimte'. Wie de gemeente een beetje kent, is dan ook niet verrast door de scores. Een gemeentebestuur kan ook niet eenvoudig de geboorte- en sterftcijfers beïnvloeden noch het aantal scholen. Het aantal scholen heeft te maken met de omvang van geboorten en het aantal schoolgaande (leerplichtige) jongeren.

Scores van woongemeenten tamelijk stabiel in de tijd

Hoe ontwikkelen de scores zich in de tijd als de maakbaarheid van wat kenmerkend is of wordt beperkt is? Het aantal winkelvoorzieningen wordt sterk beïnvloed door de omvang van het koopkrachtig publiek dat woonachtig is in de eigen gemeente en vanuit de nabijgelegen gemeenten komt winkelen. De implicatie hiervan is dat door de reeks criteria te hanteren, die in de figuur zichtbaar zijn, de kwaliteit van een gemeente gedurende vele jaren nagenoeg of geheel dezelfde zal blijven. De kans is daarom groot dat een vergelijking en rangordening over 2014 daardoor nagenoeg dezelfde uitkomst zal hebben als die over 2010 of 2012. Overigens kan een gemeentebestuur enige invloed uitoefenen op voorzieningen en het leefklimaat, maar ze is daarin niet de enige partij.

De vestigingsplaatsfactor

Een volgende kanttekening bij een terugblik op de benchmark van 'beste woongemeente' houdt in dat grootstedelijke gemeenten bijna altijd goed scoren op een reeks criteria op het vlak van basis- en plusvoorzieningen. Begrijpelijk omdat ze vaak vestigingsplaats zijn voor scholen, winkelvoorzieningen en zorgcentra. Deze steden scoren juist niet goed op 'harmonieus leefklimaat'-indicatoren. De gemeente Utrecht scoort dan ook niet erg goed op de indicatorruimte rust en ruimte. Begrijpelijk omdat deze stad een stedelijk en vervoersknooppunt in Nederland is en mede daardoor juist goed scoort op bereikbaarheid via openbaar vervoer.

9.3 De sociale staat van de gemeente

Is met de staat van beste woongemeenten alles gezegd over de kwaliteit van de leefsituatie in een gemeente? In de indicatoren is weliswaar aandacht besteed aan het leefklimaat maar het sociale gezicht van gemeenten komt in deze meting niet duidelijk naar voren. Het Sociaal en Cultureel Planbureau (SCP) vulde het hiaat echter op. Dit planbureau heeft een leefsituatie-index (SLI) ontwikkeld om het welzijn van Nederland(ers) in kaart te brengen. Het SCP brengt de uitkomsten onder de kop '*De sociale staat van gemeenten*'.

In de periode 2002-2012 is dit SLI-instrument ook door en in een aantal gemeenten gebruikt om het lokale sociale beleid te ondersteunen. Welke waarde heeft SLI voor het beleid van gemeenten? Kan het instrument functioneren als signaleringsinstrument voor sociale problemen in wijken, bijgroepen of op bepaalde domeinen van sociaal beleid? Is de SLI een geschikt instrument om te benchmarken? Valt de SLI aan te wenden om de effecten van sociaal beleid vast te stellen? Antwoorden op deze vragen zijn te vinden in '*De sociale staat van de gemeente*' (Boelhouwer en Gilsing, 2012).

De SLI vat met een getal de ontwikkelingen samen op verschillende maatschappelijke gebieden, zoals gezondheid, wonen, sociale participatie en sport. Het SCP splitst de landelijke cijfers uit naar de vier grote gemeenten (G4) en overige grotere gemeenten. De SLI is ook van betekenis voor een afzonderlijke gemeente vanwege de steeds grotere rol van gemeenten op het sociale domein.

Het onderzoek laat zien dat de leefsituatie in de vier grootste steden minder goed is dan in de rest van het land. De achterstand wordt evenwel steeds kleiner. Vanaf ongeveer 2000 verbeterde de situatie in de vier grootste steden met 11 procent terwijl de leefsituatie in geheel Nederland gemiddeld met 6 procent verbeterde. Er deed zich dus een inhaalslag voor in de G4 maar de plattelandsgemeenten zijn nog niet ingehaald. Volgens Boelhouwer en Gilsing (2012) springen de sterk verbeterde woonsituatie en de toegenomen sociaal-culturele vrijetijdsactiviteiten in de steden het meest in het oog.

Zoals gesteld, kunnen gemeenten zelf de leefsituatie-index, de SLI dus, gebruiken om de sociale staat van de eigen gemeente te volgen en met andere gemeenten te vergelijken. Gemeenten als Amsterdam, Lelystad, Eindhoven en Haarlemmermeer deden al eigen onderzoek met de SLI. Een positieve kant van de index is dat die in een cijfer een samenvattend integraal beeld geeft van de sociale kwaliteit van leven in de gemeente. De betekenis van de SLI als benchmarkinstrument wordt veel groter als meer gemeenten dit instrument toepassen, aldus de onderzoekers.

Vergelijking van de SLI-indicatoren per gemeente toont dat de sociale staat in Lelystad, Eindhoven en Haarlemmermeer beter is dan gemiddeld en in Amsterdam minder. Een reden is dat in Amsterdam meer burgers wonen die werkloos zijn en een laag inkomen hebben.

Tussenconclusie

Wat is nu het beeld van de beide vergelijkingen, namelijk van 'de beste woongemeenten' en 'de sociale staat van de gemeenten'? Het gaat om vergelijking en rangordening, dus om het opmaken van een klassement, maar in onze ogen niet om het aanwijzen en doordenken van een beste praktijk

en verklaren waarom een gemeente aan de top staat en wat een andere gemeente moet doen om daar te komen. In strikte zin is geen sprake van benchmarking. Van externe referentie of normering over wat een beste gemeente of sociale gemeente zou moeten zijn, is geen sprake. Het gaat op zijn hoogst om benchmarking 'light'.

9.4 De staat van stadsbuurten

Daarmee zijn we er nog niet. Gemeenten bestaan doorgaans uit ruimtelijke eenheden zoals stadsdelen, wijken, buurten en/of dorpen. Bestaat er een algemeen beeld van buurten?

Een algemeen beeld geven is niet eenvoudig omdat er juist zoveel verschillen bestaan tussen buurten. Er bestaan ook beelden bij burgers over buurten, zoals over de Transvaalbuurt en de Schilderswijk (Den Haag), de Kolenkit en de Indische buurt (Amsterdam), Ondiep (Utrecht), Klarendal en Presikhaaf (Arnhem), Nieuw-Crooswijk en Spangen (Rotterdam), Jeruzalem (Tilburg), Goffert (Nijmegen), Staatsliedenbuurt (Zaanstad) of Kerkelanden (Almelo). Tal van stedelijke bureaus voor statistiek vergelijken desondanks jaarlijks buurten, al of niet geïnspireerd door inzichten uit de stadssociologie (Nelissen, 1974). Dat leidt tot overzichten van rijkste en armste buurten, buurten met de meeste allochtonen en autochtonen, buurten met veel en weinig veiligheid en zo voorts. Toch is er wel iets meer bekend over buurten in tal van steden.

Van Eijk (1998) deed onderzoek naar alle buurten van 25 steden die vielen onder het grotestedenbeleid. In totaal ging het om 2049 stadsbuurten. Daarbij was sprake van vergelijking en rangordening. In strikte zin ging het niet om een benchmarkonderzoek want er werd niet aangegeven welke praktijk als beste praktijk navolging verdiende en waarom dat zo was. Desalniettemin was het een interessante studie. Een van de conclusies uit dat onderzoek is 'dat de slechtste buurten niet laag scoren omdat ze op alle deelaspecten het slechtste scoren maar omdat ze eigenlijk nergens goed scoren' (1998: 11). Ze zijn niet de armste buurten maar wel behoorlijk arm; ze tellen niet de meeste werklozen maar wel veel. Enzovoorts. Het zijn dus eigenlijk *probleemaccumulatiegebieden*, zoals wel is gesteld. Problemen zijn in probleemaccumulatiegebieden gestapeld.

Buurtbewoners van buurten onderaan de ranglijst van het algemeen klassement hebben daarom een troost, aldus Van Eijk: 'het kan nog slechter'. Als de slechtste buurten van Nederland nog verder achteruit zouden kunnen gaan, betekent dit dat ze in internationaal perspectief misschien geen paradijs zijn maar toch nog behoorlijke plusbuurten zijn in vergelijking met getto's, hier of daar. Het onderzoek laat zien dat 'slechte' stadsbuurten in Nederland uit het algemeen klassement in het algemeen niet de armste zijn. Het zijn niet de buurten met de meeste werklozen of bijstandstrekkers en ook niet de buurten met de meeste arbeidsongeschikten of eenoudergezinnen. In de top-twintig van ranglijsten voor dergelijke deelaspecten staan heel andere buurten, zoals – destijds - Heechterp (Leeuwarden), Cromhoffsbleek (Enschede), Hoppersgraaf (Heerlen).

Er bestaan ook buurten die positief in een toplijst van het algemeen klassement staan. In de kop van het klassement bevinden zich buurten als Ooievaarsnest (Eindhoven), Zilwijk-Zuid (Leiden) en Brinkhoek (Zwolle). Goede buurten willen ook nog wel eens een probleem hebben, zoals rondhangende jongeren die intimideren of als zodanig gevreesd worden. Veel bovengemiddelde positieve buurten bevinden zich in steden als in volgorde van best naar iets minder goed: Ede, Emmen, Almere, Zoetermeer, Apeldoorn, Zwolle, Amersfoort, maar niet in steden als de onderaan in het klassement bengelende steden Enschede, Groningen, Tilburg, Den Haag, Rotterdam en Amsterdam (Van Eijk, 1998: 13).

Deden de eerder genoemde gemeenten die in 2013 goed scoorden in het onderzoek 'De beste gemeente in 2013' eerder ook al positief mee? Inderdaad. Klingbosch in Wassenaar is een buurt die ook in 1998 al goed scoorde bij 'rijke buurten'.

Door de bureaus Rigo en Atlas voor gemeenten zijn overigens voor recentere momenten leefbaarheidsscores berekend voor wijken die inmiddels aandachtswijken heten.

Fig. 14: Aandeel bewoners van gebieden met een zeer negatieve, negatieve of matige leefbaarheidsscore in meerdere jaren van een aantal gemeenten, in rangorde per 1998

Gemeente	1998, in %	2010, in %	2012, in %	Af- of toename 1998-2012, absoluut
Amsterdam	50,1	31,7	24,7	-151.000
Rotterdam	48,9	43,7	40,9	-36.600
Groningen	34,1	6,3	2,8	-51.400
Nijmegen	31,4	10,7	7,6	-34.900
Arnhem	26,8	15,7	10,8	-20.100
Den Haag	26,4	20,7	22,8	5.100
Utrecht	24,2	12,9	10,3	-27.600
Lelystad	23,0	4,3	5,2	-9.600
Leeuwarden	20,5	6,5	7,4	-10.500
Dordrecht	19,3	3,9	3,3	-18.500
Schiedam	18,7	17,5	16,1	-1.200
Hoogezand-Sappemeer	17,0	5,7	4,0	-4.100
Delfzijl	16,9	6,4	5,1	-3.300
Almelo	16,1	6,5	2,9	-8.800
Roermond	15,5	10,1	4,4	-5.500
Helmond	14,3	3,6	1,4	-10.000
Enschede	12,8	8,1	6,2	-8.800
's-Hertogenbosch	12,2	3,9	3,2	-10.600
Wageningen	12,0	0,0	0,0	-3.700
Alkmaar	11,8	4,2	3,1	-7.700
Vlaardingen	10,9	7,8	5,8	-3.700
Heerlen	10,9	3,6	2,3	-7.800
Gouda	10,5	6,0	5,9	-3.200
Zaanstad	10,0	8,5	8,8	-200
Zwolle	4,9	2,2	1,1	-3.500

Gegevens ontleend aan: Rapport Onderscheid in leefbaarheid, Atlas voor gemeenten, 2012, bijlage 4.

De groeten uit Ulrum, Sellingen, Uithuizermeeden en Merselo

Hoe staat het dan met de woonkwaliteit of leefbaarheid van dorpen in Nederland? Als we op zoek gaan naar 'de beste woongemeente', 'de sociale staat van een gemeente' of een rangorde van stadsbuurten valt er te differentiëren: wat is op het vlak van de leefsituatie het beste dorp of gehucht in een Nederlandse gemeente? Die vergelijking van honderden dorpen is in een zuivere benchmark gericht op 'best practice' nog niet gemaakt. Er bestaat geen bij ons bekende actuele rangorde van de kwaliteiten van alle wijken of dorpen. Toch zou die lijst wel te maken moeten zijn. Er zijn overigens wel lijsten gemaakt van wijken die als *aandachtswijken* vielen te betitelen. Daarvan zijn kwaliteiten in kaart gebracht. Dat was nodig ook want het waren wijken die volgens het gemeentebestuur voor extra financiële steun in aanmerking kwamen. Later zijn de wijken die steun ontvingen ook wel *krachtwijken* of *Vogelaarwijken* genoemd.

Benchmarking van dorpen en wijken zou nuttig zijn omdat het grotestedenbeleid niet meer bestaat en het plattelandsbeleid op rijksniveau ook naar de achtergrond is verdwenen, ondanks de verstrekking door de minister van wonen van enige herstructureringsgelden aan krimpregio's (Hospers en Reverda, 2012; Steenbekkers en Vermeij, 2013). Maar ook omdat er gemeenten bestaan met veel dorpen of kernen. Denk aan de gemeente Súdwest-Fryslân die maar liefst 96 dorpen en gehuchten telt (inclusief Bolsward, Sneek en Stavoren), de Limburgse gemeente Gulpen-Wittem 57, de Friese gemeente Littenseradiel 29 en de Groningse gemeente Eemsmond 18. Een 'interne' vergelijking daarvan kan zin hebben, als hulpmiddel voor het gemeentebestuur. Meer dan zestig kernen impliceert niet dat het gaat om meer dan zestig dorpen. Wat is een dorp?

Het eerste kenmerk van een dorp dat in het oog springt, is dat van de kwantiteit aan woningen en inwoners. Een dorp is een kleine nederzetting. Die fungeert als vaste woon- en verblijfplaats van inwoners. Daarmee is een dorp onderscheiden van een doorgaans tijdelijk bewoonde camping. En een dorp wordt doorgaans ook beleefd als dorp. Indicaties voor die dorpsbeleving zijn: er staat een

kerk(gebouw), er is een café of gemeenschapsruimte in het centrum, en er is af en toe sprake van gemeenschappelijke activiteiten zoals in de vorm van een optreden van een eigen fanfare of harmonie of van een andere vrijwilligersorganisatie, zoals een sportclub. Kortom, er is sprake van een zekere gemeenschappelijkheid in een gemeenschap die verder gaat dan een gemeenschappelijk territoir.

Er zijn ook wel kleine nederzettingen die qua woningen en inwoners zo klein zijn dat niemand ze gaat betitelen als dorp. De bewoners spreken dan wel over 'bij ons' maar de kern is te klein om een fanfare of ander gezelschap op te richten of voetbalclub te vormen en in te schrijven voor een competitie. Dan gaat het om een gehucht of buurtschap.

Een dorp is dus groter dan een gehucht of buurtschap maar kleiner dan een stad en doorgaans ook kleiner dan een stadsdeel. Als een dorp beschikt over een kerk wordt ook wel gesproken van een kerkdorp. Als de kerk hoort bij een parochie (rooms-katholiek) of kerkelijke gemeente (protestants) dan kunnen een of meer gehuchten ook tot die entiteit behoren. Gehucht x hoort dan in religieuze zin bij voorbeeld tot parochie k.

De economische activiteiten die met een dorp verbonden zijn, onderscheiden zich gemiddeld genomen van die welke in steden plaatsvinden. Van oudsher vond het merendeel van de leefeenheden (gezinnen) in een dorp zijn bestaan op het gebied van land- en tuinbouw en veeteelt, inclusief bijzondere teelten zoals de fruit- en champignonteelt. Vaak betrof het dan een gemengd bedrijf. Daarnaast was sprake van vaklieden met een bestaan als zelfstandige, zoals kleermaker, architect, aannemer, timmerman, metselaar, elektricien, bakker, slager, kaasboer, kruidenier, groenteman, meubelmaker - interieurverzorger, schoenmaker, huisschilder, smid. De meesten werden 'middenstander' genoemd, ook de eigenaar van de sigarenzaak of kledingzaak. Dorpen kenden soms ook een postkantoor, een bank, of dienstverleners als een advocaat, boekhouder-financieel adviseur of notaris. In een stad kwamen de niet-agrarische activiteiten meer tot ontwikkeling, zoals voortgezet en hoger onderwijs, handel en bedrijvigheid, industrie, energie- en spoorvoorzieningen, slachterijen, luxere winkels, speciaalzaken, het veilingwezen, zorg.

Deze grove schets is een schets uit het verleden omdat ook dorpen veranderen en veranderd zijn. Boeren waren en zijn gehecht aan hun stek en dorpen kenden nogal tradities en een zekere behoudende, en hier en daar zelfs wat fatalistische instelling van de mensen (Hofstee, 1970; Mak, 1996). Maar het gaat in dorpen al lang niet meer toe als in het boek *'Dorp aan de rivier'*, het boek van Anton Coolen uit 1934 dat handelde over Lith aan de Maas. Er voltrokken zich stille revoluties, zoals de schaalvergroting in de agrarische sector, mentale schaalvergroting bij de homo ruralis, de deconfessionalisering en ontzuiling en de opkomst van nieuwe technologie, onder meer op het vlak van energie en duurzaam produceren (Van Braam, 1970; Hofstee, 1970; Munters, 1989; Veldheer, 1994; Duijvendak, 2002; Steenbekkers e.a., 2008; Hajer, 2013). Geert Mak (1996) schetst *'Hoe God verdween uit Jorwerd'* en laat en passant zien dat er nog meer veranderde, zoals bij voorbeeld hoe rond 1975-1978 de melkbussen bij de melkveehouders verdwenen en plaats maakten voor een meer moderne wijze van 'koeien melken' (Mak, 1996: 240). Tal van wat grotere dorpen kregen in de jaren tachtig of later een snackbar, bloemenzaak, wijnhandel, siersmederij, ijssalon (of ijsboerderij), een zaak voor kranten en lectuur, of een speciaalzaak in noten, kaas of fruit. En hier of daar kon je terecht voor moderne technologie. Later stierven dergelijke zaken soms weer een stille dood, bij voorbeeld als gevolg van een afname in het dagelijks lezen van kranten (en ontleding) of de attentering op het slechte van veel roken.

Dorpskroegen raken steeds meer in de problemen. Nederland telde in 1995 veertien cafés per 10.000 inwoners, in 2013 bijna tien. Het aantal dorpscafés en bars is in Limburgse dorpen van 1600 in 1995 teruggelopen naar ongeveer 1000 in 2013; een sluimerende sanering. Limburg heeft overigens nog steeds de grootste cafédichtheid. Dorpscafés verliezen echter door een gestage stroom sluitingen hun betekenis als sociale ontmoetingsplaats. Vooral de traditionele dorpscafés hebben het

moeilijk, in de steden valt het wel mee. Oorzaken: het kwijnende verenigingsleven (minder zang-, biljart- en beugelverenigingen in cafézaaltjes), het sterk gedaalde aantal kerkgangers dat aansluitend het café bezocht, het rookverbod, de alcoholcontroles en de gestegen consumptieprijsen.

Op winkeliersvlak was er ook nog andere dynamiek. Soms was de schaal te klein om van een beroep te bestaan en verdween de bakker of slager om plaats te maken voor een rijdende winkel of een supermarkt. Rijdende winkels (zoals die van de SRV) zijn na 2000 veelal weer uit het straatbeeld verdwenen. Tal van dorpen beschikken daarom nauwelijks meer over winkelvoorzieningen. Ook in economisch opzicht veranderde er wel iets in de land- en tuinbouw. Land- en tuinbouw groeiden bij voorbeeld uit tot agro & food business. Varkenshouderijen transformeerden tot bedrijven met (omstreden) megastallen (Termeer e.a., 2009). De duurzaamheidsdiscussie deed opgeld en het denken in termen van ketens en innovatie. Het dorp kwam zo gezien uit het isolement, uit het denken in termen van periferie en achterbleven gebied.

In stedenbouwkundige opzicht onderscheidt een dorp zich vaak van een stad door zijn organische en kleinschalige groei en opbouw; een stad is daarentegen vaak meer planmatig ontwikkeld en grootschalig. Verder heeft een dorp een open grens en een sterke verweving met het landelijke gebied eromheen. De stad heeft doorgaans een scherpere begrenzing (hier en daar met verdedigingswerken, omwalling, stadsmuur). Tenslotte wordt een dorp gekenmerkt door lage bebouwing in een lage dichtheid; een stad neigt door grondschaarste meer naar gestapelde bouw of hoogbouw en een hogere dichtheid.

Vanuit het perspectief van sociale samenhang kan gesteld worden dat een dorp een hechte gemeenschap heeft (men kent elkaar, meer sociale druk en controle) en een stedelijke samenleving vrijer en anoniemer is ('opgaan in de massa').

Vragen doemen op als: Liggen de beste dorpen of wijken in de gemeenten uit de top-tien of top-honderd van eerder benoemde woongemeenten in de top-vijftien, zoals de gemeente Rozendaal (1), Naarden (2), Haren (3), Blaricum (4), Heemstede (5), Laren (6), Bloemendaal (7), Hattem (8), Voorschoten (9), Muiden (10), Wassenaar (11), Bunnik (12), Amstelveen (13), Waalre (14) of Oegstgeest (15)? Opvallend is dat sommige van deze gemeenten niet uit meer dan één plaats bestaan, zoals Rozendaal, Naarden en Laren. Heb je de plaats in beeld, dan heb je direct de gehele gemeente in de kijker.

En een vraag is: liggen de slechtste dorpen of wijken in gemeenten die (betrekkelijk) onderaan de Elsevier-lijst van beste woongemeenten bungelen, zoals

- de Brabantse gemeenten Roosendaal (plaats 362), Zundert (377) of Steenbergen (385);
- de Drentse gemeenten Emmen (plaats 398) of Coevorden (372);
- de Flevolandse gemeente Dronten (379);
- de Friese gemeenten Weststellingwerf (361) of Ooststellingwerf (373);
- de Gelderse gemeenten Neerijnen (359), Zaltbommel (360), Millingen aan de Rijn (375), Rijnwaarden (384), of Tiel (397);
- de Noord-Hollandse gemeenten Medemblik (386), Den Helder (401) of Hollandse Kroon (404);
- de Overijsselse gemeente Almelo (380);
- de Groningse gemeenten Grootegast (369), Eemsmond (354), Oldambt (388), Stadskanaal (392), Vlagtwedde (399), Bellingwedde (406), de Pekela's (407);
- de Zeeuwse gemeenten Terneuzen (400), Hulst (402), Reimerswaal (405) of Tholen (408),
- de Limburgse gemeenten Kerkrade (394), of Heerlen (396); of
- de Zuid-Hollandse gemeenten Rotterdam (374), Schiedam (383), Hardinxveld-Giessendam (390) of Spijkenisse (391).

Het is niet zeker dat hier juist (bijzonder) zwakke wijken of dorpen te vinden zouden zijn want wat is 'zwak'. Er bestaan vanuit welke definitie men ook vertrekt typen wijken en dorpen, waarover verderop meer. Maar we kunnen al verklappen dat in gemeenten aan de randen van Nederland

(Zeeland, Oost-Groningen) tal van dorpen liggen die veraf liggen van stedelijke voorzieningen, zoals van haltes van intercityverbindingen. Voor de homo rusticus, de plattelander, ligt hier en daar een buitenlandse stad zelfs dichterbij. De inwoners van Vaals kunnen zelfs te voet naar de Duitse stad Aken.

9.5 De staat van dorpen

In opdracht van het ministerie van Economische Zaken heeft het Sociaal en Cultureel Planbureau in 2006 een dorpenmonitor ontwikkeld waarin de situatie van dorpen in kaart wordt gebracht. Deze studie sloot aan op het onderzoeksprogramma van het Sociaal en Cultureel Planbureau over de staat van het platteland. Dat resulteerde in de studie *'Thuis op het platteland'* (Steenbekkers, Simon en Veldheer, 2006), gevolgd door *'Nederlanders over het platteland'* (Steenbekkers e.a., 2008). Deze studies zijn aangevuld met *'De Dorpenmonitor'* (Steenbekkers en Vermeij, 2013). De recente rapportage schetst aan de hand van diverse statistische bronnen een beeld van de veranderingen in het laatste decennium in de leefsituatie van 5,3 miljoen Nederlanders die op het platteland woonachtig zijn. Het platteland omvat tweederde van de totale grondoppervlakte van Nederland. Op dat platteland woont 32 procent van de Nederlandse bevolking. De studies leverden de volgende uitkomsten op.

Nederlanders denken tamelijk positief over het platteland. Ruim 90 procent van de plattelandsbewoners is tevreden met de woonomgeving en de woning. Bijna een op de drie stedelingen zegt ooit naar het platteland te willen verhuizen (Steenbekkers e.a., 2008).

De leefsituatie op het platteland is het afgelopen decennium verbeterd maar minder dan in de steden (Steenbekkers en Vermeij, 2013). Er bestaat in 2010 overigens nauwelijks nog verschil in beleving van de leefsituatie tussen het platteland en de stad. Plattelanders en stedelingen zijn meer en meer op elkaar gaan lijken. Toegenomen welvaart, gestegen autobezit en een toename in mobiliteit, het maken van vakantie- of dienstreizen naar verre oorden, een toename in radio- en televisiebezit, alsmede de opkomst van de PC en sociale media maken dat burgers niet in territoriaal bepaald isolement zitten als in 1930 of 1955. 'De wereld' van goed en kwaad komt snel(ler) de huizen binnen. Toch is onmiskenbaar dat bepaalde maatschappelijke ontwikkelingen zich nog steeds iets anders voordoen op (delen van) het platteland dan in meer stedelijke gebieden (Hospers en Reverda, 2012). Om een voorbeeld te geven, actie- of pressiegroepen ontstond zelden of nooit op het platteland. En dorpen zijn gemiddeld minder etnisch gemixt. De multiculturele samenleving is vooral een grootstedelijk fenomeen. Maar de opkomst bij gemeenteraadsverkiezingen is in kleine dorpen evident veel hoger dan in grote delen van steden. Opkomst bij verkiezingen wordt op het platteland nog meer gerelateerd aan een burgerplicht. Daarom is studieuze aandacht nodig voor sociale achterstanden in bepaalde dorpen, wijken of regio's, die niet in rijks- of provinciaal beleid worden 'gevangen' of zelfs nog toenemen.

In de studie van het SCP uit 2013 (Steenbekkers en Vermeij, 2013) komen vier dorpstypen voor die van elkaar verschillen op twee dimensies: grootte (kleine en grote dorpen) en afstand tot een stad (bij de stad gelegen dorpen en afgelegen dorpen). De veraf van een stad gelegen kleine dorpen zijn uitgesproken landelijk van karakter terwijl de nabij een stad gelegen dorpen, zoals Haren bij Groningen, 'grote dorpen bij de stad' zijn. Er bestaan kleine veraf gelegen dorpen in gemeenten waar de bevolking tussen 2000 en 2010 daalde: zogenaamde krimpgemeenten. Die zijn te vergelijken met dorpen waar de groei stabiel was of zelfs groei optrad.

Daarnaast zijn de veraf gelegen dorpen te vergelijken met dorpen met een bijzonder fraaie ligging (o.a. Rozendaal, Haren, Hattem). Die vergelijkingen leren dat veraf gelegen dorpen zoals die te vinden zijn langs de grens tussen Nederland en Duitsland (Oost-Groningen, Drenthe, Overijssel) en de grens tussen Nederland en België het, uitzonderingen daargelaten, minder goed doen dan dorpen gelegen nabij een stad (o.a. Zwolle, Apeldoorn, Arnhem, Tilburg). Bewoners hebben dan gemiddeld een lager opleidingsniveau, kleinere banen en een lager besteedbaar inkomen, aldus de SCP-studie.

Ook zijn ze vaker werkloos. Daardoor hebben bewoners van afgelegen dorpen ook vaker te maken met armoede en sociale uitsluiting en komt er vaker een soberder leefstijl voor.

Meer in het algemeen wonen dorpsbewoners veel ruimer dan stedelingen en vaker ook in een eigen woning. Plattelanders wonen 'groener' dan stedelingen en hebben minder last van geluidhinder, criminaliteit en verloedering. Als plattelandersbewoners over 'minder' beschikken, passen ze zich aan de omstandigheden aan.

De lokale samenhang tussen burens is vooral sterk in de kleine dorpen. Burenrelaties en soms ook ruzies gedijen zowel in moderne gemeenschappen als in meer traditionele dorpen. De sociale gemeenschap in de kleine dorpen bij de stad is wel exclusiever: nieuwkomers integreren er even snel als in de afgelegen kleine dorpen maar migranten en bewoners met een lichamelijke beperking hebben er relatief weinig contact met de burens, aldus het SCP.

Voorzieningen hebben dorpelingen minder maar die vinden ze ook minder belangrijk.

Plusvoorzieningen, zoals een ziekenhuis, specialistische zorg, een hogeschool of universiteit, een theater, bijzondere restaurants of goed openbaar vervoer bevinden zich doorgaans op ruimere afstand dan stedelingen gewend zijn. Sinds 1960 sterk toegenomen autobezit en ontsluiting door weginfrastructuur, ook op het platteland, compenseert de afstand gedeeltelijk. Daardoor kan iemand in een uur rijtijd altijd wel een stad bereiken.

Dorpelingen die rond een stad wonen, zijn doorgaans (behoorlijk) mentaal verstedelijkt en verschillen daarmee gemiddeld enigszins van plattelanders uit veraf gelegen kleine dorpen, die gemiddeld genomen (nog) iets traditioneler zijn ingesteld. De kerkelijkheid is sinds 1965 in Nederland sterk afgenomen maar dat uit zich minder in 'the bible belt' (grofweg langs de lijn Tholen, Goeree-Overflakkee, Katwijk, Veenendaal, Staphorst, Zwartewaterland) en in de veraf gelegen dorpen. In deze perifeer gelegen dorpen wordt vaker een streektaal gesproken dan in de stad. In dergelijke veraf gelegen dorpen zetten dorpsbewoners zich sterk in voor de leefbaarheid en veel bewoners doen vrijwilligerswerk. Vrijwilligerswerk daalde het laatste decennium daar niet, volgens het SCP.

Kleine niet afgelegen dorpen met een mooie ligging - zoals Haren bij Groningen, Klimmen en Voerendaal bij Heerlen of Meerssen en Eijsden bij Maastricht - , waar de bewoners vaker hoger opgeleid zijn, zijn gewild onder huizenkopers. De huizenprijzen zijn er relatief hoog, de sociale participatie in vrijwilligersorganisaties op het gebied van sport en cultuur bruist, de streekproducten zijn er populair en restaurants nabij.

Bevolkingskrimp is vooral iets van na 2005 dat nog decennia aan zal houden. Het is vooral een plattelandprobleem. Het probleem speelt sterk in kleine afgelegen dorpen in krimp gemeenten (Hospers en Reverda, 2012). Deze dorpen, vooral in Noord-Nederland gelegen, hebben een wat lagere status, een hoge werkloosheid en toenemende armoede. Woningen zijn er ouder en minder goed onderhouden. Het al beperkte voorzieningenaanbod (bakker, slager, groenteman, fietsenmaker, en firma's van andere middenstanders zoals voor elektra, verf en behang, woninginrichting) verschaalde ondanks tal van leefbaarheidsstudies van na 1965 - onder andere over Borger en Borsele- en op leefbaarheid gericht provinciaal beleid verder (Groot, 1969, 1972; Jansen, 1972). De gemeenschapszin is er overigens niet minder dan in andere dorpen.

Leefbaarheid in de leegte

Hoe is de leefbaarheid in krimpgebieden? Onderzoek naar de leefbaarheid van dorpen op het platteland van Noord-Groningen op basis van ervaringen van 1600 burgers gaf in 2011 als uitkomst dat de dorpen nog springlevend zijn (CAB, 2011). Van alle burgers zegt maar liefst 93 procent tevreden te zijn met het wonen in dorpen als bij voorbeeld Leens, Kloosterburen, Usquert, Rottum of Ulrum. Waardoor dat komt? De sociale cohesie blijkt in de dorpen nog sterk, wat zich uit in buurtcontacten en deelname aan activiteiten in de omgeving. Voorzieningen als scholen, winkels en sportfaciliteiten worden wel gewaardeerd maar blijken minder relevant voor de leefbaarheid dan

een hechte en vitale gemeenschap waarin burgers onderlinge contacten hebben in soorten en maten ('weak and strong ties'). Burgers willen wel kunnen beschikken over een auto om bij voorbeeld kleine kinderen zo nodig naar school te brengen en op te halen en te shoppen. Het autobezit is inmiddels ook groot. Voor wie niet beschikt over een auto ligt het moeilijker. Dan is ineen ns ook het openbaar vervoer van belang.

Onderzoek in de provincies Overijssel, Drenthe, Noord-Holland en Zeeland bevestigt dat de leefbaarheid van dorpen beperkt afhangt van de aanwezige voorzieningen en uiteindelijk niet bepalend is (Thissen, e.a., 2003; Van der Meer e.a., 2008; Smit, 2009; Thissen, 2010; Van der Wouw, 2011). Burgers beschikken uiteraard over adaptief vermogen. Ze beseffen dat winkels een consumentenareaal nodig hebben dat een klein dorp veelal overstijgt.

Leefbaarheid heeft sterk van doen met de kwaliteit van de woonomgeving en sociale omstandigheden. Brengt het verdwijnen van voorzieningen het voortbestaan van kleine dorpen en leefbaarheid in gevaar? Dat blijkt niet zo (CAB, 2011: 73). Maar toegegeven: als tal van voorzieningen verdwenen zijn, wordt het in niet compacte dorpen met grotere kavels en boerderijen overdag vaak nog stiller. Voor agrariërs die op hun eigen of gepachte grond werken en burgers die naar hun werk elders – wellicht in een stad - pendelen, is dat minder een probleem maar burgers die daar niet tegen kunnen verhuizen. Zo ontstaat zekere selectie. Wie ermee om kan gaan blijft, en wie dichterbij voorzieningen wil zitten of stadslucht wil ruiken vertrekt. Wie de steden in een meer hectische Randstad wil ontvluchten, kan daarentegen juist in dorpen terecht, die bij voorbeeld in Noord-Groningen liggen, zoals een wethouder van de gemeente De Marne tegen ons zei. Deze wethouder stelde vast dat inderdaad mensen vanuit de Randstad migreren naar het gebied rond het Lauwersmeer.

Wat zijn de geluiden uit een provincie als Zeeland? Een dorp als Ossensisse in de gemeente Hulst telt om en nabij 400 burgers, waaronder recent nog ongeveer 50 kinderen onder 15 jaar. Ossensisse kent geen school en geen winkels. Maar het vier kilometer verderop gelegen Kloosterzande telt een basisschool, supermarkt en andere voorzieningen. Die plaats is bereikbaar per fiets maar ouders regelen onderling ook vervoer. De burgers van Ossensisse hechten welaan een eigen gemeenschapshuis. Ze hebben dan ook zelf gewerkt aan het onderhoud ervan (Out, 2011; Hospers en Reverda, 2012: 108). Dit gedrag zien we ook in andere dorpen, ook in andere delen van het land, zoals in het Limburgse Vaesrade of Merselo. Soms komt het op grotere schaal voor. In Elsendorp is door burgers zelf een alternatieve vorm van zorg ontwikkeld. Wat blijkt, is dat er onder burgers veel expertise zit. Soms meer dan in het gemeentehuis te vinden is (Cuperus en Lansink, 2013).

Deze activiteiten worden tegenwoordig doordat de Nederlandse staat bezuinigt geschoven onder labels als 'de eigen verantwoordelijkheid bij de mensen terugleggen', 'doe-democratie' (ministerie van BZK, 2013), burgerlijke zelfredzaamheid en 'van onderop-potentie' (Rob, 2012; Cuperus en Lansink, 2013). De RMO (2013) spreekt zelfs over een bredere noodzaak, ook in andere gebieden dan dorpen, van loslaten door de overheid. Hoe daar te komen? Bouwstenen voor bevordering van de doe-democratie zijn: stimuleren van sociaal ondernemerschap (dat wil zeggen: formuleer doelen, organiseer vrijwilligers, die je aan je weet te binden op een ethisch verantwoorde wijze); erkennen en inzetten van vakmanschap van burgers; sociaal vertrouwen (er komt pas wat van de grond als mensen binding aangaan en op elkaar kunnen rekenen).

9.6 De staat van het geluk

De wetenschap richt zich bij onderzoek naar geluk vooral op psychologisch welbevinden (SCP, 2012: 10). Echter, er wordt ook wel gezegd: top met streven naar geluk, ban de geluksobsessie.

Wie betekenis zoekt in plaats van geluk vindt het leven zinvoller. Dat zegt de Amerikaanse psycholoog Emily Esfahani in het boek '*De kracht van betekenis*'. Er zijn volgens deze auteur vier pijlers voor *een zinvol leven*:

- a) ergens bijhoren;
- b) bijdragen aan de wereld;
- c) een verhaal van je leven maken;
- d) het alledaagse overstijgen.

Ad a. 'Ergens bijhoren' is de eerste van de vier pijlers voor menselijke zingeving. Deze pijler voor een zinvol leven verwijst naar relaties, naar wederzijdse zorg, waarbij het er vooral om gaat om belangrijk te zijn voor de ander. Het huwelijk is hiervoor een geschikte invulling. 'Ergens bijhoren' wordt bevorderd door regelmatige interacties, of dat nou de betrekkingen in een familie betreft, of in een boekenclub of in iets anders.

Ad b. 'Bijdragen aan de wereld' is een tweede pijler. Deze is aan een context gebonden. Voor de een is het de mensenrechten bevorderen, voor de ander vrijwilligershulp verrichten in een verpleeghuis, voor een derde sneeuw ruimen bij de burens en voor een vierde koken voor broertjes en zusjes.

Ad c. Bij de 'eis' van een verhaal van je leven maken, het derde punt, gaat het om betekenisgeving aan wat je doet en ziet en daar iets mee te doen, bij voorbeeld door erover te vertellen, door samen ergens naar te gaan kijken of er een boek over te schrijven. Blijf niet stil zitten, als het kan.

Ad d. Een vierde pijler handelt over 'het alledaagse overstijgen'. Het transcendent toelaten, kan betekenen dat men bidt voor de goden, een drenkeling probeert te redden, samen ergens bij stil staat in een praatgroep van gelijkgezinden met een verbroken relatie of een ziekte, of samen iets organiseert of viert dat verheft.

Deze vier punten verwijzen naar zingeving. Zingeving is dus belangrijk. Zingeving en geluk kunnen botsen. Het najagen van geluk maakt ongelukkiger dan aandacht besteden aan deze vier, meer grijpbare aspecten van het zinvol leven, meent Esfahani. Geluk komt en gaat. Het geluk is op zijn best een *bijproduct* van het zingevende leven, aldus Esfahani. Deze auteur gaat dus niet aan het begrip geluk voorbij.

Wat bepaalt levensgeluk? Daar kan men zich semantisch in vergalopperen, zoals de bedoeling was aan te geven met de voorgaande beschouwing over zingeving en geluk. Beperkingen zijn hier gewenst. In deze analyse gaat het vooral om het 'geluksgevoel'. Op dat vlak is sprake van een oogst aan onderzoeksbevindingen, zoals blijkt uit de 'Atlas voor gemeenten', die in 2017 het thema 'geluk' centraal stelde. Wegblijven bij 'geluksgevoel' is dus teveel gevraagd.

Op de ranglijst van Europese landen van geluksgevoel staat Nederland na Denemarken op de tweede plaats (SCP, 2012: 11). De 'Atlas voor gemeenten' concludeert dat het geluksgevoel breed onder inwoners van Nederland te vinden is en tevens dat iets meer inwoners van dorpen (dorpelingen) gelukkig zijn dan stedelingen.

Het onderzoek richt zich overigens op de vijftig grootste gemeenten in ons land. Het geluk van de stedelingen wordt in Nederland vooral bepaald door *persoonlijke ervaringen*. Mensen zijn tot hun 35^{ste} levensjaar gemiddeld gelukkiger dan ouderen maar na het 65^{ste} levensjaar neemt het geluksgevoel weer wat toe. Leeftijd blijkt een differentiërende factor. Het verwachtingspatroon van het leven is aanvankelijk, tot 35 jaar, nog niet torenhoog. Burgers ervaren 'extra geluk' door hun eerste baan, een partner waarmee ze trouwen, de eerste auto. Gewenning, tegenslag en 'aanslagen' op de gezondheid leiden echter na verloop van tijd in hun leven tot een (lichte) terugval in geluksgevoel. Daar komt bij dat de normen ten aanzien van 'het geluksgevoel' inmiddels (verder) opgeschoven zijn. Bij menig burger stijgen ze. Dan ontstaat het verlangen naar loopbaanontwikkeling en/of een hoger salaris en/of een betere auto en ga zo maar door. Dat kan wat remmend werken op het geluksgevoel. Maar na het bereiken van de pensioengerechtigde leeftijd stijgt het geluksgevoel weer iets. Uiteraard tijdelijk want op zeker moment komen er mankementen.

Niet alleen leeftijd en levensfase maar ook *de leefstijl* is van invloed op het geluksgevoel. Matige drinkers voelen zich gelukkiger dan geheelonthouders. Mensen zonder gezondheidsproblemen zijn gelukkiger dan mensen met wel ernstige gezondheidsproblemen. Begrijpelijk is ook dat het hebben van familie, vrienden en kennissen een bijdrage levert aan het geluksgevoel. Aldus de *'Atlas voor gemeenten'* uit 2017 die handelt over de vijftig grootste gemeenten in Nederland.

De levensfase heeft in combinatie met andere factoren invloed op geluk, aldus de onderzoekers die rapporteren in de genoemde atlas. Welvarend, hoogopgeleid, christelijk, getrouwd, 65-plusser, zonder thuiswonende kinderen, dat blijken belangrijke elementen voor een geluksgevoel anno 2017 in de vijftig grootste gemeenten. Burgers met de genoemde kenmerken maken de grootste kans op een geluksgevoel. Wie een grotere kans maakt op niet gelukkig zijn? Mensen die van elders komen, de Nederlandse taal niet goed machtig zijn (misschien zelfs ongeletterd zijn), die geen betaald werk hebben, die vakmanschap missen waar vraag naar is, die niet bureaucratisch vaardig zijn ('de weg niet weten'), een ongezonde leefstijl erop nahouden, die gezondheidsproblemen kennen, die schulden hebben, die discriminatie op de arbeidsmarkt ervaren, qua leefstijl botsen met Nederlandse gewoonten of een riskante leefstijl erop nahouden, met criminaliteit in aanraking zijn geweest, eenzaam zijn. Geluk hangt evenwel niet alleen af van de inwoners zelf.

Overheden scheppen randvoorwaarden voor geluk maar ook corporaties, bedrijven, scholen en instellingen voor gezondheidszorg werken aan geluk. Almelo en Nieuwegein zijn de eerste gemeenten die *expliciet* een geluksbeleid voeren, aldus het SCP (2012) in *'Sturen op geluk'*.

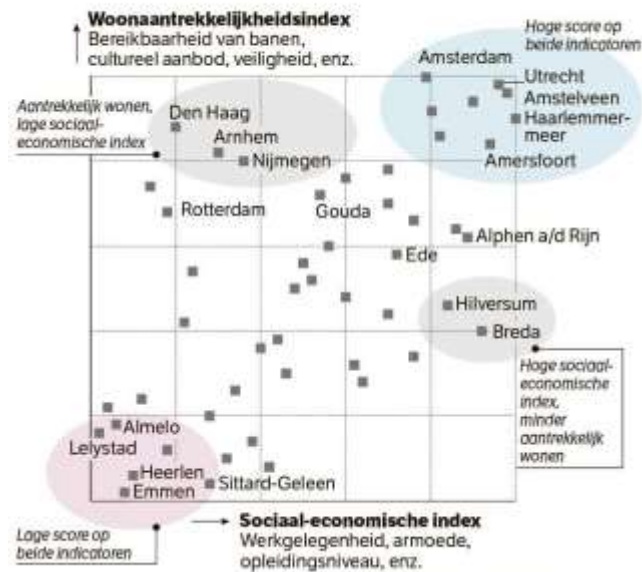
Uit de *'Atlas voor gemeenten'* van 2017 blijkt dat *de woonplaats* van belang is in relatie tot het geluksgevoel. De gemeente Ede telt relatief de meeste gelukkige mensen, volgens deze atlas. De gemeente Ede telt het hoogste percentage mensen dat zichzelf omschrijft als gelukkig, dat tevreden is met het leven en regelmatig een geluksgevoel ervaart. Bijna negen op elke tien Edenaars acht zich gelukkig.



'Gelukkige gemeenten' zijn over het algemeen gemeenten met een aantrekkelijk *woon- en leefklimaat* waar veel gezonde en werkende mensen wonen, aldus de onderzoekers. Zoals Ede. Ede wordt op de ranglijst aan de top gevolgd door de gemeenten Alphen aan den Rijn, Amstelveen, Amersfoort. Hoog scoren ook: Apeldoorn, Deventer, Roosendaal, Alkmaar, Zaandam, Den Bosch, Venlo.

Woonklimaat en sociaal-economische omstandigheden

Score van vijftig grootste gemeenten op twee indicatoren



NRC 180617 / EvG / Bron: Atlas 2017 voor gemeenten

Wat bevordert met het oog op de woongemeente een hoge score van gemeenten op geluksgevoel nog meer dan woon- en leefklimaat, en een hoog opgeleide bevolking? De hoeveelheid koopwoningen, het historisch karakter van de stad, de nabijheid van natuur, de aanwezigheid van een universiteit, het aanbod van cultuur en van goede restaurants.

Opvallend is dat de gemeente Emmen, die vaak *laag* scoort op lijstjes over 'aantrekkelijke gemeente' (zie de figuur), toch op de 16^{de} plaats van de geluksmonitor staat. Op de geluksindex staat Emmen dus in het hoogste deel van de lijst. Emmen vormt dus een uitzondering. Het is blijkbaar niet slecht wonen daar.

In de minst gelukkige gemeenten wonen relatief veel arbeidsongeschikten, werklozen, migranten, moslims en alleenstaande ouders. Onder hen zijn vaak mensen vertegenwoordigd, die van het platteland naar de stad zijn gekomen om daar het geluk te vinden en dat maar ten dele gevonden hebben. Blijkbaar speelt de kennis van rust op het platteland nog door bij het kijken naar 'de stad', als ze er wonen.



Waar verblijven de meeste gelukkige stedelingen? De kaart van Nederland komt uit de atlas en is vermeld in het dagblad Trouw (20 mei 2017). De kaart laat zien dat Rotterdam relatief lager geklasseerd staat. In Rotterdam zegt 18 procent van de bevolking niet gelukkig te zijn; maar – let wel - de meerderheid heeft nog steeds een positief geluksgevoel.

Wat werkt in steden positief in op geluksgevoel? Steden beschikken over veel voorzieningen van allerlei aard en die voorzieningen bevorderen sociale contacten. Een veilige en schone leefomgeving bevordert ook het geluk van stedelingen. Files en drukte beperken daarentegen het geluksgevoel. Ze leveren stress op bij burgers en verlies aan vrije tijd.

De woonplaats is – zoals gesteld - van invloed op de geluksperceptie. Ede scoort zo hoog in positieve zin omdat het over een aantal voordelen beschikt van het stedelijk wonen zonder veel last te hebben van de nadelen die aan steden doorgaans worden toegekend. Ede scoort goed op woon- en leefklimaat. De inwoners leven er gezond en velen hebben er ‘betaald werk’. De woonomgeving wordt gekenschetst als betrekkelijk veilig en het is er rustig wonen.

Stedelingen zijn in Nederland over het algemeen (‘gemiddeld’) iets minder gelukkig dan plattelanders maar overdrijven moeten we dat zeker niet. De percentageverschillen zijn verwaarloosbaar. Van de Nederlandse bevolking zegt 87 procent gelukkig te zijn, terwijl het vergelijkbare percentage voor stedelingen 85 is en voor plattelanders 89, aldus de atlas. Een dergelijk verschil tussen stad en platteland blijkt overigens in veel landen op te treden. Waarom? Scherp gesteld en in lijn met de onderzoekers: steden trekken een aantal ‘ongelukszoekers’ aan en verhoudingsgewijs meer alleenstaanden, werklozen en migranten. Zij vinden in steden maar ten dele hun geluk. Sommige steden kennen ook wat meer ‘verliezers van de globalisering’, zoals dat anno 2017 heet. Denk aan Nederlandse chauffeurs die hun baan verliezen tegenover concurrerende Poolse chauffeurs.

Wat kunnen gemeenten doen aan bevordering van geluksgevoel? Het gemeente bestuur van Ede zet al langere tijd in op een aangename leefomgeving en op ‘meer bewegen’. Stadbestuurders kunnen ook het cultureel klimaat bevorderen en werken aan een prettige leefomgeving. Acquisitie van

bedrijven is niet nodig want bedrijven die in een andere gemeente liggen zijn doorgaans ook bereikbaar.

9.7 Enkele conclusies

Hier passeerden de volgende vergelijkingen de revue: de beste woongemeenten, de sociale staat van gemeenten, de staat van stadsbuurten, de staat van dorpen, de gelukkigste gemeenten. Hiervan geven vergelijkingen ook een rangordening, namelijk van de beste woongemeente en de beste stadsbuurten. Ten aanzien van alle stadsbuurten en dorpen van Nederland is echter geen actuele ranglijst beschikbaar; de vergelijking van stadsbuurten is al van wat oudere datum.

In geen van de eerste vier genoemde metingen is nadrukkelijk sprake van een benoeming van een 'best practice' en een analyse daarvan. Dat wil zeggen, er bestaat wel een rangorde met beste eenheden en wat maakt dat een gemeente of buurt 'op het affiche' staat maar in geen geval wordt een handelingsstrategie aangegeven hoe een gemeente of buurt aan de top zou kunnen of moeten komen. We hebben dus hoofdzakelijk te maken met *klassementen*. Dat is wel begrijpelijk maar we moeten het toch vast stellen. Het zijn dus meer klassementen dan dat het echte benchmarks zijn. Wat een gemeente moet doen om de nationale cup van beste woongemeente te worden of beste stadsbuurt moeten lezers zelf maar proberen te bedenken. Dat neemt niet weg dat het boeiende klassementen zijn. En de onderzoekers verdienen veel waardering voor hun omvattend onderzoekswerk.

De 'Atlas van gemeenten' van 2017 geeft overigens wel een indicatie voor een gemeentebestuur wat te doen om hoger te scoren op het geluksgevoel van burgers. Daarmee is nog niet gesteld dat het eenvoudig is om dit doel te bereiken. In gemeentelijke beleidsnota's ontbreekt 'geluk' als hoofddoel doorgaans. In verkiezingsprogramma's van politieke partijen ontbreekt doorgaans het woord 'geluk' eveneens.

De rapportage over de rangordening van 408 gemeenten (per 1 jan. 2013) als 'woongemeente' beantwoordt op basis van 101 indicatoren aan het doel om inzicht te geven in wat meetelt in het klassement van 'beste woongemeenten'. Gemeenten blijken te scoren op tal van indicatoren om tot een (zoveel mogelijk) integraal beeld te komen van 'de beste woongemeente'. Scoring bleek mogelijk op basisvoorzieningen - zoals winkelvoorzieningen -, op plusvoorzieningen - zoals op het gebied van cultuur, horeca en hoogwaardige zorgverlening -, op arbeidsmarkt - in casu de werkgelegenheidssituatie en economie -, op rust en ruimte, op harmonieus leefklimaat, bereikbaarheid en woonlasten. Dat geeft een indruk van de kwaliteit van leven in een gemeente maar zegt nog weinig over het bestuurlijk klimaat of – breder - de bestuurlijke kwaliteit of over hoe beste woongemeente te worden.

De rangordening laat zien dat er gemeenten bestaan die op een aantal criteria 'goed' scoren maar geen enkele gemeente is 'perfect'.

Deze benchmark brengt onder meer aan het licht dat een gemeente een goede of matige woongemeente kan zijn ondanks tal van veranderoperaties in bestuurlijk opzicht in en van gemeenten. Een gemeentelijk proces tot het aangaan van een ambtelijke fusie met andere gemeenten verandert niet onmiddellijk de nabijheid van huisartsenposten, de kwaliteit van andere basisvoorzieningen op het gebied van onderwijs, cultuur, horeca, recreatie en zorg of de beleving van groen en natuur of de werkgelegenheidssituatie.

Denken in termen van 'best practices', bij voorbeeld suggererend dat een gemeente snel het bosareaal moet uitbreiden, is niet aangetroffen. Getuige de vele scores 'matig' in tal van vergeleken gemeenten is een harmonieus leefklimaat overigens wel een iets meer beïnvloedbaar aandachtspunt voor bestuurders en politie; hier is verbetering mogelijk.

De benchmark van beste woongemeenten kent als beperking dat enkele sociale aspecten van leefbaarheid maar beperkt aan bod komen. Daarom is een aanvulling met 'de sociale staat van de gemeente' nuttig. Aangegeven is waarom die meting nuttig is. Ook een dorpenbenchmark behoort tot de interessante mogelijkheden. Een zodanige analyse is relevant in het licht van discussies over hoe om te gaan met krimpgemeenten en hoe besturen van krimpgemeenten gevolgen van krimp kunnen beperken of opvangen.

Tenslotte, er is nog een laatste kanttekening mogelijk bij rangordes als in de atlassen te vinden is. Na de bekendmaking van de uitslag komt vooral van de kant van gemeenten die lager op de ranglijst staan naar voren dat hun inwoners helemaal niet het gevoel hebben dat in een onaantrekkelijke gemeente wonen. Dat het echt niet zo slecht wonen is, daar. Dat is wel te begrijpen. Nederland scoort internationaal op heel veel thema's steeds in de top twintig van heel vele positieve lijstjes.

10 Uitkomsten van integrale benchmarking: de bestuurlijke staat van gemeenten

Veel is daarmee al bekend over partiële, sectorgebonden benchmarking maar wat laat een integrale benchmark eigenlijk zien? Welke resultaten blijken dan? En zijn de 'beste praktijken', de 'best practices' waar zo hoog over wordt opgegeven in de literatuur over benchmarking, ook te traceren als er een rangordening van gemeenten is opgesteld? Of komt het voor dat 'best practices' niet bestaan – omdat alle gemeenten nagenoeg gelijk functioneren - of nog wel aanwijsbaar zijn, maar feitelijk onvindbaar blijken te zijn?

Tijdens de algemene ledenvergaderingen van juni en september 2008 van de Vereniging van Nederlandse Gemeenten (VNG) is het startsein gegeven voor het Kwaliteitsinstituut Nederlandse Gemeenten (KING). Hierin worden, zo is de bedoeling, benchmarking en monitoring geïntensiveerd. Verantwoording zal meer horizontaal, tussen gemeenten, gaan plaatsvinden en veel minder verticaal richting rijksoverheid en/of provinciebestuur. Benchmarking en monitoring zijn daarbij geen doelen op zich, maar middelen om te constateren waar het beter kan. In de resolutie, aangenomen tijdens de buitengewone algemene ledenvergadering van 26 september 2008, staat verder:

'Door het adequaat inzetten van deze instrumenten krijgen gemeenten goed zicht op de eigen prestaties en zijn zij in staat te beoordelen of ze over voldoende slagkracht beschikken om de ambitie van eerste overheid waar te kunnen maken. Het is aan de gemeenten zelf welke vorm zij kiezen om hun bestuurskracht te versterken: via de eigen organisatie, middels vormen van regionale samenwerking, centrum- of gastgemeenteconstructies en/of door herindeling.' KING is in 2009 definitief ingesteld. Het Kwaliteitsinstituut Nederlandse Gemeenten (KING) kreeg met ingang van 1 januari 2018 een nieuwe naam: VNG Realisatie. KING ondersteunde gemeenten bij hun informatievoorziening en dienstverlening, en was al onderdeel van de Vereniging van Nederlandse Gemeenten (VNG). Met de nieuwe naam VNG Realisatie zette de VNG verder in op ondersteuning van gemeenten op het gebied van de gezamenlijke uitvoering, naast de al bestaande dienstverlening op het gebied van belangenbehartiging, beleid en werkgeverszaken.

De Nederlandse gemeenten willen op uitvoeringsgebied meer gezamenlijk optrekken; dit is de kern van de beweging Samen Organiseren. Samen Organiseren is bedoeld als vliegwiel voor het ontwikkelen en toepassen van nieuwe generieke voorzieningen en standaarden voor gemeentelijke dienstverlening:

- Het brengt ideeën en initiatieven samen en helpt ze verder te komen
- Het maakt gebruik van wat al ontwikkeld is en past dat op grote schaal toe in de praktijk.

Het versterken van de beweging Samen Organiseren, behoort tot de kern van het takenpakket van VNG Realisatie.

10.1 De Staat van de Gemeente

We gaan even terug in de tijd. Voor wat betreft de monitoring is, onder verantwoordelijkheid van de VNG, een tijd een website in de lucht geweest: www.waarstaatjegemeente.nl. Deze website bevat(te) per gemeente een groot aantal indicatoren, afkomstig uit diverse openbare bronnen, zoals het CBS, adviesoverheid.nl, de Kamers van Koophandel etc.: de harde indicatoren. Daarnaast zijn van ruim 150 gemeenten ook zogenoemde burgerindicatoren bekend, die afkomstig zijn uit het onderzoek 'De Staat van de Gemeente' waaraan gemeenten vrijwillig konden deelnemen. Inmiddels wordt dit instrument sinds 2005 ieder halfjaar toegepast (Aardema en Korsten, 2005). Al deze indicatoren bij elkaar zouden op hoofdlijnen een beeld moeten geven over de vraag hoe gemeenten 'het' doen.

De deelnemende gemeenten namen steeds deel in groepen van om en nabij twintig gemeenten. De organisator van de benchmark zocht gemeenten die qua grootte ongeveer bij elkaar pasten. Als er dus sprake was van vergelijking dan was het een vergelijking op beperkte schaal, een rangordening

van de twintig en het zoeken naar de beste praktijken in die set van pakweg twintig gemeenten. Er was dus geen sprake van een rangorde van zeg eens 100 of meer gemeenten in één benchmark.

We analyseerden het geheel van die sets aan uitkomsten van *'De Staat van de Gemeente'* over de periode 2005 t/m 2007 en keken of het beeld voor de periode daarna stand hield. Wat voor beeld op hoofdlijnen geeft dit van de deelnemende gemeenten? Waarin zijn gemeenten volgens hun burgers goed? En waarin niet?

We beschrijven eerst kort het instrument. Daarna geven we het totaalbeeld dat uit het onderzoek rolt. Vervolgens zoomen we in op de gehanteerde indicatoren waaruit het beeld rolt waarin gemeenten volgens hun burgers goed zijn en waarin ze minder goed zijn (verder: Abma en Korsten, 2009; Abma, 2012).

10.2 Het instrument

'De Staat', zoals we dit instrument verder zullen noemen, inventariseert houdingen en opvattingen van burgers over hun gemeentebestuur op zes rollen die een burger kan vervullen: als kiezer, als klant van de gemeente, als onderdaan, als partner, als wijkbewoner en als belastingbetaler. De beoordeling wordt uitgedrukt in rapportcijfers, waarbij een 5,5 het breekpunt is. Is de score hoger dan is er sprake van een voldoende en zou een gemeentebestuur geen actie behoeven te ondernemen. Bij een lager cijfer zou dat wel het geval moeten zijn. Althans als we uitgaan van de bedoelingen van de opstellers.

De rapportcijfers per rol zijn het rekenkundig gemiddelde van ruim zestig onderliggende indicatoren. Deze indicatoren worden eveneens uitgedrukt in rapportcijfers die voor het grootste deel afkomstig zijn van oordelen van burgers uit de gemeente. Bij de rol van klant worden burgers ook ondervraagd over hun ervaringen met het gemeenteloket. Verder worden ook 'harde' indicatoren, dus zonder burgeroordelen, in cijfers uitgedrukt. Voorbeelden daarvan zijn cijfers die de opkomst bij de raadsverkiezingen betreffen, het aantal hectares met actuele bestemmingsplannen, rapportcijfers over de gemeenterekening, het weerstandsvermogen, etc. Gemiddeld gaven ruim 500 burgers per gemeente hun oordeel.

10.3 Veel gemeenten namen deel

We constateerden dat er per 1 januari 2008 152 gemeenten (35 procent van het totaal aantal Nederlands gemeenten) hebben deelgenomen aan *'De Staat'*. Omdat er gemeenten waren die twee of meer keren deelnamen kwamen we tot een totaal van 175 keren dat gemeenten *'De Staat'* lieten uitvoeren. In de figuur is de spreiding van deze gemeenten weergegeven.

Fig.15: Deelnemende gemeenten aan *'De Staat van de Gemeente'* in de periode 2005 t/m 2007



Qua deelname was er procentueel gezien weinig verschil tussen de deelname van kleine en grote gemeenten. 'De Staat' geeft in de rapporten een indeling in drie klassen, waarbij gemeenten met minder dan 25.000 inwoners worden gerekend tot de kleine gemeenten. Gemeenten met meer dan 50.000 inwoners zijn de grote gemeenten. De middelgrote gemeenten liggen daar tussenin. Van elke klasse deed ongeveer een derde van het totale aantal gemeenten mee. De grootste gemeente die deelnam was Zwolle, de kleine de Limburgse gemeenten Kessel.

10.4 Totaalbeeld

In figuur 16 hebben we de uitkomsten per rol weergegeven. Uit deze figuur blijkt dat, algemeen gesproken, de burger tevreden is over zijn gemeentebestuur en de organisatie, zij het dat de tevredenheid niet echt overloopt in enthousiasme. Op vijf rollen wordt een voldoende gescoord. Eén rol is onvoldoende: die van kiezer.

Of we nu kleine of grotere gemeenten nemen, steeds is de burger als kiezer ontevreden. Het meest tevreden zijn burgers over het algemeen in hun rol van 'partner' en hun rol van 'klant' van de gemeente.

Fig. 16: Uitkomsten van 'De Staat van de Gemeente' in de vorm van scores op een schaal van 1-10 naar burgerrollen

	Aantal gemeenten	Burger als kiezer	Burger als klant	Burger als onderdaan	Burger als partner	Burger als wijkbewoner	burger als belastingbetaler
Gemeente < 25.000 inw.	89	5,5	6,9	5,8	7,1	6,4	6,1
Gemeente 25-50.000 inw.	60	5,4	7,0	5,9	6,9	6,5	6,2
Gemeente > 50.000 inw.	26	5,3	6,9	5,7	6,6	6,5	6,1
Totaal/gem.	175	5,4	6,9	5,8	7,0	6,5	6,1

Uit figuur 16 leiden we verder af dat de schaalgrootte van de gemeente (het aantal inwoners) er niet of nauwelijks toe doet als het gaat om de rollen die de gemeenten vervullen, althans als we uitgaan van het gemiddelde.

10.5 Goede en minder goede prestaties van gemeenten

De vraagstelling is waar gemeentelijke organisaties en hun besturen goed in zijn en waarin minder goed. Daartoe moet eerst aangegeven worden wanneer we iets als 'goed' en wanneer we iets als minder goed kwalificeren.

In de systematiek van 'De Staat' doet een gemeente het goed wanneer het sein op groen staat, in wezen dus die indicatoren die op korte termijn geen speciale aandacht nodig hebben ('niet mee bemoeien'). We definiëren daarom hier die indicatoren als 'goed' als de gemiddelde score ten minste een 6,6 is én waarop geen enkele gemeente een onvoldoende scoorde. In de scoremogelijkheden bij een 5-puntsschaal is een 6,6 namelijk het midden tussen de score neutraal en tevreden, dus tussen 5,5 en 7,75.

Daartoe hebben we de scores op alle indicatoren bekeken. De rapportcijfers en hun standaardafwijkingen zijn in figuur 17 weergegeven.

Fig. 17: Rapportcijfers op indicatoren voor alle gemeenten behorende bij de zes rollen

Indicatoren	Cijfer	Standaardafwijking
<i>1. burger als kiezer</i>		
Opkomst verkiezingen	6,7	0,5
Oordeel burgers over invloed als kiezer	5,1	0,2
Oordeel burgers over vertegenwoordiging door de raad	4,9	0,2
Vertrouwen burgers in B&W	5,3	0,3
Oordeel burgers over waarmaken beloften door B&W	5,3	0,5
Mening burgers over 'goed gemeentebestuur'	5,5	0,3
<i>2. burger als klant</i>		

Berekend cijfer openingstijden gemeentehuis	3,5	0,7
Oordeel burgers over openingstijden gemeentehuis	6,5	0,5
Oordeel burgers over wachttijd voor loket	7,4	0,5
Oordeel burgers over vriendelijkheid ambtenaren aan loket	8,2	0,2
Oordeel burgers over verzorgdheid ambtenaren aan loket	8,0	0,2
Oordeel burgers over inleving in situatie door ambtenaar aan loket	7,8	0,3
Oordeel burgers over deskundigheid ambtenaar aan loket	7,9	0,2
Oordeel burgers over doorlooptijd product	7,8	0,3
Oordeel burgers over duidelijkheid verstrekte informatie via loket	8,0	0,2
Oordeel burgers over bereikbaarheid gemeentehuis	7,6	0,4
Oordeel burgers over parkeergelegenheid bij het gemeentehuis	7,0	0,6
Oordeel burgers over overzichtelijkheid van het gemeentehuis	7,5	0,3
Oordeel burgers over bewegwijzering in het gemeentehuis	7,2	0,3
Oordeel burgers over wachtruimte in het gemeentehuis	7,2	0,3
Oordeel burgers over privacy in het gemeentehuis	6,7	0,4
Oordeel burgers over prijs van geleverde dienst	5,8	0,4
Oordeel burgers over duidelijkheid van de informatie via website	6,6	0,2
Oordeel burgers over toegankelijkheid van de informatie via website	6,7	0,2
Oordeel burgers over gem. informatie via kranten	7,0	0,2
Oordeel burgers over informatie op de website	6,5	0,2
Rapportcijfer landelijke websiteranglijst	4,9	0,8
<i>3. burger als onderdaan</i>		
Rapportcijfer bouwvergunningen	8,6	1,0
Oordeel burgers over aanpak sociale veiligheid	5,5	0,2
Oordeel burgers over functioneren politie	5,4	0,3
Oordeel burgers over duidelijkheid regels	6,0	0,2
Oordeel burgers over handhaving regels	5,3	0,2
Oordeel burgers over ontbreken regels	5,0	0,1
Oordeel burgers over tegenstrijdige regels	5,2	0,2
<i>4. burger als partner</i>		
Rapportcijfer actualiteit bestemmingsplannen	6,5	2,3
Oordeel burgers over basisscholen	7,3	0,2
Oordeel burgers over het onderwijs in het algemeen	7,1	0,2
<i>5. burger als wijkbewoner</i>		
Oordeel burgers over eigen woning	8,4	0,1
Oordeel burgers over directe woonomgeving	7,5	0,2
Oordeel burgers over speelmogelijkheden	6,0	0,3
Oordeel burgers over verlichting	7,0	0,2
Oordeel burgers over wegen en straatmeubilair	5,9	0,3
Oordeel burgers over winkels in de buurt	7,0	0,5
Oordeel burgers over winkels in de gemeente	6,8	0,6
Oordeel burgers over uitgaansmogelijkheden	6,1	0,5
Oordeel burgers over culturele voorzieningen	5,9	0,5
Oordeel burgers over aanbod verenigingen en clubs	7,0	0,2
Oordeel burgers over sportvoorzieningen	7,1	0,2
Oordeel burgers over zorgvoorzieningen	6,8	0,3
Oordeel burgers over welzijnsvoorzieningen	5,6	0,3
Oordeel burgers over openbaar groen	6,6	0,3
Oordeel burgers over afvalinzameling	7,1	0,3
Oordeel burgers over onderhoud van de wijk	6,3	0,3
Oordeel burgers over bereikbaarheid van buurt met auto	7,4	0,2
Oordeel burgers over parkeren in de buurt	5,7	0,2
Oordeel burgers over openbaar vervoer in wijk	5,9	0,5
Oordeel burgers over parkeren in gemeente	5,9	0,5
Oordeel burgers over openbaar vervoer in gemeente	5,8	0,5
Oordeel burgers over fietsvriendelijkheid	6,7	0,3
Oordeel burgers over verkeersveiligheid in gemeente	5,9	0,2

Oordeel burgers over verkeersveiligheid in buurt	5,7	0,2
<i>6. burger als belastingbetaler</i>		
Stelling: Ik krijg voldoende terug voor mijn belastinggeld	5,0	0,3
Stelling: De gemeente is mijn belastinggeld waard	5,8	0,2
Algemene middelen versus rekeningresultaat	6,3	2,5
Weerstandsvermogen versus omzet	7,8	1,4
Oordeel burgers over gemeentelijke organisatie	6,1	0,2

(de cijfers van de indicatoren zonder 'oordeel over' of 'stelling' zijn niet afkomstig via meningen van burgers, maar zijn omgerekende harde cijfers)

Indicatoren die niet als 'goed' worden gekwalificeerd, behoeven de aandacht van de gemeenten. Voor de analyse maken we echter wel een splitsing tussen die indicatoren die we als neutraal kenschetsen en die indicatoren waarin gemeenten het in de breedte niet goed doen. Indicatoren met een gemiddelde score vanaf 5,5 tot en met 6,5 noemen we 'neutraal'. Daar voegen we ook de indicatoren bij die gemiddeld wel een voldoende score, maar waarbij we ook onvoldoendes van verschillende gemeenten aantreffen.

Alle overige scores, die met een gemiddelde lager dan een 5,5 zijn indicatoren waarover in de breedte burgers ontevreden zijn.

10.6 Indicatoren waarop gemeenten 'goed' scoren

Bij de rollen van kiezer, onderdaan en belastingbetaler werden *geen* indicatoren met scores aangetroffen die als 'goed' konden worden gekwalificeerd.

In de rol van klant was de burger goed te spreken over de deskundigheid, verzorgdheid, inleving, deskundigheid van de medewerker(s), evenals over de doorlooptijd en duidelijkheid van de informatie. Verder waren de bezoekers van het gemeentehuis content met de overzichtelijkheid, bewegwijzering en wachtruimte in het gemeentehuis. De burgers waren in het algemeen ook tevreden over de toegankelijkheid van informatie, de gemeentelijke informatie via lokale kranten en de informatie via de website.

De burgers waren ook tevreden over de basisscholen en over het onderwijs in het algemeen. Deze indicatoren behoren bij de rol van partner.

In zijn rol van wijkbewoner is de burger tevreden over zijn woning en de directe woonomgeving de verlichting, verenigingen, clubs, sport- en zorgvoorzieningen, afvalinzameling en de bereikbaarheid van de buurt met de auto.

10.7 Neutrale indicatoren: niet goed en niet slecht

Bij de rol van kiezer worden de opkomst bij verkiezingen en de mening van burgers over goed gemeentebestuur neutraal door de burgers beschouwd.

Bezoekers van het gemeentehuis zijn in hun klantrol neutraal over de openingstijden, de wachttijd voor het loket, over de privacy bij het loket en over de prijs die voor de gemeentelijke producten en diensten betaald moet worden, evenals over de bereikbaarheid van en de parkeermogelijkheden bij het gemeentehuis.

De verhouding tussen gegronde en ongegronde bezwaar- en beroepschriften, het telefonisch meldpunt, de aanpak van de sociale veiligheid en de duidelijkheid van de gemeentelijke regels worden als niet goed en niet slecht gewaardeerd, evenals het percentage actuele bestemmingsplannen in de partnerrol.

In de rol van wijkbewoner zijn de burgers in het algemeen tevreden noch ontevreden over de speelmogelijkheden voor de kinderen, over de wegen en het straatmeubilair in de buurt, de winkels, de uitgaansmogelijkheden, de cultuur en de welzijnsvoorzieningen in de gemeente, over het openbaar groen, het onderhoud van de wijk of buurt, het parkeren en het openbaar vervoer in de

wijk en in de gemeente, de verkeersveiligheid in de buurt en de gemeente en over de fietsvriendelijkheid in de gemeente.

De neutrale indicatoren in de rol van belastingbetaler zijn: de mening van de burgers over de vraag of de gemeente hun belastinggeld waard is, de gemeentelijke organisatie, het weerstandvermogen versus de omzet en de algemene middelen versus het rekeningresultaat.

10.8 Waar burgers ontevreden over zijn

Burgers zijn het meest ontevreden in hun kiezersrol: de invloed van de kiezer, de vertegenwoordiging van de burger in zijn gemeenteraad, het waarmaken van beloften door B&W en het vertrouwen in B&W worden als onvoldoende gekwalificeerd. Voor een lijst met een rangorde van scores op basis van opvattingen van burgers van vóór 2010 over 'goed bestuur' verwijzen we naar bijlage 3.

In de rol van onderdaan zijn burgers niet tevreden over het functioneren van de politie, het ontbreken van regels, de tegenstrijdigheid en de handhaving van die regels.

In de rol van belastingbetaler wordt de mening dat de burger voldoende terugkrijgt voor zijn belastinggeld niet door diezelfde burgers onderschreven.

Bij de rollen van partner en wijkbewoner werden geen onvoldoendes gescoord.

10.9 De ranglijst van gemeenten

Volgens de beginselen van benchmarking valt er ook een ranglijst te maken van de totaalscores van deelnemende organisaties. De gemeente B (we hebben de namen geanonimiseerd) scoorde het hoogste op 'De Staat van de Gemeente' met een totaalscore van 41,6 een optelsom van de zes afzonderlijke scores op rollen. Dat is minder dan een score van gemiddeld 7. Uit deze lijst valt nog meer te concluderen.

Geringe verschillen in totaalscores

Ten eerste dat de totaalscores van de beste gemeenten dicht bij elkaar liggen. De verschillen in totaalscores tussen de eerste elf gemeenten zijn niet zo groot. Dat betekent dat de gemeente die zich wil verbeteren, over zes rollen en zes scores gezien, zich niet speciaal bij gemeente B zou hoeven te melden om te achterhalen wat het geheim van deze smid is maar evenzeer het licht kan opsteken bij nummer 3, 6 of 7. De scores zijn over het algemeen ook niet zo hoog dat we één gemeente kunnen kwalificeren als een excellente gemeente. De excellente gemeente bestaat niet, volgens deze benchmarking (zie ook Korsten, 2010a).

Wat eerste op het totaal wil zeggen

Ten tweede valt te kijken naar de afzonderlijke scores die er qua hoogte uitknallen. Dan valt op dat de nummer 1 van de lijst op de partnerrol een 8,3 scoort, evenals gemeente 4 maar gemeente 120 ook nog een 8,0 scoort. Het gemeentebestuur van nr. 120 kan dus zijn trots hebben. Blijkbaar kan een gemeente die het redelijk doet op een bepaalde rol op andere rollen duidelijk zwakker zijn. 'Een keer goed, altijd goed' geldt dus hier niet. Eerste op een rangorde wil nog niet zeggen dat de 'best practice' niet elders kan liggen.

Beste praktijk soms niet goed te achterhalen of bestaat helemaal niet

De eerste vijftien gemeenten op de ranglijst scoren bij vijf van de zes rollen allemaal minder dan een 8. Het blijkt bij nader inzien voor een gemeentebestuur met een lerende houding en een daarbij passend zoekgedrag nog niet zo eenvoudig om de 'best practice' per rol of cluster te achterhalen. Bij welke gemeente moet een ambtenaar of wethouder te rade gaan als hij of zij zoekt naar de gemeenten die qua klantgerichtheid eruit springen? Bij gemeente N die de tweede plaats bezet op de algemene lijst maar qua klantgerichtheid de beste score heeft van de eerste vijftien gemeenten, met een cijfer 7,4 op een scoremogelijkheid van 1-10. Zou daar veel te leren vallen, in vergelijking met gemeenten die een 7,2 scoren?

Van alle 151 gemeenten scoort geen enkele gemeente op de rol van wijkbewoner een 7,0 of meer. Bijna alle cijfers van alle gemeenten liggen tussen 6,0 en 6,9 (Abma en Korsten, 2009: 190-192). Op dit vlak bestaat een beste praktijk niet eens maar ook geen 'bad practice'.

De kiezers zien geen beste praktijk

Ten vierde blijkt dat heel veel gemeenten die aan de benchmark deelnamen op de rol van kiezer matig of slecht scoren. Van 151 gemeenten scoren slechts zeven gemeenten een 6,0 of iets hoger. Op het eerste gezicht lijkt gemeente H, die op de zevende plaats van de algemene scores prijkt met de score 6,4 een praktijk te hebben die weliswaar beter is dan die van andere gemeenten maar om 6,4 als 'best practice' te kwalificeren gaat erg ver.

Voor scores op het item 'goed bestuur' wordt verwezen naar bijlage 3. Er is volgens die scores van burgers uit deelnemende gemeenten reden voor bijna alle gemeenten die deelnamen aan de benchmark 'De Staat van de Gemeente' om harder te werken aan de uitvoering van de 'Code voor goed openbaar bestuur'.

Probleemgemeenten scoren matig of slecht

Ten vijfde valt op dat gemeenten die gedurende langere tijd vanaf 1995 het stempel 'bestuurlijke probleemgemeente' kenden (Schoemaker, 2011) een slechte score laten zien in de benchmark op de rol van kiezer. Kiezers hebben niet veel vertrouwen in een dagelijks bestuur als de coalitie vaak valt of een individuele bestuurder gedwongen wordt tot vertrek omdat hij of zij niet langer het vertrouwen geniet van de gemeenteraad. Op deze rol zijn volgens de scores veel 'bad practices' waarschijnlijk.

Het merkwaardige is dat van Vlaamse steden bekend is dat ze op dit vlak ook lang niet altijd voldoende of tamelijk voldoende functioneren maar lage scores krijgen van kiezers. We konden in een visitatie van Vlaamse steden, in opdracht van de Vlaamse overheid, vaststellen dat dagelijks bestuurders en individuele fractieleiden hiermee niet goed raad wisten. Ze wisten niet wat te doen en ontvouwden dan ook geen actieplan over hoe het vertrouwen van veel kiezers te herwinnen (o.a. Korsten, 2012).

Groot niet beter

Het zal de lezers overigens niet ontgaan dat de grotere gemeenten geen aparte positie innemen (ook Boogers, 2013). Grotere gemeenten scoren niet zonder meer goed bij burgers vanuit welke burgerrol dan ook bezien. Gemeenten met meer dan 30.000 inwoners komen niet voor bij de eerste vijftien gemeenten van de ranglijst. Voor 'best practices' hoef je, kijkend naar gemeentegrootte, niet en zeker niet

Fig. 18: Ranglijst van gemeenten naar totaalscores op 'De Staat van de Gemeente'

nr.	Gemeente	aantal inw.	Kiezer	Klant	onderdaan	Partner	wijkbewoner	belastingbetaler	Totaal
1	Gemeente B	19475	6,0	7,2	6,2	8,3	6,6	7,3	41,6
2	Gemeente N	16441	5,5	7,4	6,3	7,9	6,6	7,0	40,7
3	Gemeente O	24019	5,7	7,1	6,2	7,8	6,8	7,0	40,5
4	Gemeente G	14456	5,5	7,0	5,9	8,3	6,3	7,4	40,4
5	Gemeente A	12221	5,8	7,0	6,1	8,0	6,4	7,1	40,3
6	Gemeente T	7164	6,2	7,1	5,9	8,3	6,6	6,3	40,3
7	Gemeente H	11630	6,4	6,8	5,9	7,6	6,8	6,9	40,3
8	Gemeente D	25605	5,6	7,3	6,2	7,1	6,7	7,2	40,2
9	Gemeente D	25586	5,6	7,1	5,9	7,3	6,7	7,4	40,1
10	Gemeente V	21947	5,9	7,0	6,2	7,6	6,6	6,8	40,0
11	Gemeente P	23173	5,5	7,1	6,0	7,7	6,9	6,8	40,0
12	Gemeente M	7660	5,8	6,9	5,8	7,7	6,5	7,1	39,8
13	Gemeente N	26583	6,0	7,0	5,9	7,8	6,7	6,4	39,7
14	Gemeente D	18114	5,6	6,9	6,2	7,4	6,5	7,1	39,7

15	Gemeente H	19107	5,6	7,1	6,4	7,3	6,8	6,4	39,6
20	Gemeente K	28783	5,8	6,9	5,8	7,6	6,8	6,5	39,4
30	Gemeente L	10850	5,5	6,8	6,0	7,5	6,4	6,8	39,0
40	Gemeente C	36040	5,4	7,0	6,0	7,2	6,5	6,7	38,8
50	Gemeente H	42732	5,5	6,8	6,3	7,4	6,6	5,8	38,5
60	Gemeente D	34353	5,2	7,0	6,0	7,7	6,3	6,0	38,3
70	Gemeente B	9514	5,6	6,9	5,5	7,3	6,3	6,3	37,9
80	Gemeente O	21887	5,3	7,1	6,1	6,5	6,4	6,3	37,7
90	Gemeente U	9401	5,6	7,0	5,3	6,9	6,4	6,1	37,3
100	Gemeente M	33282	5,5	6,7	6,1	5,5	6,7	6,7	37,2
110	Gemeente Z	20797	5,0	6,8	5,9	6,3	6,5	6,2	36,8
120	Gemeente Zw	21872	5,8	6,7	5,7	8,0	6,2	4,3	36,6
130	Gemeente Gr	9685	5,2	7,3	5,2	7,3	6,1	5,3	36,3
140	Gemeente Be	66325	4,6	7,0	5,3	7,4	6,4	5,0	35,7
150	Gemeente Ha	17799	3,8	6,8	5,5	7,2	6,2	4,6	34,1

Bron: Korsten en Abma, 2009. Van gemeenten die meerdere malen deelnamen zijn de rapportcijfers van het laatste jaar van deelname weergegeven.

vanzelfsprekend bij de gemeenten boven 50.000 inwoners te zijn. De slechtste positie op de lijst van 151 gemeenten wordt ingenomen door een gemeente van 75.000 inwoners die op vier van de zes rollen een onvoldoende scoort. Deze gemeente bezet plaats 151.

Ook een door Marcel Boogers uitgevoerd onderzoek op basis van data die ontleend zijn aan Waatstaattjgemeente.nl ('De Staat van de Gemeente') toont dat er 'geen duidelijk lineair verband is tussen de omvang van gemeenten en de tevredenheid van inwoners met de verschillende aspecten van het bestuur ervan'. Als het gaat om de responsiviteit van het gemeentebestuur is er zelfs helemaal geen verband. De tevredenheid met de buurtgerichtheid van het bestuur (de mate waarin aandacht is voor leefbaarheid en veiligheid van de woonomgeving) is in de grootste én kleinste gemeenten wat lager. 'Het oordeel over de kwaliteit van de leefomgeving is in grotere gemeenten iets minder positief, terwijl het omgekeerde geldt voor de tevredenheid met het voorzieningenniveau' (Boogers, 2012: 14; 2013).

'De algemene kwaliteit van het gemeentebestuur wordt het best beoordeeld in middelgrote gemeenten: in de kleinste én de grootste gemeenten is men hier gemiddeld genomen minder tevreden over. Dat wil zeggen dat de democratische legitimiteit van het gemeentebestuur hier het meest onder druk staat' (Boogers, 2012).

Fig. 19: Kwaliteit van gemeenten in rapportcijfers, naar gemeentegrootteklassen

Aspecten	< 20.000 inw.	20-50.000 inw.	50-100.000 inw.	>100.000 inw.
Responsiviteit	6.24	6.26	6.24	6.24
Buurtgerichtheid	6.04	6.12	6.17	6.05
Leefomgeving	6.83	6.88	6.75	6.67
Voorzieningen	6.84	7.02	7.14	7.17
Algemene kwaliteit	6.05	6.11	6.18	6.06

Bron: Boogers, 2012, tabel 9.

Bevordering van bestuurlijk vertrouwen door benchmarking?

De opeenvolging van deze benchmarks in zelfde gemeenten heeft het vertrouwen van burgers in het gemeentebestuur of de afzonderlijke organen, zoals het college van burgemeester en wethouders en de gemeenteraad, niet bevorderd. Dat blijkt uit een vergelijking van gemeenten die meerdere malen deelnamen aan 'De Staat van de Gemeente' (zie ook Abma en Korsten, 2009; Abma, 2012). De scores tenderen in het algemeen niet naar een positieve correctie en ook niet op het aspect 'vertrouwen in het college' en 'vertrouwen in de gemeenteraad'. Dat de benchmarkscores in bepaalde probleemgemeenten duidelijk lager zijn en blijven, is een weerspiegeling van de praktijk. Die ontgaat burgers blijkbaar niet.

De veronderstelling dat de benchmark *'De Staat van De Gemeente'* op zich al bevorderend zou kunnen werken op het herstel van het over het algemeen in de gemeenten lage vertrouwen in bestuurlijke organen en hun handelen, is overigens betrekkelijk naïef. Vertrouwen is niet louter te herstellen via terugkoppeling van uitkomsten van benchmarkonderzoek naar burgers en een discussie daarover. Vertrouwensverlies en vertrouwensherstel blijkt van tal van factoren, gebeurtenissen en andere omstandigheden afhankelijk, zoals uit *'Bouwen aan vertrouwen in het openbaar bestuur'* blijkt (Korsten en De Goede, 2006). Er lijkt nog steeds wat te zeggen voor de stelling dat 'vertrouwen te voet komt en te paard gaat', wat zoveel wil zeggen als dat vertrouwen geleidelijk kan groeien maar dat het vertrouwen in het lokaal bestuur en landsbestuur ook zo verdwijnt.

Zijn er gemeenten geweest die zich zijn gaan bezinnen op de score op de kiezersrolen daarmee op de vertrouwenspositie? Voor zover we konden waarnemen zijn er amper gemeentebesturen gevonden die in de tussentijd, tussen enkele deelnames aan de benchmark, een expliciete herstelpolitiek op het vlak van bestuurlijk vertrouwen voerden (Abma, 2012). Vertrouwen was bij voorbeeld wel een aandachtspunt van de besturen van de Drechtsteden.

Relatie met eerdere onderzoeken

De uitkomsten zijn in de lijn met de uitkomsten van andere onderzoeken. Zo gaven burgers bij een onderzoek van de SGBO en de VNG uit 2005 gemeenten overall een 6 of een 7. De onderzoekers noemden de gemeenten een *'onopvallende leerling'* waarvan het de burgers niet scherp voor de geest stond wat nu precies de prestaties waren. Net als in het onderzoek *'De Staat van de Gemeente'* waren burgers vrij positief over hele concrete, veelal uitvoerende activiteiten van de gemeente. Dit gold vooral voor de beoordeling van de 'service' en het vertrouwen dat burgers hebben in de inzet door de gemeenten bij het verbeteren van de veiligheid. Hoe concreter het onderwerp, hoe groter het vertrouwen.

In een onderzoek van de Universiteit van Tilburg uit 2005 kwam naar voren dat burgers over het geheel genomen betrekkelijk negatief waren over het functioneren van de lokale democratie: a) burgers voelden zich te weinig serieus genomen; b) waren ontevreden over hun invloed op het bestuur van hun gemeente; c) voelden zich weinig betrokken bij de lokale politiek; en d) vonden dat deze weinig herkenbaar is (Boogers e.a., 2005).

Groot verschil met de eerdere en meer algemene onderzoeken is dat bij *'De Staat van de Gemeente'* per gemeente de burgertevredenheid werd gemeten, zodat gemeenten elkaar onderling kunnen vergelijken. Het algemene beeld wordt dus bevestigd, maar er zijn onderling wel verschillen tussen gemeenten. Zo waren er gemeenten die wel degelijk een voldoende scoorden op hun rol als kiezer, maar ook in de financiële indicatoren werden grote verschillen aangetroffen. Maar over het algemeen is er sprake van een redelijk constant beeld van de tevredenheid van burgers.

10.10 Wie oriënteert zich op wie

Naar aanleiding van de integrale benchmarking is nog een vraag op te werpen over rangordening als methode. We zagen immers dat soms bij het vergelijken van 150 gemeenten al nauwelijks een 'best practice' te vinden was. Is vergelijking van groepjes gemeenten – steeds een twintigtal - in een benchmark, zoals bij voorbeeld bij *'De Staat van de Gemeente'*, zinvol of niet zinvol? Bestaat niet het risico dat een aantal kleine gemeenten op elkaar lijkt en allemaal op zelfde punten redelijk aan de maat is of matig scoort? We geven enige pros en cons.

Een vergelijking van een kleine groep gemeenten is verdedigbaar omdat dan nog te overzien wat van welke diensten, producten of processen vergeleken wordt. Dan is nog te achterhalen wat er schuil gaat achter de cijfers. Dus bij voorbeeld waarom er verschillen in ratio's zijn of bij voorbeeld waarom een gemeente 750 of 37 euro vraagt voor een bouwvergunning voor het aanbrengen van een

dakkapel op het dak van een woning en wat de in dit verband relevante legespolitiek is van de deelnemers.

Bovendien wordt er in de praktijk niet of nauwelijks tegen benchmarking in kleine groepjes geprotesteerd. Dus ook zo gezien, is werken met kleine groepen verdedigbaar.

Ten derde, ambtenaren voelen wel aan dat als er pakweg 100 gemeenten vergeleken worden door een onderzoeker van een of ander centraal planbureau of een andere grote andere organisatie de hogeschool van benchmarking betreden worden. Dan zal een onderzoeker multivariate analyses uitvoeren, dan vliegen de regressiecoëfficiënten je om de oren en dan gaan tal van niet in methoden en technieken van onderzoek geschoolde lezers, zoals een gemiddelde jurist, afhaken.

Maar er zijn ook tegenargumenten te geven, dus waarom het werken met kleine groepen minder zinvol is.

Ten eerste is het mogelijk dat ongeveer gelijke gemeenten ook nagenoeg gelijk scoren. Dat er dus nauwelijks verschillen zijn en er helemaal geen beste praktijken zijn te vinden. 'Von einer herausragende Gemeinde ist nicht die Rede'. Dan zit je aardig te kijken met je benchmark gericht op verbetering op basis van de beste praktijken.

Ten tweede, is er misschien wel een betere gemeente te vinden maar die is niet goed genoeg. Deze gemeente maakt te weinig het verschil, ze springt er niet voldoende positief uit. Dit verschijnsel namen we ook waar bij de rangordening van scores bij 'De Staat van de Gemeente'. Denk maar aan de scores op de rol van 'partner' of 'klant'.

Ten derde, spreekt tegen kleine aantallen gemeenten met ongeveer dezelfde grootte qua inwoneraantal dat gemeenten soms vergeleken wensen te worden met een wat grotere gemeente waarvan actoren het gevoel hebben dat ze daarvan iets kunnen leren. De gemeente Amsterdam wenst zich nog wel te laten vergelijken met Hamburg of München maar niet te laten vergelijken met Assen, Den Bosch, Papendrecht of Emmen. De praktijk toont dat een delegatie die iets van veiligheidsbeleid wil weten steevast richting New York, Los Angeles of Washington gaat. Bloomberg ontvangt Van der Laan, dat werk. Juist daar denkt men de grotere schaal te vinden waar dingen in gang gezet worden of juist verdwijnen, zoals dat wellicht in Amsterdam ook zou kunnen gebeuren. Met andere woorden, gemeentebestuurders, politiechefs en andere representanten van de veiligheidsdriehoek zoeken contrast om zo afstand in denken te kunnen nemen van bestaande vastgeroeste of ingesleten praktijken.

Dat betekent dat een benchmark wel degelijk een zoekproces is naar het betere. Soms doen gemeenten die vergeleken worden het collectief matig, en is geen 'best practice' te vinden, of al collectief vrij goed, waardoor vanuit de gedachte 'behoud het goede' niet eens een zoekproces naar een 'best practice' nodig is. Het kan nuttig zijn te overwegen of juist het zoeken naar het beste of het betere niet op een andere wijze dan via benchmarking kan geschieden. Benchmarking is geen universeel paracetamolletje; genezen kan ook op een andere manier.

10.11 Enkele conclusies

In een aantal opzichten tevredenheid

De conclusie is dat de burgers het meest tevreden zijn als ze voor een product of dienst in het gemeentehuis komen. Ze zijn tevreden over de medewerkers die hen te woord staan en over de informatie die ze daar krijgen. Ze zijn ook tevreden over de mogelijkheden om informatie te krijgen via de gemeentelijke sites, al blijkt dat niet direct uit het berekende cijfer via de landelijke website-ranglijst. Ook over het gemeentehuis zelf zijn de burgers in het algemeen content, op enkele punten na. Ook over de onderwijsmogelijkheden zijn de burgers goed te spreken. Over de directe woonomgeving, inclusief het onderhoud daarvan en de afvalinzameling, en over de voorzieningen in buurt en gemeente zijn eveneens weinig klachten.

Onvrede

Burgers van veel gemeenten zijn absoluut en relatief ontevreden over hun dagelijks bestuur (college) en over de gemeenteraden. Burgers van een groot aantal gemeenten voelen zich gemiddeld genomen niet voldoende of goed vertegenwoordigd in de gemeenteraad, hebben de indruk dat ze weinig invloed hebben op besluiten en dat toezeggingen door het college niet worden nagekomen (ook Korsten, 2010a). Het vertrouwen in het college van B&W is dan ook niet erg groot. Ze vinden bovendien dat ze te weinig terugzien van hun belastinggeld. Op een specifieke vraag over 'goed bestuur' geven burgers bijna overal het gemeentebestuur niet het voordeel van de twijfel.

Ook bij de gemeentelijke regelgeving hebben burgers tamelijk wat bedenkingen en weinig vertrouwen. De duidelijkheid scoort nog neutraal, maar de regels zijn hier en daar tegenstrijdig en op bepaalde terreinen ontbreken juist regels. Ook over de handhaving van de regels is de burger ontevreden, al is het niet duidelijk of dat nu juist gaat om een teveel of een tekort aan handhaving.

Gemeentebesturen kunnen dus nog niet achterover leunen, want hoewel veel rapportcijfers voldoende zijn, is er geen sprake van ruime voldoendes. Ze moeten in ieder geval hun pijlen richten op het vergroten van de tevredenheid van de kiezers en vertrouwensgroei en niet alleen de kaarten zetten op de verbetering van de dienstverlening.

De excellente gemeente bestaat niet

De rangordening van gemeenten bracht nog enkele andere interessante aspecten van lokale benchmarking aan het licht. Zo blijkt de perfecte of excellente gemeente niet te bestaan. Er zijn in de ogen van oordelende burgers blijkbaar altijd wel zwakkere punten, bij kleinere gemeenten en grote gemeenten. De totaalscore van de gemeente met positie één op een ranglijst van 151 gemeenten heeft op zes vergelijkingspunten iets minder dan 42 punten, wat een gemiddeld betekent onder de score 'zeven'.

Geen uitschieters

Het aantal positieve uitschieters ('product champions') is beperkt. Dat maakt begrijpelijk dat 'best practices' in bepaalde opzichten niet te vinden zijn. Ze bestaan eenvoudig niet. De scores van al de gemeente liggen hier en daar - op een beoordelingspunt - dicht bij elkaar, tussen een 'voldoende' en 'ruim voldoende', waardoor ook 'bad practices' ontbreken. Het was juist de bedoeling van benchmarking dat die de beste praktijken aan het licht zou brengen. Wie zich als gemeente wil verbeteren kan dus onder deze condities niet op zoek naar een andere 'beste' gemeente. Dat geldt niet op alle punten. Er bestaan beoordelingen waaruit wel degelijk de beste praktijk naar voren komt.

Het valt verder op dat een gemeente die niet in de 'top tien' van totaalscores staat in een bepaald opzicht te midden van 150 andere gemeenten wel 'aardig kan scoren', zelfs als beste. Dat is begrijpelijk. Er is geen reden om te veronderstellen dat er een gemeentebestuur met een organisatie zou bestaan die in tal van opzichten de beste praktijken zou kennen.

Bij alle beoordelingen door burgers moet overigens beseft worden dat een verschil in sociaal-culturele en religieuze samenstelling van de bevolking kan uitmaken voor de scores. Als de burgers over het algemeen een sterk 'pro overheidsgevoel' hebben, neigen ze ertoe om gemeenten op bepaalde punten eerder het voordeel dan de twijfel te geven.

Wat ontbreekt

De ontwikkeling van 'De Staat van de Gemeente' was een arbeidsintensieve en moedige poging, waarvoor de initiatiefnemer, de VGS, lof verdient. Deze staat heeft echter ook een begrenzing. De staat maakt niet duidelijk waarom de ene gemeente in een rol beter scoort dan een andere. Vanuit het kwaliteitscriterium 'een verschillenanalyse is nodig om de rangorde te begrijpen, scoren deze integrale benchmarks niet voldoende. In andere woorden, een oorzakenanalyse ontbreekt. Deze benchmarkrapporten verklaren niet. Daar staat tegenover dat met de opzet van de benchmark een goede prestatie is geleverd.

'De Staat' geeft ten tweede wel een klassement maar geen 'best practice'. De rapporteurs leggen niet uit wat navolging verdient. De gemeentebesturen krijgen daardoor geen handelingsperspectief aangereikt om stappen naar de top te kunnen zetten.

Zo gezien, maakt benchmarking enkele pretenties die in de literatuur geventileerd worden in de gemeentelijke sfeer niet waar.

11 Bespreking van benchmarkrapportages: gebruiken, leren en verbeteren?

We hebben nu verschillende benchmarks bekeken. Op basis daarvan is nog niet te generaliseren in termen van nut en noodzaak van benchmarking, of leren en verbeteren. Toch zijn er pogingen gedaan tot generalisatie over meer benchmarks en meer gemeenten heen. Wat is het beeld van benchmarking in termen van leren en verbeteren?

11.1 Sceptis: geen directe benutting

We gaan eerst maar eens buiten het bestaande kader van benchmarking treden. De Amerikaanse onderzoekster (en latere Harvard professor) Carol H. Weiss vertelde ooit dat ze enkele jaren onderzoek gedaan had naar een bepaald thema rond 'The War on Poverty' (onderzoek in Harlem, New York), het driedelige rapport naar Washington stuurde en toe afwachtte wat er ging gebeuren. Hoezeer 'Washington D.C. als centrum van de federale politieke discussie in de Verenigde Staten ook geroezemoes kent, rond haar rapport bleef het stil, heel lang oorverdovend stil. 'Not a whisper from Washington, not a hint that anyone had even read a word', aldus Weiss (zie internet; ook Engbersen, 2009: 11). Een verrassende opmerking voor iemand als Weiss die al sinds het einde van de jaren zeventig bezig was met de relatie tussen onderzoek en beleid en dus gepokt en gemazeld was tegen (over)optimisme. Zij had onderkend, met anderen, dat onderzoeksresultaten vaak pas op *langere* termijn benut worden. Dat de benutting, als die al plaatsvindt, niet instrumenteel maar soms conceptueel is. En dat de benutting niet direct is, maar veelal indirect tot gebruikers in de bestuurlijke arena's doordringt. Dat betekent dat een onderzoeksrapport na verschijning eerst tegen het licht gehouden wordt en gerecenseerd wordt, dat onderzoeksuitkomsten samen met ander onderzoek gewikt en gewogen worden in een branche of sector en op conferenties, en dat inzichten vermengd raken met beschouwingen van anderen in persmedia, soms ook met belanghebbenden of pressiegroepen. Onderzoek werkt dus vaak niet een-op-een, direct en instrumenteel op korte termijn door. Die veronderstelling wordt ontkracht (Korsten, 1983).

Onderzoekers moeten, zo wist Weiss, ook oog hebben voor de functies van onderzoek, zoals

- de functie van tactisch gebruik: onderzoek dient ertoe om beleid uit te stellen en geen beleid te hoeven maken; spelen met tijd is in de politiek belangrijk;
- symbolisch gebruik: wel lippendienst aan een waarde bewijzen maar beleid niet instrumentaliseren;
- gebruik als munitie: een rapport om te strijden tegen anderen;
- of een rapport gebruiken voor verheldering van zaken: 'enlightenment';
- gebruik als rechtvaardiging voor het bestaande: 'legitimation';
- of als vertaling in concrete doelstellingen en instrumenten: 'engineering' genoemd.

Kortom, onderzoek en politiek bestuur zijn onderscheiden werelden. Politici en bestuurders gebruiken onderzoek soms op het moment dat het hen uitkomt. En zo nodig op de manier die hen uitkomt (ook: Derksen, 2011). Onderzoek wordt soms ook zwart gemaakt. De keer dat een gemeente als Schiedam evident slecht scoorde in '*De Staat van de Gemeente*' (Bouwmans, 2009) en er enige vragen vanuit de gemeenteraad over werden gesteld, reageerde het college van burgemeester en wethouders van de gemeente uit het Rotterdamse havengebied schriftelijk dat het onderzoek gebaseerd was op data van enkele jaren terug. De indruk werd gewekt dat de situatie nu veel beter was, waarvoor overigens geen bewijs werd aangedragen. Daarmee was ook voor de raadsmeerderheid de kous af want de coalitie mocht geen verdere schade oplopen.

Vanuit de ervaringen van Weiss moeten we juist niet naïef zijn door te veronderstellen dat iedere opdrachtgevende bestuurder en betrokken ambtenaar zit te smachten naar het rapport met een vergelijkend verhaal en met vermelding van beste praktijken en het waarom daarvan. Wellicht zijn ze wel nieuwsgierig maar veroorzaakt een benchmarkrapport, waarvan de politiek-bestuurlijke prioriteit misschien gering is alleen maar onrust of roept het rapport verdeeldheid op. Van de andere kant, onder een van de kabinetten-Balkenende is de '*Code voor goed openbaar bestuur*' opgesteld, die

inhoudt dat overheden eigenlijk moeten leren en dus zich stelselmatig moeten verbeteren. Dat klinkt direct al wat opgewekter.

Leren is dus belangrijk. Enkele onderzoekers van PricewaterhouseCoopers (PwC), die veel benchmarks uitvoerden op het vlak van zorg en beroepsonderwijs, zetten nog een tandje bij. Ze stellen zelfs:

‘Het succes van een benchmark staat of valt (...) met de mate waarin de instelling daadwerkelijk verbeteringen weet door te voeren’ (Poerstamper e.a., 2007: 35).

11.2 Aanvaarding en belang van benchmarking

Wat blijkt? ‘Opmerkelijk genoeg lijkt juist het leereffect van benchmarking niet altijd te worden bereikt’ (Poerstamper e.a., 2007: 35). Lijkt....Met ‘lijkt’ kunnen we niet tevreden zijn. We gaan eens verder op zoek en proberen vooral te kijken naar gemeenten.

Benchmarking in gemeenten is als een van de managementvernieuwingen van na 1985 tamelijk positief ontvangen. Dat blijkt ook uit bewoordingen als: ‘benchmarking kan’; ‘benchmarking is niet verkeerd’, zo zeggen ambtenaren en bestuurders in koor (Ter Bogt, 2005). ‘Altijd nuttig om over de grenzen te kijken’. Maar als benchmarking kan, wat dan met de uitkomsten van onderzoek te doen? Tenslotte gaven tal van overheden opdracht tot benchmarking, wat onder meer bleek uit waarstaatjegemeente.

Het oordeel over het effect van benchmarking als instrument op de doelmatigheid en effectiviteit van gemeentelijk handelen is eveneens ‘voldoende tot positief’, concludeert Henk ter Bogt (2005: 54). Zo gezien, is sprake van ‘diffuse’ *erkenning en steun* onder ambtenaren en ook wel onder bestuurders voor benchmarking.

Ook verbetering na herhaalde meting? Uit de scores op *waarstaatjegemeente*, waaraan nogal wat gemeenten meerdere malen deelnamen, blijkt geen positieve groei in scores. Het is niet zo dat gemeenten drie jaar na de eerste meting bij de tweede of derde meting evident beter scoren op de rol van klant, onderdaan, wijkbewoner, partner en kiezer. Voor het bereiken van duidelijk betere of hogere scores in de ogen van burgers zijn de verbeterpogingen, als ze er al waren, blijkbaar niet groot genoeg geweest; althans ze waren niet zichtbaar voor burgers. Dat zegt overigens niet alles omdat er naast deze integrale meting tal van andere, partiële benchmarks plaats vonden, zoals op het vlak van bouw- en woningtoezicht. En wellicht traden dáár wel evidente verbeteringen op, maar hebben tal burgers die ook niet opgemerkt.

11.3 Kritiek

Als benchmarkonderzoek wel bekend raakt, worden uitkomsten dan van betekenis gevonden of niet? De uitkomsten van in het begin van de jaren nul uitgevoerd benchmarkonderzoek ‘werden eveneens veelal van belang gevonden’, aldus de Groningse hoogleraar Henk ter Bogt in zijn oratie, hoewel benchmarkonderzoeken ‘niet zo zeer precieze informatie oplevert over “goed of slecht” presteren van een organisatie’ en ‘zeker niet over de achtergronden ervan’ (Ter Bogt, 2005: 51). Respondenten blijken dus *aarzelender* over ‘specifieke’ erkenning van en steun aan benchmarkonderzoek in relatie tot hun *eigen* organisatie.

De volgende aspecten zijn in dat verband belangrijk.

Het vergelijkbaarheidsprobleem: Veel geïnterviewde ambtenaren en bestuurders gaven in survey-onderzoek aan dat de gegevens in gemeentelijke benchmarkstudies ‘lang niet altijd’ vergelijkbaar zijn (De Groot e.a., 2004; Ter Bogt, 2005: 51). Misschien zijn de diensten, producten of processen nog wel (ongeveer) gelijk maar gaan er achter organisaties verschillende doelstellingen en oogmerken schuil. Het is nuttig als de benchmarkonderzoeker daar weet van heeft. De prijs voor een vergunning voor de bouw van een dakkapel, bij voorbeeld, kan in de een gemeente pakweg 40 euro bedragen en in

een andere gemeente om en nabij 750 euro. Gaat er een gelijke redenering achter beide prijsstellingen schuil? Nee, de organisatie die vanwege een besluit van de gemeenteraad 750 euro vraagt, meent dat zo de vergunningaanvraag volledig kostendekkend is terwijl het gemeentebestuur dat het lage bedrag in rekening brengt er geheel anders over denkt. Zo gezien, is er ook sprake van een zekere reserve bij de veronderstelde gemakkelijke transfereerbaarheid van 'best practices'. Conclusie: voor een goed begrip van rapporten en het kunnen zoeken naar de beste praktijken is het nodig dat aangegeven wordt wat wel vergelijkbaar is en wat niet, of niet goed. Gemeenten die volledige kostendekkendheid van bouwaanvragen nastreven, kunnen vergeleken worden. Maar ook de vergelijking tussen het wel of niet nastreven van kostendekkendheid is zinnig: hoeveel besturen van organisaties willen dat eigenlijk en waarom?; is dat slechts een minderheid van organisaties? Meestal wordt in benchmarkstudies geen aandacht besteed aan de (on)zinnigheid van normeringen en argumentatie die achter diensten, producten en processen schuil gaan.

Het interpretatieprobleem: Veel geïnterviewden geven aan dat de wel vergelijkbare gegevens uit lokale benchmarkstudies 'lang niet altijd goed te interpreteren' zijn (Ter Bogt, 2005: 51). Dit is evident een probleem in tal van rapporten. Soms lijkt een benchmark door een specialist opgesteld ten behoeve van specialisten. Maar onder de potentiële gebruikers bevinden zich anderen dan specialisten. Politici zijn geen specialisten maar generalisten.

Het onzichtbaarheidsprobleem: Bij de analyse van de rangorde van gemeenten die deelnamen aan *waarstaatjegemeente* is vastgesteld dat in bepaalde opzichten helemaal geen beste praktijken waarneembaar waren. Veel gemeenten scoorden op of rond een gemiddelde score. Een voorbeeld. De scores van alle 240 gemeenten gingen een score van zeven op een schaal van 1-10 ten aanzien van wat gemeenten doen voor wijkbewoners en hoe burgers als wijkbewoner naar gemeenten kijken niet te boven. Als er geen 'best practice' naar voren komt, dus geen score die evident boven een 7 uitkomt, kan een gemeente ook niet (verder) verbeteren op basis van een 'best practice'.

Het uitblijven van de specificatie per gemeente: In tal van benchmarkrapporten wordt duidelijk dat een deelnemende organisatie iets aan de prestaties (bij voorbeeld van producten of diensten) kan en misschien ook moet doen en ook *waar* (de richting van verbetering) maar niet *wat* die moet doen om in de richting van de 'best practice' op te schuiven. In andere taal gesteld: de rapporten maken niet duidelijk aan welke knoppen een organisatie door middel van een verbeteractie moet draaien. Om een voorbeeld te geven: als een gemeente A matig scoort op overlastbestrijding in vier van acht wijken, als onderdeel van veiligheidsbeleid, en een andere gemeente B heel goed, wat doet B dan wat A niet goed doet en wat moet A dan doen om de score van B te benaderen? Of waarom is Littenseradiel al enkele jaren een van de veiligste gemeenten van Nederland? Wil een benutting van een gemeentelijk benchmarkverslag mogelijk zijn, dan is vaak *nader onderzoek* nodig naar wat een 'best practice' inhoudt en naar de achtergronden van een bepaalde goede of slechte score. Wat doet deelnemer X die laag scoort niet dat deelnemer Y die (op een bepaald vlak) een 'best practice' bereikt blijkbaar wel doet? Alleen een 'best practice'-cijfer sec geven of een plaats van een gemeente te midden van alle andere gemeenten, maakt nog niet duidelijk wat die praktijk inhoudt en zo goed maakt en waarom die score bereikt is, dus wat erop inwerkt. Aldus de communis opinio onder kenners (Martens e.a., 2002; Ter Bogt, 2005: 51; Commissie-Van Rijn, 2007). Dat is iets dat ook al bleek naar aanleiding van een benchmarkstudie over het presteren van rechtbanken. Maar dat nader onderzoek blijft vaak uit, stelden we eerder, en rapportages over dat nader onderzoek zijn in het algemeen schaars. Het verrast zo gezien dan ook niet dat we geen studie kennen die gaat over de omvang van verbeteracties en de kwaliteit daarvan naar aanleiding van eerdere benchmarkrapportages. Conclusie: wil benchmarking kwaliteit hebben dan moet een benchmarkrapport uitleggen wat een 'best practice' inhoudt en bij voorkeur ook wat een 'bad practice' inhoudt en hoe een deelnemer een 'best practice' kan bereiken. En dan is specificatie van uitkomsten, verklaringen en bij voorkeur ook van verbeterrichtingen per deelnemende organisatie nodig. Dat betekent, volgens onze veronderstelling, een omslag ten opzichte van veel

benchmarkpraktijken. Als die omslag gemaakt wordt, wordt ook de kans op benutting van rapporten (en vervolgtrajecten) groter.

De geringe ontvankelijkheid bij wethouders voor prestatiegegevens in schriftelijke vorm: Jansen (2008) concludeert in een onderzoek naar het gebruik van prestatie-informatie in drie gemeenten dat politieke bestuurders een onderscheiden perspectief hebben op prestaties en prestatiemeting, in vergelijking met managers en andere ambtenaren. Ambtenaren hanteren een intern perspectief, terwijl politieke bestuurders meer gericht zijn op de financiële kant en op hun verhouding met burgers. Jansen sloot daarmee aan op bevindingen uit eerder onderzoek van Ter Bogt, die aantoonde dat veel bestuurders (wethouders) weinig betekenis toekennen aan prestatiegegevens. Voor zover dezen wel informatie willen hebben, willen ze die vooral mondeling te horen krijgen van iemand die het rapport kent. Wethouders zijn over het algemeen geen types die zich vol ijver studieus door benchmarkrapporten en andere 'performance'-rapporten heen ploegen (Ter Bogt, 1999; 2000; 2001a, b; 2002, 2003; 2004a). Wie wil bereiken dat bestuurders toch schriftelijke benchmarkinformatie tot zich nemen moet daartoe een prikkel inzetten. Bij voorbeeld door erop te wijzen dat burgers (kiezers) om voorbeelden zullen vragen van publiekelijke verantwoording.

Bekijk het eens vanuit de ambtelijke achtergrond: Is het altijd persé nodig dat een benchmarkrapport wordt behandeld door een college van burgemeester en wethouders of de gemeenteraad? Niet altijd vinden managers of directeuren dat nodig. En, zoals we hiervoor ook zagen, niet altijd zijn bestuurders en politici ontvankelijk voor dergelijke rapporten. Maar laten we het eens bekijken vanuit de positie van de ambtelijke kant.

Veel benchmarks zijn aan te merken als bedrijfsvoering en daarmee meer een zaak van de organisatie dan van het gemeentebestuur. In de klassieke denkbeelden over een gemeentebestuur bestuurt de politiek en voert de ambtelijke organisatie uit. Anders geformuleerd: politiek en bestuur gaan over de 'wat-vraag' en de ambtenaren over de 'hoe-vraag'. Met dit wat stellige uitgangspunt is het ook te verdedigen dat de ambtelijke organisatie zelf onderzoeken uitvoert om die 'hoe-vraag' efficiënter te beantwoorden. Want daar kan toch niemand op tegen zijn? Ambtenaren die een benchmarkrapport laten samenstellen om te kijken hoe het beheer van de riolering in hun gemeente staat ten opzicht van andere gemeenten, dat kan alleen maar voordelen opleveren. En baat het niet, schaden doet het ook niet.

Maar een benchmark kan ook georganiseerd worden om te anticiperen op mogelijke vragen van een politiek bestuur. Wanneer er bezuinigd moet worden, kijkt een raad of het college al snel naar de post salarissen. De roep om te bezuinigen op het ambtelijk apparaat ligt voor de hand. Wat is er dan handiger om de resultaten van een benchmark formatievergelijking of een benchmark overhead alvast paraat te hebben?

Kortom: er zijn rapporten van benchmarks die vrijwel uitsluitend het ambtelijke domein betreffen. Nogal eens wordt dat aangeduid als bedrijfsvoeringsdocumenten. Het is goed verdedigbaar dat dergelijke rapporten binnen het ambtelijke domein blijven en hooguit 'ter kennisname' naar het bestuur worden gestuurd.

Zo gezien, vormen deze factoren en omstandigheden een rem op benutting van benchmarkstudies in de concrete dagelijkse gemeentelijke praktijk en op verbeteren (lees: leren). Als benchmarkstudies niet of weinig benut worden, kan dat dus liggen aan kenmerken van de onderzoeken en rapportages daarover (dus aan moeilijk te interpreteren bevindingen; nader onderzoek gewenst).

We gaven eerder al aan dat de constatering van verschillen tussen organisaties uit de publieke en private sector ook een rem is op gebruik van benchmarking in overheidsorganisaties, zoals gemeenten (cf Tillema, 2006). Gemeenten worden hoofdzakelijk bestuurd door politiek gekozen bestuurders, politici hebben te maken met achterbannen en belanghebbenden, en besturen - althans in algemene besturen, zoals in gemeenteraden -, in het volle licht van de openbaarheid. En dan is er ook nog het geheel van complexe producten en processen en de meervoudigheid van doelstellingen. Dat maakt het allemaal complexer dan in de (meeste) private organisaties.

Maar benchmarking kan ook ineens verrassend gebruikt worden, veelal *in combinatie* met andere informatiebronnen. Dat heet *het percolatoreffect*, alsof informatie uit verschillende bronnen zich vermengt. Dat betekent – vanuit een andere metafoor gezien – dat als benchmarking in een bredere context aan de orde komt dit onderzoek een lift krijgt.

Een voorbeeld hiervan betreft een waterschap in het Noorden des lands dat had deelgenomen aan een benchmarkonderzoek naar de verhouding directen-indirecten in het personeelsbestand. Uit dat onderzoek bleek in 2013 dat het waterschap veel indirecten had, meer concreet een te grote ambtelijke ‘top’. Daarbij was sprake van een onwerkbaar situatie. Het gevolg was dat het bestuur ingreep, onder meer door de ambtelijke top te halveren. Concrete aanleiding voor ingrijpen was dat directieleden niet meer door één deur konden. De oorzaak voor ingrijpen lag echter dieper. De werkwijze van de organisatie, de organisatiecultuur, was dusdanig dat moest worden ingegrepen volgens de dijkgraaf, die toevoegt: ‘Er zijn onderzoeken geweest waaruit tekortkoming bleken, onder meer de vele indirecte functies. (...) We willen van bovenaf beginnen’. De dijkgraaf verder: ‘De cultuur moet ook veranderen. Er moet effectiever en efficiënter gewerkt worden. Het zit hem vooral in minder vergaderen en overleg. Het gaat ook om de directe aansturing. Nu is vaak onduidelijk wat iemands taak is of wie de leiding heeft, de procesmanager of de unitmanager? Er moet directe aansturing zijn’. Het waterschap staat ook teveel met de rug naar de samenleving (BB, 5 juli 2013). Zie zien we dat een benchmarkonderzoek bestuurlijk volop aandacht kreeg.

Benchmarking kent ook tactisch gebruik. Benchmarkrapportages kunnen gebruikt worden als munitie voor lokale politieke gevechten. Een gemeentesecretaris vertelde ons dat hij eens de gemeente liet deelnemen aan een benchmark die ‘overhead’ meet. Dat was een heel nuttige deelname want als vanuit de gemeenteraad weer eens werd geroepen om bezuiniging te laten neerslaan op de ambtelijke organisatie en beperking van de personeelsomvang geopperd werd dan kon hij als gemeentesecretaris deze suggestie pareren met de benchmark die liet zien ‘dat het met de overhead reuze meeviel’. En een hoofd openbare werken vond een benchmark van rioleringszorg nuttig om te laten zien dat verbeteringen nodig waren of om te beargumenteren dat ‘de tarieven’ moesten stijgen. Dat betekent dat benchmarking een indirecte betekenis kan hebben. Rapporten worden uit de kast gehaald als zich politieke suggesties aandienen.

11.4 Benchmarking en twee soorten leren: marginaal leren voorop

Overheidsorganisaties moeten volgens de ‘Code voor goed openbaar bestuur’ lerende organisaties zijn. Hoe zit het nu met dat leren in gemeentelijke organisaties en door bestuurders? 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 organisatie taak 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 (beleids)strategie en is verbetering iets op het vlak van instrumentatie van beleid, dus iets van de middelen (mensen, budget, voorzieningen). Bij ‘dubbel cirkel leren’ (‘double-loop learning’) worden de waarden, strategie, doelstellingen en referentiekaders achter beleid juist wel ter discussie gesteld. Hier is sprake van een radicalere bezinning op beleid (Argyris & Schön, 1978; Bouckaert e.a., 2011: 108). Het gaat dus om marginaal leren (een cirkel leren; model one) of fundamenteel leren (dubbel cirkel leren; model two).

Wat leert onderzoek naar verbeteringen als gevolg van benchmarking ten aanzien van leren? Alle aanwijzingen uit verricht empirisch onderzoek geven grond aan de grote waarschijnlijkheid dat benchmarking *zelden* leidt tot ‘dubbel cirkel leren’ of fundamenteel leren (Ter Bogt, 1999, 2000, 2001a, b, 2005; Aardema, 2002; Moynihan, 2005; Tillema, 2006; Ananias, 2007; Lapperre, 2007; Noordhoek, 2013: 98). 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’ (Van der Zee, 2005).

De door de Unie van Waterschappen geïntroduceerde benchmarkprojecten betreffende afvalwaterverwerking waren zowel gericht op externe verantwoording als op leren en verbeteren. Prestaties werden vanuit vier optieken aan analyse onderworpen: operationele doelmatigheid; de financiële prestatie; de ecologische prestatie; en de mate van innovatie. Analyse van twee uitgevoerde benchmarkstudies leerde: 'de functie van externe verantwoording bleek meer uit de verf te komen dan die van leren en verbeteren' (De Bruijn en Van Helden, 2007: 23; ook Tillema, 2006).

Wat draagt ertoe bij dat de kans op 'double loop learning' als gevolg van benchmarking 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, aldus Van der Zee (2005: 146).

Gemeentelijke benchmarking leidt niet tot fundamentele veranderingen, niet tot radicale beleidsbreuken. Op papier is gemeentelijke benchmarking een middel om te voorkomen dat gemeentelijke organisaties en het gemeentebestuur in zichzelf gekeerd zijn, maar in werkelijkheid zijn organisaties in het algemeen wel sterk met zichzelf bezig, of het nu om gemeenten, onderwijsorganisaties, kerkelijke organisaties, hulporganisaties of welke organisatie dan ook gaat. Iedere organisatie wordt op zeker moment een gesloten wereld met een eigen cultuur waarvan de leden niet uitsluitend maar wel heel sterk met zichzelf bezig zijn en te weinig met waarvoor ze eigenlijk bestaan of overleven in turbulente tijden. Benchmarking breekt niet vanzelfsprekend of krachtig door de muur van geslotenheid heen. Benchmarking blijkt geen machtsfactor van betekenis te zijn.

Organisaties blijken –in het licht van de kijk van Niklas Luhmann hierop - *zelfreferentieel*, dat wil zeggen dat de neiging daarbinnen sterk is om de omgeving te zien vanuit een geheel eigen perspectief op wat belangrijk of onbelangrijk is zonder dat men in het oog heeft of dit ook zo is. Organisatieleden zoeken steun bij elkaar en houden zich met de eigen organisatie bezig. Organisaties neigen er dus toe om als het even kan nog lang spijkers te maken ook als er al lang geen behoefte meer aan. Gevolg van die geslotenheid: de veranderingsgezindheid en innovatieve kracht zal beperkt blijven en de collectieve reflectie zal gering zijn.

Kan benchmarking dat bij gemeenten doorbreken? Niet zo dat sprake is van fundamenteel leren. Crises, zoals een bankencrisis, woningcrises of een ernstige breuk in het consumentenvertrouwen in de samenleving, leiden veeleer tot meer radicale veranderingen in beleid dan één rapport over één benchmarkonderzoek. De kans dat door benchmarking sprake is van 'één cirkel leren' is dus veel groter dan dat sprake is van 'twee cirkel leren'.

Het lijkt goed om nog iets toe te voegen, namelijk om een onderscheid te maken tussen leren van partners in een benchmarking en leren van het sectorbeeld dat oprijst uit een of meer benchmarks. We hebben hier te maken met een *niveaoverschil*. Een voorbeeld. Zorgverzekeraars kunnen leren van het totaalbeeld dat naar voren komt uit het functioneren van ziekenhuizen als blijkt dat bepaalde ziekenhuizen veel en veel meer declareren dan andere voor activiteiten a, b en c dan drie of vier jaar terug. Dan kan op het mesoniveau van de sector geleerd worden en wel op een andere manier dan individuele ziekenhuisorganisaties dat doen. Niet uitgesloten lijkt dat 'leren' in de sector op mesoniveau zich onderscheidt van leren op microniveau. 'The organizational level influences the mode of use' heet dat bij De Lancer Julnes (2008).

Wat weten we hierover nog meer? Wat zou dat 'een cirkel leren' kunnen inhouden? 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). We wijzen er overigens wel op dat Peter Senge

(1992) vaststelt dat er vaak minder wordt geleerd dan mogelijk is. Niet dat het ontbreekt aan goede bedoelingen maar 'we zitten ons zelf in de weg'. De eigen 'mental models' zitten ons dwars, de eigen geestelijke gevangenis speelt ons parten. Zijn advies: maak die mentale voorstelling van zaken bespreekbaar. Chris Argyris en Donald Schön (1978) hebben aandacht voor teams, die niet zelden verdedigend gedrag vertonen.

11.5 Geen breuken in beleid maar wel deliberatie?

Goede prestatie meting levert relevante, actuele, geldige, betrouwbare en begrijpelijke informatie op die tot bezinning en discussie kan leiden over daadwerkelijke verbeteringen in de instrumentatie van beleid en dus werkwijzen. Een stelling kan daarom zijn: Benchmarking is een van de innovatieve instrumenten voor reflectie en deliberatie (Denters en Toonen, 2010: 226). Goede benchmarkrapporten zijn een opstap voor een discussie in een forum en daarmee een hulpmiddel in 'the deliberative democracy', maar ze zijn op zich *niet genoeg* om onmiddellijk een verbetering te veroorzaken. Ook voor een incrementele verandering ('stukje bij beetje') is over het algemeen *meer* nodig (Korsten, 1983). 'Measurement by itself will *not* improve performance', zegt Callahan (2007: 70). Dit wordt ook benadrukt door auteurs die menen dat prestatie meting door middel van benchmarking niet het belangrijkste is maar juist wel *de praktijk van bezinning* op prestaties. In die praktijk kan een op leren gerichte cultuur en forumwerking aanwezig zijn of groeien waarin meerdere instrumenten worden gebruikt (Noordegraaf, 2011).

11.6 Van invloed op benutting of gebruik

Wat is dan zoal van belang? Wat werkt in op intensief of matig gebruik of non-gebruik van bevindingen uit benchmarkstudies? Daarover is veel te doen in de literatuur (o.a. Van de Vall, 1980; Korsten, 1983; Ter Bogt, 2000, 2005; Askim, 2007, 2008; Askim, Johnson & Christophersen, 2008; Van Dooren en Van de Walle, 2008; De Lancer Julnes, 2008; Dixon e.a., 2013). We leggen ons in dit verband enkele beperkingen op (zie verder Korsten, 1983).

Allereerst is van belang te weten wat *het doel* is van een benchmark en rapportage daarover. Is het doel slechts om verantwoording af te leggen, dan heeft het rapport een externe functie en is de 'audience' niet primair de interne ambtenaar die - met instemming van het bestuur - het bestaande beleid kan proberen te verbeteren. Is het doel daarentegen juist om te leren van vergelijking en rangordening dan zijn de ambtenaren die gaan over het object van benchmarking wel een belangrijke doelgroep. Dan zal de insteek eerder zijn om te pogen een rapport intern te bespreken en dus te benutten. Daarom treffen we in de literatuur het advies aan: let als je wilt weten of van benutting sprake is eerst op het doel en de doelgroep van een benchmark, want doel en doelgroep hangen samen met het te verwachten gebruik en de gebruiksfuncties (Behn, 2003; De Lancer Julnes, 2008; Noordhoek, 2013). Dit impliceert een nuancering bij wat opgemerkt is over 'double loop learning' en 'single loop learning'. Als het doel van benchmarking verantwoording is en de doelgroep van een benchmark de samenleving is, dan valt niet direct te verwachten dat ambtenaren en bestuurders een poging tot 'learning', van welke aard ook, ernstig zullen overwegen of zelfs ondernemen.

Wat is nog meer van belang? Sommigen stellen dat de aard en inhoud van een benchmarkrapportage ook een heel grote invloed heeft op de benutting (Korsten, 1983). Een benchmark moet organisaties in beeld nemen die ook *qua context vergelijkbaar* zijn. Als een echte vergelijking niet geremd wordt door valkuilen waar een onderzoeker instapt (zie verderop) en een onderzoeker dus verschillen in scores ook weergeeft *en* verklaart, en 'best practices' en aanbevelingen geeft, dan is een kans op benutting van bevindingen in een deelnemende organisatie aanwezig.

De context is een belangrijke factor die van invloed is op gebruik of non-gebruik van benchmarkuitkomsten en ook op de gebruiksvorm (Noordhoek, 2013: 63). Rousseau (2006) merkt op dat organisaties soms bepaald bevindingen, bij voorbeeld uit benchmarking, niet oppakken omdat ze de eigen situatie als uniek zien. Die redenering is lang niet altijd terecht.

Een ander aspect is of een ontvanger *vertrouwen* kan hebben in de kwaliteit van een benchmarkrapport (De Lancer Julnes, 2008: 64). Als de klippen omzeild zijn en de valkuilen bij benchmarking vermeden, kan het vertrouwen groter zijn dan als dat niet het geval is.

Maar het onderkennen van het doel van benchmarking en de kwaliteit van de analyse en van het rapport zijn nooit de enige factoren die van belang zijn. Hakvoort & Klaassen (2010: 118) wijzen er op dat het voor de receptie van een benchmarkrapport gunstig is als de analyse zowel zwakke als sterke kanten laat zien van een in een benchmark betrokken geheel van productgroepen, voorzieningen of processen.

Verder lezen we dat de *politiek-bestuurlijke gevoeligheid* van bepaalde gemeentelijke taken, die in een benchmark aan bod komen, van belang is om het gebruik van bevindingen uit benchmarkonderzoek te bevorderen: hoe meer politiek-bestuurlijke *prioriteit* een gebied heeft dat aan benchmarking is onderworpen, hoe groter de kans dat van bezinning op bevindingen ook serieus werk gemaakt wordt. Van Elsacker (2007) komt in zijn proefschrift naar prestatie meting in gemeenten in stelling 6 tot de volgende conclusie: 'Leren van opgedane ervaringen, om hiermee prestatieverbetering na te streven, gebeurt alleen onder specifieke omstandigheden, namelijk wanneer doelstellingen en prestatie maatstaven geïntegreerd zijn en betrekking hebben op routinematige en politiek gevoelige uitvoeringstaken'.

Wie geeft opdracht en wie ontvangt het benchmarkrapport? De Lancer Julnes (2008: 60) is een van de onderzoekers van prestatie meting die het belang van *leiderschap* benadrukt. : '.... executive branch leadership positively influences strategic planning and performance management.' Leidinggevend zijn een belangrijke factor bij wel of niet benutting van benchmarkinformatie, zo toont Noordhoek op basis van vier gevalstudies aan (2013: 63).

Er zijn ook nog andere omstandigheden nodig die gunstig zijn. Verdedigers van de deliberatieve democratie schenken aandacht aan het *ontstaan en voortbestaan van op leren gerichte praktijken* ('learning forums' etc.; Moynihan, 2008a; Askim, 2007, 2008; Askim, Johnson & Christophersen, 2008). Nutley e.a. (2007) geven een lijst met aanbevelingen over hoe gebruik van onderzoeksresultaten te bevorderen. We noemen er enkele. Onderzoeksresultaten moeten altijd naar de eigen praktijk worden vertaald en in brede kring geloofwaardig zijn. Tevens is het gewenst dat er voorvechters opstaan die de vaak noodzakelijke veranderingen bepleiten. En de onderzoeksresultaten moeten passen in politieke schema's en niet haaks staan op politieke wenselijkheden.

Ook 'mindsets' doen ertoe. 'A sense of professional pride is a strong motivator for using performance information' (Moynihan, 2008a: 35). Wie een team samenstelt met voor prestatie management ontvankelijke ambtelijke en eventueel ook bestuurlijke leden zal merken dat de benchmarkrapporten beter bekeken worden dan door een team met leden met een andere 'mindset' (Boyne, James, John & Petrovsky, 2011). Sommigen stellen zelfs: vorm een team van uitsluitend juristen en je weet dat een benchmark niet bekeken wordt want juristen hebben in hun opleiding nooit te maken gehad met sociaal-wetenschappelijk onderzoek, noch met zoiets als benchmarking, indicatoren of kengetallen. Ze hebben geen oog voor benchmarking en er geen gevoel voor. Kies dus op zijn minst voor een gemengd team. Beter nog: kies voor een lerende organisatie. Te stellen is dat sprake is van een paradox: 'The outstanding condition for learning in organisations is the creation of the "learning" organisation in structural and cultural terms', aldus Noordhoek (2013: 98). In lijn hiermee valt te stellen: 'Benchmarking is only as effective as the people who apply it'. Samenvattend stelt Marike Noordhoek (2013: 101): 'The literature suggests that the municipal benchmarking is likely to deliver improved organisational performance when it is coupled to organisational learning.'

Wat is dan een lerende organisatie? Peter Senge en Patria de Lancer Julnes (2008) helpen ons hier. 'A learning organization is a place where people continually expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free, and where people are continually learning how to learn together' (2008: 69).

De rol van *de pers* is eveneens van belang. Perspublicaties kunnen de urgentie voor lezing van een benchmarkrapport vergroten. Dat lichten we toe.

11.7 Druk van de pers van invloed?

Een discussie betreft de vraag of er niet een schakeltje tussen een vergelijking en rangordening in een benchmarkstudie zit en de receptie door bestuurders, namelijk de ontvangst door de persmedia. Maakt publicatie van cijfers en kritiek op 'rankings' in de pers indruk en heeft die factor vervolgens

invloed op bestuurders? Deze discussie speelt op het vlak van onderwijs in verschillende landen als gevolg van PISA-cijfers en PIRLS. Dixon e.a. (2013: 486) deden er onderzoek naar in vier landen en formuleerden enkele hypothesen.

Hypothese 1: de reactie in persverslagen over onderwijsrankings is doorgaans negatief; en dat heeft volgens een toelichting door de auteurs weer invloed op de publieke opinie, die daardoor ook negatiever wordt (Dixon e.a., 2013: 486).

Hypothese 2: verslagen in persmedia over onderwijsrankings zijn negatief vertekend (er is een 'negativity bias').

Hypothese 3: politieke bestuurders op centraal niveau worden persoonlijk aangesproken op en verantwoordelijk gesteld voor scores in onderwijsrankings ('personally blamed in press reports of educational rankings'; 2013: 486).

Hypothese 4: rapportages over onderwijsrankings lokken een politiek-bestuurlijke response uit die beïnvloed wordt door de intensiteit en teneur van de behandeling van de rankings in de pers.

Welke hypothesen worden bevestigd of verworpen?

Hypothese 1 wordt in elk van de onderzochte onderwijsrankings in de vier landen bevestigd.

Perspublicaties blijken in het algemeen negatief kritisch. De uitkomsten van onderwijsrankings geven geen aanleiding om 'bij de pakken te blijven neerzitten'.

Hypothese 2 wordt niet bevestigd. Het blijkt niet zo te zijn dat de PISA en PIRLS studies als irrelevant, onjuist of geflatteerd worden beschouwd.

Hypothese 3 wordt zeker niet voor alle vier de landen bevestigd. 'There was relatively little blame directed personally at politicians' (Dixon e.a., 2013: 499). Er blijken wel verschillen in de intensiteit en manier waarop politieke bestuurders aangesproken worden op de uitkomsten van de rangordening op onderwijsgebied. 'Politicians were treated very different by the press of the four countries', aldus Dixon e.a. (2013: 495). Hier en daar zitten parlementariërs ministers meer op de huid dan elders. Niet alleen de uitkomsten van de rankings maar ook de verschillen in politiek systeem en coalitieverhoudingen spelen hierbij een zekere rol. Verbeteringen in onderwijs als gevolg van de rankings kunnen overigens beter via een direct prikkelsysteem verlopen dat inwerkt op onderwijsinstuties dan via de lange weg van interventie van een minister, aldus Dixon e.a. (2013: 497).

Tenslotte hypothese 4: lokken rapportages een politiek-bestuurlijke response uit die beïnvloed wordt door de intensiteit en teneur van de behandeling van de onderwijsrankings in de pers? Deze hypothese gaat gedeeltelijk op.

De algemene conclusie is dat perspublicaties weliswaar niet zonder meer bepalend zijn, maar er wel toe kunnen doen (Tillema, 2006: 305). Ze beïnvloeden de publieke opinie en die maakt uit hoe politici reageren (Dixon e.a., 2013: 499). Tevens blijkt er *geen eenvoudig rechtstreeks* verband tussen resultaten van onderwijsrankings en politiek-bestuurlijke responses, maar de betekenis van perspublicaties moet zeker niet worden onderschat. De pogingen tot verbeteringen in het onderwijs zijn – op nationaal niveau - in relatie tot rapportages over rankings hoogstens *indirect* (Dixon e.a., 2013: 500).

Kijkend naar gemeentelijke benchmarks die in Nederland zijn uitgevoerd of naar het functioneren van provincies en waterschappen, valt op dat er geen substantieel onderzoek is verricht naar de betekenis van persmedia in lokale leerpraktijken en naar het eventuele verband met benutting van uitkomsten van benchmarking. Dat is wel begrijpelijk want de sterke indruk bestaat dat in de periode 1990-2014 weinig perspublicaties (in landelijke en regionale dag- en weekbladen) aan deze benchmarks zijn gewijd. Een kritisch forum ten aanzien van benchmarking ontbreekt vaak. In het voorgaande zijn al aanwijzingen verkregen waarom dat zo is. Als er maar weinig rapporten op de tafel van een gemeente- of waterschapsbestuur verschijnen met begrijpelijke conclusies en

verbeteraanwijzingen ontbreken, hoeft het niet te verbazen dat artikelen in dag- en weekbladen hierover heel schaars zijn. De pers wordt dan ook niet 'gevoed'. Omgekeerd, regionaal of lokaal opererende journalisten speuren niet nadrukkelijk naar benchmarkrapporten. En als ze er zijn, vragen ze ambtenaren of bestuurders niet wat de kern is en welke betekenis ze eraan toekennen.

11.8 Nogmaals: ontvankelijkheid voor benchmark bij bestuurders

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 politieke en ambtelijke cultuur gericht op reflectie is ook van belang (Boyne e.a., 2011). Dit zijn slechts enkele variabelen die inwerken op benutting of gebruik van bevindingen uit benchmarkrapporten.

Maakt de bestuurlijke functie ook uit? 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. Bestuurders lezen niet graag veel maar laten zich wel graag bijpraten over bestuurlijk relevante 'essenties'. Wethouders financiën tonen doorgaans iets meer ontvankelijkheid voor rapportages met impact op de begroting (uit de planning & control cyclus). Vaak wordt overschat dat bestuurders stukken grondig lezen of bestuderen. Dat valt in werkelijkheid nogal eens tegen. Zoals een wethouder (Beurskens) van een middelgrote stad (Venlo) 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, zeker als een vervolganalyse uitblijft, 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 – we stelden het al -, een dergelijk afzonderlijk, verbindend rapport.

11.9 Witte raven

Desalniettemin, bepaalde benchmarkrapporten worden hier en daar wel instrumenteel en direct gebruikt (lees: benut) 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 rechtmatigheid en 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 denken over leren naar de eenvoudige werkelijkheid, dan blijkt ook nog het volgende.

11.10 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 kringbijeenkomsten plaatsvinden over bevindingen nog voor een rapport uitkomt, zoals de genoemde gemeentelijke sportbenchmark uit hoofdstuk 7 laat zien.

Maar het hoofdbeeld geeft vooralsnog 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; Ter Bogt, 2005).

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 of ze zijn twijfelachtig;
- c) benchmarkrapporten 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 vervolgprioriteiten 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.

11.11 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, zeker als ze over een politiek gevoelige materie handelen, 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. college en gemeenteraad) over de kwaliteit van wat beoordeeld is. Als een rapport ook van een begeleidende oplegnotitie van – in dit voorbeeld - het college van 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 waarnemer 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 toelichting op enig kengetal. De politiek-bestuurlijke interesse in kengetallen was in deze gemeente van bijna 20.000 inwoners kortom volkomen nihil.

11.12 Benchmarks: geen bananenschillen voor bestuurders

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 – vanaf 1990 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 en/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; Hoogwout, 2010; Korsten, 2010b; Aardema, Korsten, Riezebos en Van Dam, 2011; Schoenmaker, 2011). Er is ook nog geen dijkgraaf gesneuveld als gevolg van benchmarking van waterschappen en evenmin een minister van justitie als gevolg van een vergelijking van prestaties van rechtbanken.

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).

11.13 Verklaring

'Benchlearning is een nog ver verwijderd ideaal'. Aldus de benchmarkspecialisten Jos Blank, Flóra Felsö 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, en zelden of nooit leidt tot fundamenteel leren, heeft verschillende redenen. Een aantal werd er al genoemd. We concentreren ons hier op de bestuurlijke benutting respectievelijk het leren in bestuurlijke kringen.

Een eerste reden is dat benchmarkrapporten soms een dagelijks bestuur, zoals een college van burgemeester en wethouders, wel bereiken maar dat daarin geen handvatten in zijn opgenomen over 'hoe nu verder?' De rapporten bevatten niet of nauwelijks (hanteerbare) suggesties of aanbevelingen over 'wat te doen?' 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. Dan is een dagelijks bestuur daar dus snel klaar mee: het benchmarkrapport wordt voor kennisgeving aangenomen en verder terzijde gelegd. Als het wordt doorgestuurd naar het algemeen bestuur, zoals een gemeenteraad vanwege de actieve informatieplicht, zal een oplegnotitie met duiding van de uitkomsten en een beleidsvoorstel vaak ontbreken. En de dan volgende reactie van raadsfracties uit een coalitie blijkt vervolgens vaak te zijn: 'als het college geen reden ziet om iets te doen met de uitkomsten, dan laten we het daarbij'. Benchmarkrapporten bevatten dus vaak geen (politieke) brandstof 'to make fire'. We komen daar verderop op terug als we spreken over analytische en politieke rationaliteit van de rapportages.

Een tweede reden voor geringe benutting van benchmarkrapporten in de bestuurlijke gremia verwijst opnieuw naar de inhoud van de rapporten: de verschillen in prestatiescores worden niet (begrijpelijk) verklaard. Het ontbreekt aan een verklaring voor de hoge, gemiddelde of lage prestatiescores van de organisaties die aan de benchmark deelnamen. Er is vaak ook geen aparte rapport per opdrachtgevende overheidsorganisatie beschikbaar (die de uitkomsten naar de specifieke organisatie toetrekt, de scoring begrijpelijk maakt en aanbevelingen geeft) terwijl dat volgens de benchmarkliteratuur wel een 'eis' is (Martens e.a., 2002: 39; Commissie-Van Rijn, 2007). Het zijn veelal aanbieders van benchmarks die aanbieden om een benchmarking uit te voeren en deelnemers aan een benchmark zoeken. Deelname van zeg vijftien gemeenten leidt dan tot één rapport, niet tot 15 afzonderlijke rapporten of 15 afzonderlijke rapporten en een overall-rapport.

Achter deze praktijk gaat schuil dat opdrachtgevers hun opdrachtgevende rol niet substantieel invullen. Onderzoekers activeren die opdrachtgeversrol ook niet want ze bieden een benchmark aan waarop een opdrachtgever kan intekenen. Als dan pakweg vijftien deelnemers gevonden zijn, kunnen de kosten gedeeld worden en in feite gedrukt worden. Doordat deze productbenadering het wint van de meer op interactie gerichte procesbenadering (De Bruijn en Van Helden, 2009) worden

bepaalde basiscondities voor latere benutting van uitkomsten uit benchmarkonderzoek al voor de start van een benchmarkonderzoek weggedrukt. Dit voert ons naar de derde reden.

Een derde reden voor zwak of afwezig gebruik van rapporten en gebrekkige 'benchlearning' in de bestuurlijke context is dat opdrachtgevende gemeentebesturen 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 vierde reden is dat colleges van B&W zelden veel weet hebben van de deelname van de gemeente aan een analyse van prestaties. De deelname is een zaak van een hamerslag in een collegevergadering 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 of zich een lerende organisatie te tonen en *zelden* het gevolg van een politieke prioriteit uitgesproken in een coalitieakkoord of geaccentueerd door een aangenomen raadsmotie. Er is rond dit onderwerp benchmarking dus weinig 'druk op de ketel'. Daarbij moet opgemerkt worden dat er uitzonderingen zijn. De publicatie van kwaliteitskaarten van onderwijsinstellingen door het dagblad Trouw werd vaak nauwlettend door schoolbestuurders en directeurs gevolgd vanuit het idee 'hoe staan wij ervoor?' Maar dan hebben media ook een rol gespeeld. Benchmarkrapporten over gemeentelijke prestaties halen zelden een krant en slechts af en toe een vakblad. Dat vergroot de druk dus niet.

Een vijfde 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 lagere score in het openbaar bestuur 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 zesde 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 van prestaties? 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 of zelden intensief.

De zevende reden voor gebrekkige 'benchlearning' is dat ambtenaren en/of bestuurders slechts beperkt ontvankelijk zijn voor 'leren' of 'verbeteren' tenzij er sprake is van een crisis of andere urgentie (Tillema, 2006). 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 benchmarkrapport 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).

Een achtste reden voor gebrekkige benchmarking betreft de analytische en politieke rationaliteit die in de benchmarkrapportages naar voren komen. Zelfs de betere benchmarkrapporten getuigen meer van analytische rationaliteit dan van politieke rationaliteit. Dat benchmarking niet echt past in de politiek- bestuurlijke rationaliteit van het algemeen bestuur van een gemeente hoeft niet te verbazen omdat ook rapporten van een lokale rekenkamer lang niet altijd passen, zoals Rolf Willemse (2009) voor Rotterdam vaststelde. Studies van de Rotterdamse rekenkamer omarmden over het algemeen een analytische werkwijze, die uitmondt 'in een logisch opgezet beleid'. In de ontwerpfase van beleid wordt dan, volgens die rekenkamerrapporten, gekeken naar doelen en middelen, in een logische samenhang. In de uitvoeringsfase wordt gezien of het beleid uitgevoerd wordt zoals het bedoeld is. En zo voort (Willemse, 2009: 30). Dat soort rekenkamerrapporten wordt in politieke contexten evenwel vaak gerelativeerd. 'Achterblijvende prestaties' worden in de politieke context 'aanloopproblemen' genoemd, aldus Willemse. En meetinstrumenten zijn niet altijd bedoeld om te meten maar zijn eerder 'een kwaliteitsinstrument om beelden te vatten'. Zo kan er ook gekeken worden naar benchmarkrapporten. De inhoudelijke relevantie van benchmarkrapporten wordt in politieke contexten blijkbaar makkelijk weggemasseerd. Vanuit de politieke rationaliteit gezien, ligt ook aan benchmarkrapporten teveel doelgerichtheid en instrumentaliteit ten grondslag en te weinig dat beleid een stolling van het haalbare is, een tijdelijk besluit of een politiek compromis.

Tenslotte, het komt ook voor dat de vermelde beoogde prestaties veelal nauwelijks beheersbaar zijn voor bestuurders en ambtenaren. De variabelen die verklaren zijn niet altijd ook geschikt om te manipuleren. En manipuleren impliceert een tijdpad. Het is niet mogelijk direct 'aan de knoppen te draaien'. In geval van 'governance' of netwerkbestuur wordt dit nog versterkt. Een gemeentebestuur heeft niet alle partners waarvan het bestuur afhankelijk is aan een touwtje (Van Helden & Johnsen, 2002; Ter Bogt, 2006: 172).

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) en niet goed aan verbeteringen in beleid. Toch gaat het benchmarken door. De mythes blijven bestaan.

11.14 Enkele conclusies

De gemeentelijke praktijken rond benchmarking zijn nog te versterken. Deze conclusie sluit aan op een bevinding van Ter Bogt uit 2005 (Ter Bogt, 2005: 68), Tillema (2006) en anderen.

Gemeentelijke benchmarking is een potentieel geschikt instrument voor bezinning op de kwaliteit van dat gemeentelijk handelen, dat zich leent voor kwantificering, vergelijking, rangordening en verbetering. Dit instrument wordt door bestuurders en ambtenaren ook (diffuus) positief benaderd. Desalniettemin zitten in de dagelijkse praktijk weinigen in Nederland, specifiek in gemeenten, echt op een benchmarkstudie en -rapport te wachten. De bestuurlijke prioriteit om benchmarking te starten is beperkt en de urgentie om rapporten te benutten bijgevolg ook. Instrumenteel gebruik van benchmarking is desondanks niet afwezig maar komt niet op grote schaal voor; we noemen het gebruik daarom beperkt (Ter Bogt, 2005; Tillema, 2006). College en politieke fracties vragen er slechts incidenteel naar en wanneer ze deze ontvangen, doen ze er weinig mee. De kansen op aan benchmarking gekoppelde sturing die men heeft, worden vaak niet benut. Dat komt ten dele ook door de gebrekkige kwaliteit van de benchmarkstudies en de onvolledige rapportages. Maar ook als gevolg van werking van de *wet van Kingdon*: de kans dat een onderwerp op de bestuurlijke agenda komt is heel groot als er een click ontstaat tussen een probleem, een oplossing en een politiek moment. Het verbaast dan ook niet dat een benchmarkrapport over bijvoorbeeld de afvalomvang blijkens verwijzingen soms pas twee jaar na verschijning opduikt in een beleidsvoorstel over afvalbeleid.

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.

Er bestaat een paradox: organisaties blijven aan benchmarking doen, ook al benutten ambtenaren en bestuurders de uitkomsten maar beperkt of niet. Waarom? De prijs is niet hoog. En als de samenleving verantwoording vraagt, dan krijgt men verantwoording. En van meting en verantwoording gaat de illusie van controle en beheersing uit.

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).

Ook als geen spoor van verbetermaatregelen als gevolg van benchmarking te ontdekken is, staat dit de populariteit van het instrument niet in de weg (Poerstamper e.a., 2007: 35). Blijven vergelijken is blijkaar toch uitdagend, ook al gooit het bestuur beleid niet om als gevolg van benchmarking (dat wil zeggen het uitblijven van 'double loop learning'). De nadruk op de ambitie van 'deugdelijk bestuur' en 'good governance' in huidige tijden en de hardnekkigheid van de mythes rond benchmarking maakt ook dat deelname aan benchmarking voortgaat. De bestuurlijke wereld kent zijn stilzwijgende ficties.

12 Valkuilen bij benchmarking

Voor we definitief concluderen dat gemeentelijke benchmarking vaak niet in alle deelnemende gemeenten brengt wat ervan verwacht wordt (alleen klasseringen; best practices niet opgespoord; verklaring van verschillen blijft mager of achterwege; leren zelden radicaal) nog iets over valkuilen. Immers, als valkuilen *niet* vermeden worden, versterkt dat het begrijpen van wat benchmarking in werkelijkheid vermag te betekenen. Dan moeten we een minpunt noteren. Maar als valkuilen juist wel vermeden wordt, draagt dit bij aan de kwaliteit van benchmarking

Het is mogelijk dat benchmarkonderzoekers dus in valkuilen lopen. Als dat gedrag in de toekomst vermeden worden, zou daarin een reden gelegen kunnen zijn voor een stap voorwaarts ten aanzien van benchmarking. Laten we zien. We onderscheiden twee typen valkuilen. Voor valkuilen op methodisch vlak wordt een aparte paragraaf gereserveerd.

12.1 Valkuilen bij benchmarkonderzoek op een rij

Benchmarkonderzoek vertoont nogal eens tekorten. Die zijn op te sporen door te bezien of benchmarking wel toegepast wordt 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 valkuilen bleken te traceren.

Fig.20: Valkuilen bij gemeentelijk benchmarkonderzoek

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.
12. De valkuil van het vergeten op overeenkomsten te letten.
13. De prestatieparadox als valkuil.
14. Myopia als valkuil.
15. De rechtvaardiging van zwakte.
16. De valkuil van de opgedreven prestatienormen.

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 Binnenlandse Zaken en Koninkrijksrelaties 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 wegenonderhoud, 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 bijvoorbeeld 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).

Het is ook maar de vraag of afgesproken indicatoren zullen bijdragen aan een verbetering van een verschijnsel zoals de veiligheid. Er kunnen meettechnische problemen opdoemen. Het is niet nuttig om voor de vaststelling van vergroting van de veiligheid in dorp of wijk alleen te kijken naar de politie omdat ook andere organisaties bijdragen aan (on)veiligheid, zoals het openbaar ministerie, stadswachten en organisaties voor hulpverlening. De waarneming dat de politie meer boeven vangt, kan teniet gedaan worden als het OM ze moeten laten gaan wegens cellengebrek of beperkte vervolgingscapaciteit (Volvaard, 2003).

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.

Bij het ontwerpen van 'De Staat van de Gemeente' (later: waarstaatjegemeente) is overigens getracht deze valkuil te vermijden.

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 beoordeelt, 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 afdekkingsgraad van een indicator (Bouckaert e.a., 2011: 123).

Dat betekent dat soms ten onrechte verondersteld wordt dat een indicator het hele (brede) 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

We schreven al dat benchmarkonderzoekers zelden rekening houden met de strategie en missie van organisaties die in een vergelijking betrokken worden (De Vries en Lako, 2009: 75). Is er echter wel sprake van zicht op concrete beleidsdoelstellingen? Ook niet. Een 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.

Over beleidsdoelstellingen of verschillen in doelstellingen tussen gemeenten lees je lang niet altijd in verslagen van benchmarks. Het gevolg hiervan is dat er niet zomaar te spreken valt 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 bereiken 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 op verbetermogelijkheden? Wordt in een rapport een handelingsperspectief aangereikt? 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 zelden iets omdat geen streefcijfers of een andere normering, of 'best practices' vermeld worden. De lezer van menig benchmarkrapport krijgt *geen idee of het totaal aan scores onder een kritische grens zit of juist er boven*. Het politiek bestuur van een deelnemende gemeente is in zo'n geval dus onthand en moet zelf tot een nadere analyse proberen over te gaan.

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 goede praktijk representeert terwijl dit niet juist is. We stootten op dit verschijnsel toen we deelnemers aan *waarstaatjegemeente* op vijf burgerrollen vergeleken. De scores lagen op een aantal rollen dicht tegen elkaar aan en waren 'voldoende' of hoger. De suggestie dat er grote kwalitatieve beleidssprongen 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 bepaald 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 plausibele 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.

12 De valkuil van het vergeten op overeenkomsten te letten

We noemen nog een valkuil. Die houdt in dat bij vergelijking wel op rangordening en het begrijpen van verschillen gelet wordt tussen de aan de benchmark deelnemende organisaties (rangorde en vaststelling van 'best practices') maar vergeten wordt het *totaal aan overeenkomsten* in beeld te nemen. Misschien wordt een rangorde gegeven en wordt geschetst welke gemeente de beste is maar wordt vergeten dat eigenlijk alle gemeenten al goed, matig of slecht functioneren. Dan heeft men dus teveel oog voor verschil terwijl de meest markante uitkomst juist de overeenkomst is in scores van alle organisaties samen. We lichten dit toe.

Aan benchmarkrapporten zijn vaak drie dimensies te onderkennen: een macro-, een meso- en een microdimensie. De macrodimensie verwijst naar de analyse van uitkomsten van alle vergelijkbare benchmarks op het niveau van een land of sector, de meso-uitkomsten naar de individuele onderzochte organisaties in één benchmark en de micro-analyse naar een onderdeel van de onderzochte en vergeleken organisaties. Stel er is sprake van 'De staat van de gemeente'. Dan levert een algehele analyse een rapport op met uitkomsten van alle gemeenten in Nederland die dit instrument hebben toegepast (macro). Deze analyse is ook verricht (Abma en Korsten, 2009; Abma, 2014). Dan worden alle benchmarks van steeds ongeveer vijftien tot twintig vergeleken gemeenten samen genomen en wordt een overall-beeld geschetst. Maar er is ook een analyse mogelijk van zeg de vijftien gemeenten die op een bepaald moment deelnamen en vergeleken zijn (de eigenlijke benchmark; meso) en er is de vergelijking op een deel mogelijk, zoals de beoordeling van het functioneren van gemeenten vanuit de rol van de kiezers op vertrouwen in het gemeentebestuur (micro). Waar nu op gelet kan worden, is op welke niveau de analyse wordt verricht. Het ministerie van BZK kan als systeemverantwoordelijke organisatie voor de bestuurlijke organisatie in het land juist aandacht hebben voor het totaal aan uitkomsten: dus voor de overeenkomsten op macroniveau en niet zozeer voor de verschillen op macro- of mesoniveau. Als een analyse plaatsvindt van rechtbanken in Nederland in vergelijking met drie andere landen, zal de interesse uitgaan naar de vraag of alle rechtbanken in Nederland samen het beter doen dan die in elk van de andere landen en waarom. Kortom, benchmarking laat verschillenanalyse toe maar evenzeer een overeenkomstenanalyse en soms is die laatste belangrijker. De kans bestaat om dit over het hoofd te zien in een drang tot rangordening.

13 De prestatieparadox als valkuil

Een bekende valkuil is de prestatieparadox (Van Thiel en Leeuw, 2003). Stel dat uit benchmarking komt dat op papier de gewenste doelen worden gehaald dan betekent dit nog niet zonder meer dat een gemeente ook goed bezig is. Om een voorbeeld uit de hulpverlening te geven: mogelijk toont een hulpverleningsorganisatie zich met registratie en het bereiken van makkelijk te halen doelen wel een goed boekhouder maar daarmee is ze nog niet vanzelfsprekend een goede hulpverlener aan

burgers. Het gevaar bestaat dat cijfers de blik afhouden van waar het werkelijk in de hulpverlening om gaat. Schijnsuccessen dreigen. Een goede organisatie moet de goede dingen goed doen (Tonkens, 2004: 49).

Prestatiemeting beoogt wel het opportunisme 'om makkelijk te scoren' buiten de deur te houden maar tegelijkertijd ligt dergelijk gedrag op de loer. Een gemeente die van een schouwburgdirectie een minimaal gemiddelde stoelbezetting van drievierde per zaal per voorstelling vraagt, lokt uit dat geen toneelvoorstellingen in de vorm van tragedies meer plaatsvinden, maar slechts komedies of optredens van populaire zangers en bands. Culturele verarming dreigt. Misschien leidt dit tot McDonaldization. De schouwburgen bieden evenals de fast-foodketen overal hetzelfde aan.

Een organisatie die meegaat in gemakkelijk te halen doelen, maakt zich schuldig aan 'afroming' (Tonkens, 2004: 50). Het vervolg ligt op de loer. Voor je het weet, doet een hele sector hetzelfde: insteken op makkelijke successen. Dan is sprake van wat Tonkens *successisme* noemt. Er wordt nog slechts over succes gesproken maar in feite is sprake van verkeerd leren ('wassen neus').

14 Myopia als valkuil

Bijziendheid (myopia) is een verschijnsel (ametropie) waarbij iemand voorwerpen ver weg niet scherp kan zien, maar wel nabij gelegen voorwerpen. Daarom spreken we van (dicht)bijziendheid. In de prestatiemetingsliteratuur wordt met myopia bedoeld dat er zoveel outputindicatoren de revue passeren dat belangrijke en onbelangrijke zaken niet meer uit elkaar gehouden kunnen worden. Men ziet door de bomen het bos niet meer. Het gevaar dreigt dat een discussie over een onderwerp dan leidt tot selectief winkelen om een vooruitgang te bewijzen of te bestrijden (Tonkens, 2004: 50). Dit gevaar kan bestaan bij benchmarking maar de kans daarop is doorgaans niet groot omdat er juist sprake is van een gering aantal indicatoren.

15 De rechtvaardiging van zwakte

Als een organisatie het slecht doet, is het mogelijk om een aantal naar verwachting nog slechter functionerende organisaties te selecteren en te vergelijken. Dan kan uit de vergelijking en rangorde als bevinding komen dat de je als organisatie de beste bent, maar dan wel de beste van de slechte. Zo wordt de benchmarking 'a justification of poor performance', een truc (Hakvoort & Klaassen, 2010: 116).

16 De valkuil van de opgedreven prestatienormen

De valkuil van de opgedreven prestatienormen houdt in: de mogelijkheid dat het vergelijken en rangordenen ertoe leidt dat de prestatienormen hoger gaan worden als de huidige normen gemakkelijk bereikt worden. Het effect dat betiteld kan worden als: '*Rankings drive up standards*' (Dixon e.a., 2013) Ook het omgekeerde is mogelijk. Als aanvankelijke normen te scherp waren en niet haalbaar bleken, kan een normverlaging optreden.

12.2 Wat nu van de genoemde valkuilen te denken?

We zetten ze nog even op een rij.

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.
12. De valkuil van het vergeten op overeenkomsten te letten.
13. De prestatieparadox als valkuil.
14. Myopia als valkuil.
15. De rechtvaardiging van zwakte.
16. De valkuil van de opgedreven prestatienormen.

Deze valkuilen zijn zonder meer serieus te nemen. We noemen in het bijzonder hier 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.

Het is niet uit te sluiten dat nog meer valkuilen in verder onderzoek opduiken. Een zo'n valkuil die nog niet genoemd is betreft de selectie van organisaties ten behoeve van de vergelijking.

12.3 Methodische valkuilen bij benchmarkonderzoek op een rij

Er bestaan naast de genoemde valkuilen 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), Bouckaert & Halligan (2008), 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 om 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:

Fig. 21: Enkele methodische valkuilen bij gemeentelijk benchmarkonderzoek

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.
10. De valkuil van de onvergelijkbare organisaties.

We lichten deze toe.

1 De valkuil van Pangloss

Deze ziekte of valkuil, de '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 van dat type 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 ook maar één opera van

Sydney. Zo is er ook maar een gemeente Súdwest-Fryslân. Er is geen equivalent voor deze gemeente met 69 kernen.

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 gecontroleerd moet uitbranden. Dan ligt de lat anders. 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 *gedetailleerde* 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 verkeersslachtoffers 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.

10 De valkuil van de onvergelijkbare organisaties

Het is bij benchmarking mogelijk dat organisaties uit een verschillend land of een verschillend bestel (institutioneel arrangement) vergeleken worden. Dan ontstaat onvergelijkbaarheid. Wie de winstgevendheid van luchtvaartmaatschappijen wil vergelijken, moet ze vergelijken op dezelfde vliegtuigen, dezelfde routes, dezelfde bestemmingen en dezelfde basisfaciliteiten. Wie denkt dat dit uiteraard gebeurt, vergist zich (Hakvoort & Klaassen, 2010: 133). Het onderzoek naar de vergelijking van rechtbanken in vier landen, eerder als voorbeeld opgevoerd, is een treffende illustratie van deze valkuil (Dijkstra en Toonen, 2004; Blank e. a., 2004). Als deze valkuil niet vermeden wordt, heeft dit tot gevolg dat de verkeerde best practices worden aangewezen en aanbevelingen tot verbetering in feite niet te geven zijn.

12.4 Valkuilen vermijden in ontwerpprocessen van meting

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 en publiceren erover. Ook in de wereld van corporaties is in het rapport ‘*Beter meten*’ wel aandacht hiervoor gevraagd (Commissie-Van Rijn, 2007). Er is een commissie Benchmarking ingesteld die een leidraad voor benchmarkaanbieders opstelde. 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 zoals raden van toezicht die benchmarks af en toe onder ogen krijgen, bemoeien zich doorgaans 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; Minderman, 2007; Boers en Van Montfort, 2009; 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 en andere (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 maar niet expliciteren of het hen gelukt is) 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 en die elke positieve uitkomst verwelkomen maar die de methode waarop die uitkomst tot stand kwam voor lief nemen). Het bestaan van deze gescheiden werelden stemt niet optimistisch vanuit het oogmerk van kwaliteitszorg rond benchmarking.

Het lijkt in het algemeen gewenst de genoemde valkuilen zoveel mogelijk te vermijden. Is een nuloptie van ‘elke valkuil moet vermeden worden’ een haalbare kaart? Sommigen menen dat prestatiemeting door benchmarking ondanks dreigende valkuilen of feitelijke perverse effecten toch doorgaat vanwege ‘de onmiskenbaar positieve effecten’ (De Bruijn en Van Helden, 2007: 20). Benchmarking is van belang als ‘kwaliteitszorg’ hoog op de agenda staat. Benchmarking kan heel concreet tal van bespreekpunten opleveren.

13 Enkele conclusies over de keizer en de kleren

13.1 Benchmarking en New Public Management

Doet het streven naar een zakelijk opererende bedrijfsmatig werkende overheid, dus 'New Public Management', ertoe? Deze vraag stelden Gow & Dufour in 2000 en rond die tijd ook velen in Nederland. Wat is het antwoord in relatie tot benchmarking?

'New Public Management' (NPM) heeft sinds een aantal decennia toepassing en doorwerking gevonden bij Nederlandse overheden (Ter Bogt, 2005; Pestman, 2007; Ter Bogt e.a., 2010; Van Thiel, 2009; Pollitt e.a., 2010; Noordhoek, 2013). Die NPM-beweging uit zich onder meer in toegenomen aandacht voor ambtelijk en bestuurlijk ondernemerschap of -voorzichtiger – initiatieven, een naar buiten gerichte organisatiecultuur, outputsturing, prestatie meting, prestatieafspraken, pogingen tot beheersing en verbetering. Die aandacht leidde niet altijd tot succes. Zo maakten Nederland kennis met de Ceteco-affaire, die zich in de provincie Zuid-Holland voltrok. Een ambtenaar kreeg veel ruimte om van geld geld te maken, bleek aanvankelijk zeer succesvol, werd daarover 'over zijn bol gestreken' maar uiteindelijk bleek sprake van onvoldoende verantwoording en controle. Foute zaak dus. Deze affaire pakte niet goed uit voor de betrokken ambtenaar en kostte enkele provinciebestuurders de kop. Wat is er in het kader van deze studie te zeggen over NPM: een opgewekter beeld?

Benchmarking, oorspronkelijk ook wel bedrijfsvergelijking genoemd, past ontegenzeggelijk als ambitie in de ontwikkeling naar meer 'zakelijkheid' en 'bedrijfsmatig werken'. Waarom? Omdat het in potentie en qua achterliggende ambitie, althans volgens de literatuur, een onderzoeksmethode is voor interne of externe prestatievergelijking tussen organisaties of onderdelen of aspecten daarvan, gericht op het vinden en verklaren van 'best practices' en uiteindelijk op verbeteren van praktijken in organisaties. Benchmarking sluit als onderzoeksmethode aan op door NPM geïnspireerd 'performance management'. Oké, het benchmarken past bij NPM maar werkt het ook zo uit?

13.2 Met benchmarking beleidshomeopathie uitbannen

Benchmarking past niet alleen bij een beweging naar meer bedrijfsmatig werken, waarbij methoden en technieken van bedrijfsanalyse uit de private sector worden overgenomen in de publieke sector, maar ook in een streven naar 'Evidence Based Policy' (EBP; Pawson, 2006; Leeuw, 2008; Coolsma, 2009) en naar 'Evidence Based Public Management' (EBPM; Pfeffer & Sutton, 2006; De Groot, 2010; Shillabeer e.a., 2011; Rousseau, 2006, 2012). Dat betekent dat de verwachtingen over benchmarking nog groter worden. De beweging naar zoveel mogelijk op degelijk bewijs gebaseerd beleid wordt gemaakt omdat teveel managementinterventies te 'dun' of 'zwak' gefundeerd zijn, bij voorbeeld op basis van persoonlijke observaties, selectieve voorbeelden of beperkt generaliseerbare onderzoeksresultaten (De Groot, 2010: 21). Er is teveel sprake van sprongen in het duister volgens de verdedigers van de wenselijkheid van 'evidence-based public management'.

Hans de Groot (2010) legt in zijn Twentse oratie 'Evidence-Based Public Management' expliciet die link tussen benchmarking en op bewijs gebaseerd publiek management. Er is volgens hem behoefte aan beproefde aanpakken; meer dan aan theoretische blauwdrukken of andere interventiepogingen die in de praktijk niet blijken te werken. Benchmarking is aantrekkelijk voor bestuurders want deze methode levert zicht op beproefde aanpakken, op wat werkt (De Groot, 2010: 23). Tenminste, dat denkt hij. Waarom die relatie?

Omdat de bedoeling van benchmarking is om door a) vergelijking van organisaties (producten, diensten of processen), b) rangordening van organisaties van goed naar slecht, en c) van analyse van het waarom van de rangorde en wie eerste is, en d) van uitleg van wat waarom 'best practice' is te streven naar op evidentie gebaseerde verbetering van die producten, diensten of processen. Benchmarking betekent in de praktijk dat een onderzoeker of een team onderzoekers in feite namens de deelnemende organisaties in andermans keuken gaat kijken om te zien of en zo ja, wat

daar een prima praktijk is. Evidentie houdt in dat er *overtuigend bewijs* moet zijn dat iets in een bepaald opzicht 'werkt', dus bij voorbeeld effectief of doelmatig is of goed in ander opzicht. Denk in geval van universitaire opleidingen bij voorbeeld aan de instelling die de beste is op het gebied van de organisatie van contacten met opleidingsstudenten in opleiding X, bij voorbeeld qua tijdige melding van roosters van colleges en collegezalen of van tentamenuitslagen.

De veronderstelling achter benchmarking is dat goede of zelfs excellente 'bewezen' praktijken te transfereren zijn van de ene organisatie naar de andere. Dit is ook wel de 'copy & paste'-aanpak genoemd, dus van knippen en plakken of - anders gesteld - van imitatie. Voor de organisatie van een opleiding is de transfereerbaarheid nog groot. Hier kan kopiëren nog. De ambities rond benchmarking blijven echter niet steken bij dit soort simpele uitkomsten. Vanuit een streven naar evidentie in beleid en management, waarin benchmarking past, wordt ook gepoogd *beleidshomeopathie* uit te bannen (Leeuw, 2008). De kans dat beleid ontwerpen of uitvoeren dan 'een slag in de lucht wordt' of 'besturen op de tast wordt', zou door evaluatieonderzoek en benchmarking verminderen; niet alleen door benchmarking maar ook door benchmarking. Dat betekent dat onderzoek in de justitiële sfeer bij voorbeeld duidelijk moet maken hoe de reïntegratie van tbs-ers in de samenleving verloopt: heeft stap voor stap gedurende korte tijd weer 'aanwezig zijn' in de samenleving, met bepaalde begeleidings- of meldingsvormen, het gewenste effect? Of is dit niet meer dan een beredeneerde inschatting, een 'best possible guess'? En dat onderzoek naar enkele tientallen krachtwijken ('Vogelaarwijken') duidelijk moet maken of de leefbaarheid in deze wijken ook toeneemt, zoals de bedoeling was.

Volgens EBP en EBPM moet korte metten gemaakt worden met slechte praktijken (in de vorm van zwakke prestaties), met ongefundeerd beleid, met beleid 'op de tast', en met beleid dat een sprong in een onbekende diepte is. Dus er moet afscheid genomen worden van beleid dat op zwakke of ronduit verkeerde veronderstellingen berust, en met ondoelmatigheid of verspilling gepaard gaat (De Groot, 2004). Die prestaties moeten plaats maken voor evidente effectieve(re) en doelmatige(re) beleidspraktijken, die te ontleen zijn aan 'best practices'. Opsporing en kennisname van goede praktijken zou dus bijdragen aan de degelijkheid van het beleidsfundament. Goede praktijken zijn praktijken die zich positief bewezen hebben. En benchmarking moet die aan het licht brengen.

Wetenschappers als Meier en O'Toole jr (2009) en De Groot (2010) willen via grote databestanden van gemeenten en andere organisaties analyses verrichten en bijdragen leveren aan op evidentie gebaseerd beleid. Anderen, zoals Nelen (2008), zijn daarover sceptisch. Ze geloven niet zo in de transfereerbaarheid van 'best practices' en wijzen steeds op een verschil in juridische, organisatorische, culturele en politieke context die een 'pijnloze' overname van praktijken bemoeilijkt. Beleid maken en uitvoeren blijft mensenwerk in een bepaalde situatie en bepaalde tijd. Hun argumenten zijn relevant maar dan nog is het mogelijk om te vergelijken en te bezien wat inwerkt op verschillen in prestaties en wat beste praktijken zijn en waarom. Dat inzicht geeft indicaties voor de knoppen waaraan managers kunnen draaien of beter niet kunnen draaien, of netter gesteld wat ze bij het (her)ontwerpen van beleid als beïnvloedbaar zien.

Zie hier de bredere verwachtingsvolle situering van benchmarking in NMP en EBP/EBPM. De vraag was en is of benchmarking die ambities en aanspraken waar kon maken. De pretenties waarmaken, zou impliceren dat 'de keizer kleren aan heeft'. Maar dan moeten ambtenaren en bestuurders wel gevoelig zijn voor 'bewijs'.

De pretenties zijn in deze beschouwing nader uitgewerkt in de formulering van enkele mythen, dat wil zeggen in de vorm van 'hardnekkige' verwachtingen over benchmarking.

13.3 De ambitie en werkelijkheid van benchmarking: een demasqué

Benchmarking is – zoals gesteld - een potentieel interessant instrument voor vergelijking, rangordening en het aanwijzen van goede praktijken, analyse en verklaring van verschillen, en een

manier van verantwoording en leren in de publieke sector maar beantwoordt volgens deze studie nog *niet* lang aan een aantal verwachtingen (of hypothesen). De ambities zijn vaak te groot in verhouding tot de georganiseerde dagelijkse benchmarkpraktijken. Anders gesteld, de benchmarkpraktijken blijven in de praktijk ver achter bij de ambities. Waarom? Benchmarking blijkt vaak niet meer dan het *opmaken van een klassement* (lees vergelijking en rangordening). Een analyse van het waarom van de verschillen in het klassement ontbreekt en het overtuigend aanwijzen van een goede praktijk en hoe je uit moet komen bij die goede praktijk al helemaal. Dat geeft een gevoel van onbehagen. Want als je weet dat bij voorbeeld PSV voetbalkampioen werd en daarmee hoogste in een klassement, ken je het geheim van de (voetbal)smid nog niet. En daar gaat het juist wel om. Wat zou de club die op de vijfde plaats eindigde moeten doen, kijkend naar de kwaliteiten van PSV, om het volgend jaar *zelf* kampioen te worden: andere trainings- en voedingsmethoden toepassen, de jeugdopleiding verder ontwikkelen zoals PSV deed (aanname), een betere spits kopen, een betere trainer aanstellen, zorgen voor betere balans in het elftal, voor elke basisplaats een goede reserve in de selectie opnemen en de concurrentie om de opstelling bevorderen, het budget voor transfergelden verruimen, altijd zorgen dat je wint van een subtopper, hopen op wat geluk, ja wat?

Ergo, wat onder de vlag van benchmarking vaart, zoals 'de beste woongemeente' of 'de sociale staat van de gemeente', biedt in veel gevallen op zijn best een klassement. Wie het klassement kent, heeft zijn nieuwsgierigheid kunnen bevredigen maar weet nog niet hoe te handelen om zelf eerste te worden. Veel benchmarkrapporten bieden geen handelingsperspectief omdat de analyse van de 'best practice' ontbreekt. Een gemeentebestuur weet dus niet wat het hoe moet verbeteren. Dus van het overnemen van een goede praktijk komt vaak niets terecht. De context doet er wel degelijk toe.

Dit onderzoek toont dat deze conclusie geldt voor veel benchmarkonderzoek en zeker ook voor benchmarking in de sfeer van het lokaal bestuur. Dat impliceert kritiek op NPM-denkbeelden. Blijkbaar is een succesvolle praktijk van benchmarking in de private sector niet zo maar succesvol in de publieke sector. En ook een succesvolle praktijk in pakweg Veendam is niet zo maar over te nemen in Bergen op Zoom, Vlissingen of Zaandam. Benchmarking paste in NPM en als benchmarking niet brengt wat de bedoeling is, dan is dat zo gezien ook een kritische evaluatie van NPM.

Benchmarking leidt nog niet zo maar 'evidence-based policy', dus naar beleid dat zich bewezen heeft. De analyse van *waarstaatjegemeente*, zoals hier gepresenteerd is, toonde dat in sommige opzichten geen beste praktijk te vinden is. De verschillen in rangordening waren heel gering. Wat dan aan te bevelen en te transfereren naar een andere organisatie? Als je de beste praktijken wel vindt, blijken nastrevenswaardige beleidspraktijken bovendien niet zo maar over te hevelen van de ene gemeente naar een andere; nogmaals, de context is ook belangrijk. De Maastrichtse hoogleraar Nelen heeft dus wel een punt. Dat was ook al gebleken toen jaren terug het Tilburgs model van 'goed bestuur' (via contractmanagement) in de mode raakte. Bestuurders van kleinere gemeenten namen dat concept niet over. Zoals (oud-) burgemeester van de gemeente Voerendaal Huub Strous een keer tegenover ons opmerkte:

'De ambtenaren worden hier door het gemeentebestuur en de ambtelijke top rechtstreeks aangestuurd, daar hebben we geen contractmanagement en decentrale vrijheidsgraden voor nodig. Aan een uitwerking van het Tilburgs model naar allerlei diensten bestaat bij ons geen behoefte. We bereiken al heel wat als we de ambtenaren hier in de goede richting kunnen sturen en dezen ook leveren wat de bedoeling is'.

Hoe komt dat, die teleurstelling? De kwaliteit van het benchmarkonderzoek is in de gemeentelijke praktijk maar ook daarbuiten lang niet altijd degelijk genoeg is om bewijs te leveren voor een nastrevenswaardige praktijk. Als Ajax voetbalkampioen wordt in de hoogste klasse van het betaald voetbal weten trainers, managers en bestuurders van andere clubs nog niet direct wat te doen om ook kampioen te worden, ook al is makkelijk waar te nemen dat de verdediging moeilijk passeerbaar bleek en de voorhoede vaak goed 'op schot' was. Als een gemeentebestuur de doorlooptijd voor afhandeling van een bijstands aanvragen in complexe situaties de helft korter inricht, moet de

gebruiker van de benchmark wel vernemen hoe dat kan. Telt het inzetten van maatschappelijk werkers dan misschien niet mee?

Niet altijd zal een before-after model van onderzoek met een experimentele groep (die de interventieprikkel ontvangt) en controlegroep (die de interventieprikkel niet ontvangt) nodig zijn, er zijn ook tal van andere onderzoeksopzetten, maar *waarstaatjegemeente* brengt toch als onderzoeksmethodiek niet naar voren dat de beste praktijk ook zeker de beste is, waaruit die bij nader inzien bestaat en waarom die de beste is. Op toepassing gerichte vergelijkende studies als benchmarks blijven teveel hangen bij het opstellen van een klassement per indicator of een spinnenwebprofiel. Ze komen dat *klassementsdenken* nauwelijks of niet voorbij. Benchmarks kennen dus te weinig diepgaande verklarende analyse en prescriptie. Het zijn halfproducten. Er moet eigenlijk na binnenkomst van de benchmarkresultaten steeds een vervolganalyse gemaakt worden. Voorzien de bij benchmarks georganiseerde gesprekskringen hierin? Blijkbaar niet genoeg. De bevindingen uit dergelijke kringen vervliegen vaak.

Met benchmarking valt meer te bereiken dan tot nu toe bleek en blijkt. Dan moet wel rekening worden gehouden met aan benchmarking in de publieke sector te stellen eisen. Voorkomen moet worden in de valkuilen lopen. Een van die eisen is dat als er – als uitkomst van vergelijkend onderzoeken - sprake is van '(evidence-based) best practices' de benchmarkonderzoeker die ook echt als 'bewezen' praktijken in de rapportage naar voren moet brengen en dat op een inzichtelijke (en verklarende) wijze moet doen. Dat gebeurt blijkbaar – uitzonderingen daargelaten – niet steeds voldoende. De beste praktijken worden menigmaal niet eens in rapporten vermeld. En niet zelden, als dat wel het geval is, ook niet inzichtelijk genoeg. Het zijn de gebruikers van benchmarkrapporten die zelf achter de beste praktijken aan moeten. Ze moeten op pad om onder meer te gaan spreken met actoren die opereren in die bewezen praktijk. En dat laten ze ook nogal eens na; beter gesteld, heel vaak na.

13.4 De beste praktijk en de beste organisatie

Wat hier parten speelt is, dat een beste praktijk in werkelijkheid vaak in benchmarkrapporten ontbreekt of, als die niet ontbreekt, een verzameling is van dit aspect of element (volgens de rangorde) dat goed is bij organisatie A, dat aspect of element dat goed is bij organisatie B en dus bij C en zo bij D, en zo voorts. Dus een 'hele' organisatie die een beste organisatie is, is blijkbaar niet (zonder meer) te vinden. De beste organisatie is hooguit een *virtuele* organisatie, een fictie die op papier uitgetekend is. Dit is eerder ook al eens vanuit een analyse van bestuurskrachtrapporten van gemeenten vastgesteld (Korsten en Abma, 2009; Korsten, 2010a; Abma, 2012). Maar daar staat tegenover dat ook vanuit een virtuele organisatie te onderkennen is of er verbeterruimte bestaat.

Zijn er meer argumenten te geven die verklaren waarom een beste praktijk in een rapport ontbreekt?

Een eerste argument is: wie het beste is, is dat misschien maar heel even. De gemeente Tilburg heeft met haar Tilburgs model van contractmanagement wel gefungeerd als referentiepunt maar dat was maar tijdelijk. Bestuur en ambtenaren van de gemeente Tilburg kwamen zelf terug op 'het Tilburgs model'. Na enige tijd was Tilburg niet meer het 'Mekka van het Nederlands openbaar bestuur' (Korsten, 1995).

Een tweede argument houdt in dat de perfecte organisatie nergens te vinden is. Benchmarkstudies in de sfeer van thuiszorgorganisaties, verzorgings- en verpleeghuizen, onderwijsorganisaties voor middelbaar beroepsonderwijs en analyses van visitaties in de woningcorporatiesector maken duidelijk dat *de perfecte organisatie* niet bestaat (Van Bortel e.a., 2012; Poerstamper e.a., 2007). Er is altijd wel iets dat minder functioneert. Naast sterkten blijken er altijd zwakten (Abma, 2012; Schutgens, 2014).

Kijken naar de top vijf van positief er positief uit springende organisaties of de 'bottom five' qua presteren, is overigens wel mogelijk. Het valt te proberen. In de woningcorporatiesector blijkt dan

dat de organisatiebesturing ertoe doet om tot de top vijf te kunnen behoren, alsmede 'de nadruk op willen leren en verbeteren en communicatie met belanghebbenden' (Van Bortel e. a., 2012: 78). Deze conclusies komen grotendeels overeen met conclusies uit *'Het geheim van de goede corporatie'* (De Klerk, 2012).

13.5 Voor- en nadelen van benchmarking

Benchmarking is een instrument om - via hoofdzakelijk kwantitatieve prestatievergelijking tussen organisaties (of delen daarvan zoals productclusters, diensten of processen) en rangordening van prestatiescores - , uiteindelijk te zoeken naar goede of 'beste' praktijken, deze begrijpelijk te maken en mede op basis daarvan te komen tot verbeteringen. Benchmarking richt zich op prestaties (outputs) en daarmee is het een onderzoeksapproach die past binnen prestatiebesturing en -management. In de kern gaat het bij benchmarking natuurlijk om de vraag of een overheid de gewenste (beleids)effecten (outcomes) bereikt.

Benchmarking heeft als praktische en op toepassing gerichte onderzoeksapproach (verwachte) voor- en nadelen. Ze zijn eerder op een rij gezet.

Tot de aan benchmarking toegedichte voordelen behoort dat het in oorsprong en potentie een vernieuwende approach is waarvan organisaties en medewerkers inspiratie en energie kunnen krijgen.

Er is hierbij ook sprake van een gerichtheid op verantwoording en verbetering. Gemeentebesturen laten door deel te nemen aan een benchmarkonderzoek zien wat bereikt is en waar de gemeente staat ten opzichte van andere gemeenten. Verantwoording is echter bijna nooit het enige of eerste doel van benchmarking, er is altijd een 'meer' dat benchmarking rechtvaardigt. Daar staat tegenover dat burgers vaak niet of slechts heel moeizaam een benchmarkrapport te pakken kunnen krijgen. Een fanclub van betrokken burgers heeft benchmarking dan ook zelden.

Benchmarking is vervolgens ook een basis, een vertrekpunt om met elkaar in gesprek te gaan over bevindingen. Dat betekent dat benchmarking in cognitief en interactief opzicht betekenis heeft.

Benchmarking kan ook betekenis hebben in een interbestuurlijke context. Wie van doen heeft met verpleeghuizen kan op basis van benchmarks stellen dat er voor een aantal verpleeghuizen verbeterruimte te vinden moet zijn omdat andere organisaties kwaliteit leveren bij inzet van zus en zoveel (minder) middelen. De voor verpleegzorg verantwoordelijke staatssecretaris Van Rijn (kabinet-Rutte II) deed dat dan ook.

Het is dus – samenvattend - een signaleringsinstrument voor het opsporen van zwakke plekken en sterke plekken in organisaties, en van producten, diensten en processen, en is daarmee een opstap voor bezinning en discussie daarover. Praktisch gezien biedt benchmarking mogelijkheden om te verbeteren door elders te gaan kijken, dus zonder zelf te proberen om het wiel opnieuw uit te vinden.

Benchmarking past zo gezien in een het concept van een 'lerende organisatie' en sluit ook aan op de *'Code voor goed openbaar bestuur'* uit 2009. Minister Guusje ter Horst en staatssecretaris Ank Bijleveld hebben de kernprincipes voor goed bestuur geformuleerd.

Tot de nadelen van benchmarking rekenen we dat dit instrument, zeker als het gaat om gemeentelijke benchmarking, meer de burgers als klant in relatie tot een organisatie in beeld brengt dan burgers als kiezer, participant, partner in samenwerking, wijkbewoner of onderdaan (cf Hiemstra, 2003: 78; vergelijk ook: Bogumil & Kuhlmann, 2006:362). Bijgevolg kunnen we bijna alle benchmarks betitelen als 'partieel'. Dat betekent dat ze een deel van de gemeentelijke opgaven en taken in beeld brengen. Denk maar aan een benchmark over rioleringen. Een uitzondering hierop vormt *waarstaatjegemeente*, een integraal benchmarkinstrument dat uitspraken pretendeert te

geven over de staat van het gehele gemeentelijk taakgebied. Deze benchmark beziet de kwaliteit van gemeentebesturen en hun organisaties vanuit vijf burgerrollen (Abma, 2014).

Een ander nadeel is dat niet alleen niet alle gemeentelijke processen, diensten of producties aan benchmarking worden onderworpen (kent u een benchmark daklozenbeleid of coffeeshopbeleid?) maar vooral datgene wat zich 'kwantitatief laat vangen', dus in getallen wordt gevat.

Benchmarking heeft nog een ander nadeel, namelijk dat het in de praktijk over het algemeen niet netwerkproductiviteit in beeld brengt terwijl gemeenten voortaan van taken en opgaven juist wel afhankelijk zijn van medeoverheden en andere actoren, zoals voor arbeidsmarktproblemen (bestrijding van jeugdwerkloosheid), en voor de ontwikkeling en het beheer van bedrijventerreinen (De Rynck, Voets en Van Dooren, 2010; Noordhoek, 2013). Daarom kan de volgende stelling in een rapport als *'Rotterdam vergelijkenderwijs'* (2002: 2) niet onderbouwd worden. De stelling luidt:

'Alle processen en functies van een organisatie kunnen in beginsel in aanmerking komen voor een benchmarkonderzoek'.

Die stelling gaat uit van de misvatting dat alle prestaties op alle taakgebieden zich volledig in getallen laten vangen. We kennen evenwel geen benchmarks waarin werkelijk alle gemeentelijke opgaven en taken worden meegenomen (zie ook bijlage 2). Ook *waarstaatjegemeente* kent grenzen (Aardema en Korsten, 2005; Abma en Korsten, 2009; Abma, 2014). Logisch, niet alles is kwantificeerbaar. En de stelling gaat er ten onrechte van uit dat het gemeentebestuur van Rotterdam op alle gebieden alle touwtjes in handen heeft. In werkelijkheid functioneert de gemeente Rotterdam in bestuurlijke netwerken als de Veiligheidsregio. Prestaties van interorganisationele verbanden als beleidsnetwerken ('governance') zijn wel te benchmarken, maar daar moet dan wel speciale aandacht aan worden besteed. Governance-gerichte benchmarks, waarbij van elkaar afhankelijke partners samen zoeken naar prestatieverbetering door governance-benchmarks zijn echter niet aangetroffen.

Een nadeel van benchmarking is dat het niet goed past in de 'ongekende samenleving' (Van Gunsteren, 1994; Van Gunsteren en Van Ruyven, 1995). Bepaalde auteurs zijn van mening dat de samenleving en ook het openbaar bestuur zich niet helemaal laten kennen en dus niet helemaal te vangen zijn in statistieken en kengetallen. De ene gemeente is anders dan de andere, ook al lijken ze op het eerste en tweede gezicht erg op elkaar.

We stellen vast dat benchmarking in de publieke sector niet onmogelijk is maar wel geremd wordt door de eigenheid van de publieke sector, zoals de meervoudigheid van politiek-bestuurlijke doelstellingen en door de complexiteit van producten, diensten en processen en de beoordeling met andere maatstaven dan prijzen of kosten en baten. Door de toepassing van benchmarking kunnen ook neveneffecten of averechtse effecten optreden, zoals een verwaarlozing van niet-meetbare taken en een jacht op gemakkelijke productie (de politie jaagt op bekeuringen?), of van productie van producten waarvoor in interbestuurlijk verband financiering wordt verkregen.

Een voorbeeld van dit laatste. Als gemeentebesturen (nagenoeg) volledige financiering krijgen van de rijksoverheid voor het aanleggen van verkeersdrempels en andere mogelijke maatregelen, zoals wegversmallingen, ter vergroting van de verkeersveiligheid zullen de verkeersdrempels als paddenstoelen uit de grond schieten. Aannemend dat de rijksoverheid geen enorme stapel verantwoordingen vraagt en de bureaucratiekosten hoog zijn.

Tweede voorbeeld. Als universiteiten financiële incentives ontvangen per goedgekeurde dissertatie, zullen universitaire en facultaire bestuurders in tijden van teruglopende middelen ook de effectuering van dissertaties gaan begroten en zullen ze promotoren (eventueel subtiel) aanzetten om voort te maken met de realisatie van een proefschriftenstroom. Dan ontstaat er zelfs discussie of de omvang van proefschriften niet beperkt moet worden onder het motto: 'dunner is sneller klaar'. En er ontstaat

wellicht een systeem van voortgangsmelding onder de vlag: bericht eens hoe het met de voortgang ervoor staat.

Wie de grenzen van benchmarking kent, kan echter wel dergelijk benchmarkonderzoek opzetten en voordelen behalen.

13.6 Gedijen

Als niet alle taakgebieden zich eenvoudig laten 'vangen' in een benchmark, en dus sprake is van grenzen aan benchmarking, waar kan partiële of integrale benchmarking dan in feite wel het beste gedijen? Benchmarking is vooral te richten op de prestaties van machinebureaucratieën (een organisatietype uit de typering van Mintzberg), dus op grote productiegerichte organisaties met min of meer gestandaardiseerde operationele processen, die zich voltrekken 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 of verstrekking van paspoorten.

Benchmarking is als manier van prestatiemeting *minder* geschikt voor het opsporen en beoordelen van prestaties van co-creatie of coproductie, van onderhandelingsdeals, van 'ad hocracies', 'high performance organizations' of 'zero fault organizations', politieke partijen of kerkelijke organisaties. Immers, die prestaties zijn vaak uniek, en/of ze moeten maximaal zijn (denk aan een kernenergiecentrale), en/of ze vereisen veel creativiteit, en/of er is sprake van geheimhouding, en/of de prestaties zijn niet te kwantificeren. Of zo een organisatie (geheime dienst?) wil niet beoordeeld worden in termen van producties en prestaties (Wilson, 1989; Mintzberg, 2006).

Benchmarkonderzoek is weliswaar gericht op verantwoording naar bestuur en samenleving maar in werkelijkheid worden benchmarks niet vaak door burgers geraadpleegd. Dat heeft ermee te maken dat de rapportages nogal eens technisch zijn. Ze worden doorgaans ook niet actief richting burgers 'op transport gezet'. Het blijkt ook in tijden van internet en google moeilijk om gemeentelijke benchmarkrapporten te verkrijgen. Dat is ook wel begrijpelijk omdat we vaststelden dat veel benchmarkrapporten niet eenvoudig leesbaar zijn en de analyse over verschillen in scores vaak sober of afwezig is.

Tal van aspecten van benchmarking zijn behandeld, maar welke bevindingen levert onze analyse over de mythen rond benchmarking op? 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. We lichten dit toe.

13.7 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, ook al pleitte de commissie-Van Rijn (2007) hier nog voor in verband met benchmarking van woningcorporaties. Soms is overigens de hele populatie van organisaties in een benchmark betrokken, zoals voor mbo-instellingen geldt; dan valt er ook weinig tot niets te selecteren. Maar dat is een uitzondering.

De belangrijkste reden voor de constatering van een afwijking van de mythe is dus de sterke aanboddrevenheid bij benchmarkaanbieders. Dit is ook wel de productbenadering genoemd. De opdrachtcontext is in interactief opzicht matig ontwikkeld. Waterschappen en gemeenten nemen heel vaak deel aan benchmarking door aanmelding en inschrijving.

Benchmarking blijkt in het openbaar bestuur desondanks wel een in zekere mate interactief proces te zijn omdat de individuele deelnemers prestatiegegevens verzamelen in het kader van een aanpak zoals het onderzoeksbureau die met benchmarking voor ogen staat. De uitkomsten worden vervolgens weer centraal verwerkt door het extern onderzoeksbureau. De uitkomsten in de vorm van tabellen en grafieken worden daarna geanalyseerd en er wordt over gerapporteerd, maar soms worden de tabellen en grafieken ook aan de deelnemers ter beschikking gesteld, zodat deze de resultaten van de 'rangordening' zelf verder kunnen analyseren. Van dat laatste komt vaak weinig terecht, bij voorbeeld omdat de politieke prioriteit van het onderzochte proces of product of de onderzochte dienstverlening niet groot genoeg is. Benchmarkrapporten hebben dus soms iets weg van halfproducten; ze zijn niet af.

13.8 Tweede mythe

De tweede mythe die we over benchmarking formuleerden, luidde: *'Benchmarking leidt tot het opsporen van "best practices" en bezinning daarop in benchmarkrapportages'*.

Op grond van onze huidige inzichten wordt de stelling niet bevestigd. Daarvoor zijn enige argumenten aan te voeren.

1) In veel benchmarkrapporten komt een aanduiding van beste (dus excellente of goede) praktijken helemaal *niet* expliciet voor. Maar in een beperkt aantal dus wel.

Meergenoemde benchmark over de geestelijke gezondheidszorg (Riagg 's) kende geen beste praktijken en de raad van toezicht waarin een van ons participeerde, 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.

2) Als die aanduiding van een goede of beste praktijk in benchmarkrapporten wel voorkomt, worden 'best practices' later vaak toch niet overgenomen. Dat gebeurt niet en kan ook niet goed gebeuren omdat de interne context (kenmerken van de organisatie zelf: bestaande structuur, cultuur, de samenstelling van de medewerkersgroep, lopende veranderprocessen, prioriteiten) en externe context (maatschappelijk, cultureel, fysiek, demografisch etc.) waarin een organisatie zich bevindt verschilt van die waarin de organisatie met de beste praktijk zich bevond.

Maar gaan organisaties dan wel op zoek naar die beste praktijk, gaan ze proberen te leren door bij andere organisaties te kijken, door als het ware te spieken en te pogen om een vertaalslag te maken? Dit is wel voorgekomen. Het komt incidenteel voor dat door de leiding van een organisatie een arrangement wordt aangegaan met uitvoerders van benchmarkonderzoek om juist wel voor een vertaalslag te zorgen of daaraan met eigen mensen van de organisatie bij te dragen. En we weten dat tal van gemeentelijke raadsgriffiers, al of niet vergezeld van een delegatie, in het eerste decennium van deze eeuw zijn afgereisd naar de gemeente Almere om daar te horen over het als uitdagend beschouwde concept van de politieke markt, dat daar was uitgevonden. Maar het betreft een uitzondering. Uit onderzoek naar waterschappen komt naar voren dat medewerkers van onderzochte waterschappen beter presterende waterschappen zelden of nooit bezochten om te leren van de onderliggende bedrijfsprocessen (Tillema, 2006: 304). Zodanige acties konden in geval van *waarstaatjegemeente* ook nauwelijks worden waargenomen (Abma en Korsten, 2009; Abma, 2012; Abma, 2014). Begrijpelijk omdat er door de uitvoerders van dit benchmarkinstrument nauwelijks in de gemeentelijke keuken werd gekeken.

Er kan overigens een rem bestaan op het bezoek van goed presterende organisaties. Welke? Een poging tot overname van een elders gegroeide praktijk kan jaren vergen. Als gevolg van langjarige contracten voor huisvuilophaal kan een goede praktijk van een andere organisatie inzake doelmatig huisvuilophaal niet zo maar overgenomen worden. Bovendien, als verandering wel mogelijk is, wil dit nog niet zeggen dat die ambtelijke of politiek-bestuurlijke prioriteit heeft. Er is dus blijkbaar een remweg voor correctie van bestaand beleid en een opstart naar nieuw beleid.

3) Er is ook meer in het algemeen (nog) geen studie te vinden die aanwijst wat aan goede praktijken is overgenomen door bij voorbeeld ziekenhuizen, thuiszorginstellingen, politiekorpsen, rechtbanken, universiteiten, waterschappen of gemeenten (De Groot, 2010; Korsten, 2010a; Abma, 2012). Toch komt navolging (imitatie) op zich wel voor, zo leert de organisatiekunde (DeMaggio & Powell, 1983; Lammers, 1993; Van Doorn e.a., 1993), maar het bewijs is niet gevonden dat positieve imitatie in sterke mate het directe of indirecte gevolg is van uitkomsten van benchmarking.

Dat het opsporen en overnemen van goede of beste praktijken van presteren overigens niet zo makkelijk is, valt te begrijpen uit de kritiek op 'evidence-based policy' (Pawson, 2006; Nelen, 2008; Coolsma, 2009). We menen: 'een goede praktijk is geen schroevendraaier die je in de winkel koopt'.

13.9 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 stelling wordt niet bevestigd. Daarvoor zijn de volgende argumenten aan te voeren.

1) Benchmarking bevredigt de nieuwsgierigheid. Van de inhoud van benchmarkrapporten wordt door ambtenaren aan opdrachtgeverskant alleen al daarom wel kennis genomen. Een dergelijke minimale mate van benutting is wel frequent traceerbaar. Maar is er meer? Benchmarkrapporten worden vaak – dus lang niet altijd - ter kennis gebracht van dagelijks 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. Als organisaties een zwak of slecht ontwikkeld zoekgedrag ten aanzien van het waarom en hoe van 'best practices' ten toon spreiden, duidt dat niet direct op een sterke aandring om te verbeteren.

2) Los van dit argument kan de vraag gesteld worden: stel dat de mythe geïnterpreteerd wordt als dat 'één cirkel leren' en 'dubbel cirkel leren' voorkomt, zijn daar dan aanwijzingen voor verkregen? Voor dubbel cirkel leren ('double loop learning'), waarbij waarden en doelstellingen achter beleid ter discussie komen, zijn geen harde aanwijzingen verkregen in studies over benchmarkonderzoek die handelen over afzonderlijke sectoren, als die van gemeenten, waterschappen, scholen 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, dus incrementeel leren. Wat houdt dat incrementeel leren eigenlijk in? We hebben gekeken naar 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. Lokale benchmarks zetten gemeentelijk beleid niet op de kop.

Om welke onderzoeksfuncties gaat het bij bezinning, als die wel plaatsvindt in 'een cirkel leren', dan? Benchmarking dient menigmaal ter legitimering van bestaand beleid of reeds ingezette of overwogen marginale aanpassingen in beleid of werkwijzen. Ambtenaren zeggen dan bij voorbeeld: 'we waren dit al van plan'. Anders gesteld, benchmarkuitkomsten dragen als deze al bestuurlijk benut worden, meer bij aan ondersteuning van een ingezette koers of aan incrementele veranderingen (verandering bij stukjes en beetjes) dan aan synoptische veranderingen.

Conclusie: de derde 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 overigens een bestuurlijke kwestie: welke benutting is goed genoeg?

13.10 Vierde mythe

De vierde genoemde mythe hield het volgende in: *'Verbeteren is in elk geval financieel-economisch gericht'*. Zijn benchmarks inderdaad (ook) financieel-economisch gericht of niet? Worden verbeterpogingen ondernomen en zijn verbeterpogingen (mede) financieel-economisch gericht? Gebleken is dat een aantal benchmarks inderdaad gedeeltelijk financieel-economisch is gericht.

De onderzoekers van PricewaterhouseCoopers stellen: *'Het inventariseren van de bedrijfsvoering van organisaties is jarenlang een onderdeel van de zorgbenchmarks geweest'* (Poerstamper e.a., 2007: 81). De redenering was: als een organisatie zich spiegelt aan de prestaties van andere organisaties en onderkent dat bepaalde andere organisaties het beter doen, dan zal het wenselijk zijn te verklaren hoe dat kan en hoe de organisatie die wil verbeteren daar dan kan komen. Dat voert naar (interne) organisatiegebonden verklaringen, zoals (aspecten van) bedrijfsvoering, en verklaringen die daar buiten liggen. De interne verklaringen zijn van belang want die zijn wellicht meer te beïnvloeden dan de andere verklaringen. De onderzoekers zijn daarom op zoek gegaan naar welke verbeteringen ook bij organisaties met een 'best practice' daadwerkelijk betere prestaties opleverden. Ligt het aan de hiërarchische lagenstructuur van sturing of juist aan een zekere platheid in de verhoudingen (horizontalisering)? Is sturen op hoofdlijnen van betekenis om tot beter presteren te komen? Heeft extra belonen zin of juist niet? Ligt het aan de communicatie of de organisatiecultuur? Aan welke knoppen zouden leidinggevend en anderen van de lerende organisatie moet draaien?

Er is bij voorbeeld een samenhang gevonden tussen financiële prestaties van een publieke organisatie en een aantal aspecten van de kwaliteit van het werk. Medewerkers voelen zich prettiger in organisaties die financieel goed presteren (zoals in verpleeg- en verzorgingshuizen, thuiszorginstellingen, organisaties van beroepsonderwijs). En omgekeerd is aannemelijk dat medewerkers willen bijdragen aan financieel gezonde organisaties.

Er is ook een relatie gevonden tussen financieel gezonde organisatie en kwaliteit van het werk, hoewel dit niet overal aantoonbaar is gebleken. Efficiency en kwaliteit gaan hier en daar samen maar niet overal (Poerstamper e.a., 2007: 78).

Enkele voorbeelden van verbanden tussen prestaties en bedrijfsvoering zijn de volgende.

- 'Instellingen met een beperkt aantal managementlagen scoren beter op kwaliteit van werk.'
- 'Hoe groter de zorgteams, hoe minder vaak de begeleidingsplannen met de cliënt worden opgesteld.'
- 'Hoe groter de zorgteams, hoe minder positief het cliëntoordeel.'
- 'Hoe meer inzicht in de kostprijs, hoe minder vaak er gekort wordt op zorg- en dienstverlening.'
- 'Hoe meer vrijwilligers, hoe negatiever de cliënten.'
- 'Hoe sterker de sturing en betrokkenheid van het middenmanagement, hoe positiever het cliëntoordeel.'
- 'Het nakomen van afspraken is een element dat voor cliënten zwaar weegt in het totaaloordeel over kwaliteit van zorg'.
- 'Als medewerkers negatief zijn over het planningsysteem, verzui men ze vaker.'
- 'Een decentraal planningsysteem gaat samen met een hogere productiviteit van medewerkers. Bij een dergelijk systeem kunnen medewerkers makkelijker voor elkaar inspringen als het nodig is.'
- 'De aanwezigheid van gedeelde waarden en normen blijkt van invloed op prestaties van organisaties.'
- (Zie voor gegevens hierover Poerstamper e.a., 2007: 83; ook Denhardt, 1993; Rose & Lawton, 1999; Denhardt, Denhardt & Aristiqueta, 2002; Boyne, Meier, O'Toole jr. & Walker, 2006).

Maar zijn verbeterpogingen daar dan ook op gericht? De vierde mythe gaat uit van enig gebruik en dus enige benutting van beste praktijken zoals opgenomen in benchmarkrapporten. Benchmarking van vergelijkbare organisaties draagt volgens deze mythe bij aan meer bedrijfsmatig functionerende overheidsorganisaties. Benutting van benchmarkrapporten zou dan, in deze operationalisatie, moeten leiden tot meer oog voor de afweging van kosten en baten, voor meer doelmatige uitvoering van taken, en hier en daar tot bezuinigingen. De reactie hierop luidt: tal van benchmarks zijn inderdaad wel gericht op financieel-economische aspecten van productie van producten of diensten (De Groot e.a., 2004; Poerstamper e.a., 2007: 81). Denk bij voorbeeld maar aan de studie *'Zuinig met*

zorg', over de productiestructuur van verpleeghuizen. Ook is bekend dat benchmarks tactisch gebruikt worden om bezuinigingen in een sector juist te voorkomen en meer personeel te vragen.

Maar voor nadrukkelijke verbeterpogingen op dit vlak is – in aansluiting op de tweede mythe - nog geen substantiële aanwijzing gevonden. Benchmarking van zuiveringsschappen heeft daarentegen wel bijgedragen aan stappen voorwaarts op dit vlak. Voor gemeenten hebben we dit (nog) niet kunnen vaststellen.

Benchmarking doet geen afbreuk aan bedrijfsmatige stappen voorwaarts, zeker niet in tijden van crisis en als de valkuilen bij benchmarking vermeden worden. Hoe meer adequate benchmarking is ingebed in een lerende cultuur en hoe meer expliciete stimulansen van bestuurders en topambtenaren om te willen leren geventileerd worden (door forumwerking te organiseren) en hoe hoger de politieke prioriteit, hoe groter de kans op goede kwaliteit van de toepassing van benchmarking, hoe meer kans er is op een bijdrage aan bedrijfsmatig functionerende overheidsorganisaties.

Maar sla niet door in doelmatigheidsstreven. Een goede en daarmee ook (betrekkelijk) doelmatige organisatie is als een winnende schaatser in een klassement voor allrounders. De winnaar wint meestal maar een of twee afstanden maar zelden is de suprematie zo groot dat de succesvolle schaatser alle vier de schaatsafstanden wint. De winnaar van de meerweekse wielrenwedstrijd zoals de Tour de France wint niet vanwege het uitsluitend beschikken over tijdritkwaliteiten, goed kunnen rijden in waaiers of bergbeklimmerskwaliteiten. De goede organisatie moet als een Sven Kramer of Peter Sagan in multidimensioneel opzicht goed zijn (Poerstamper e.a., 2007). Gebleken is zelfs dat een uitermate hoge doelmatigheid (van verpleeghuizen, verzorgingstehuizen, thuiszorginstellingen en centra voor gehandicaptenzorg) vaak samengaat met een middelmatige kwaliteit (Poerstamper e.a., 2007: 79). De implicatie hiervan is dat er een goed evenwicht moet worden gevonden tussen de betekenisvolle bouwstenen voor een 'goede' organisatie en doelmatigheid niet moet worden verabsoluteerd. 'Geld is ook niet alles.'

13.11 Na vier mythes

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 Ter Bogt, 2006; De Bruijn en Van Helden, 2007; 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 worden gevraagd. Ergo: de politieke urgentie en prioriteit is vaak niet groot genoeg. En het gevoel voor eigenaarschap van benchmarkonderzoek is bij colleges van burgemeester en wethouders en raden en individuele actoren te gering. Een poging ondernemen om dat eigenaarschap al voor de start van het onderzoek te ontwikkelen, verdient serieus overweging (zie ook Wijnen, 2006).

Benchmarkrapporten 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. Bestuurders sneuvelen niet op benchmarkrapporten. Als ze sneuvelen, en dat gebeurt in de gemeentelijke sfeer sinds 2000 vaker, dan is het op wat anders.

13.12 Kanttekening

Een kanttekening moet gemaakt worden. Een sterke gerichtheid op het ideaalbeeld van benchmarking kan teleurstelling over wat niet waargemaakt is tot gevolg hebben. Die houding kan zelfs het zicht verminderen of ontnemen op wat in de praktijk van organisatie wel is veranderd, aldus Henk ter Bogt (2006: 177). Wat zou er dan veranderd kunnen zijn?

Een analyse van ontwikkelingen van de laatste twintig jaar toont volgens geïnterviewde lokale bestuurders en ambtenaren wel dat het hele denken in gemeenten, zeker in gemeenten boven grofweg 25.000 inwoners, is veranderd naar meer gerichtheid op cultuur ('mindsets'), competentiegerichtheid van medewerkers en veranderingen, meer projectmatig werken, meer prestatiegerichtheid, meer kijken naar orde en veiligheid, meer zoeken naar externe legitimiteit voor handelen, meer co-creatie initiatieven, meer verantwoording (Korsten, 2009a, b). Onder invloed van NPM en de komst van de dualisering in 2002 is ook het aantal onderzoeken, al of niet in het kader van de controle door de gemeenteraden, toegenomen. Denk aan onderzoek van lokale rekenkamer(commissie)s (Baas en De Vries, 1988; Ter Bogt, 2006; Van der Woude, 2011). De aandacht voor interne veranderingen in de organisatiestructuur is daarentegen verminderd (Ter Bogt, 2006: 176-177; Aardema en Korsten, 2009).

13.13 Algemeen beeld aangevuld

Deze ervaringen bevestigen het volgende beeld dat uit praktijkervaringen in de afgelopen decennia in brede zin naar voren komt en zeker ook lijkt te gelden voor gemeenten. Deze praktijkervaringen van ruim twee decennia benchmarking in het Nederlandse openbaar bestuur op decentraal niveau hebben we gerubriceerd naar de kenmerken voor benchmarking zoals we die eerder 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). De oorzaak ligt hierin dat effecten laat of later dan prestaties blijken en effecten niet goed in benchmarks te vangen blijken.

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 leidinggevenden 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. Kwalitatieve vergelijking, die rekening houden met of aangevuld worden met verklaringen van substantiële verschillen in scores tussen organisaties, komen veel minder voor.
- Verklaringen voor scores ('het waarom?') ontbreken veelal in rapportages. Aanvullende rapportages over verschillen ontbreken vaak.
- De interpretatie van uitkomsten is in benchmarkrapportages nogal eens sober of afwezig. Dat betekent dat de kwaliteit van benchmarkrapportages niet zelden zwakten vertoont.

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 (ook internationale vergelijkingen, cf Hakvoort & Klaassen, 2010: 133).

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 beleidswijziging 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.

13.14 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. Enkele empirisch gefundeerde suggesties zijn te geven. Wat bevordert benutting van benchmarks in de Nederlandse context praktisch gezien wel?

1. Zorg er als overheidsorganisatie voor dat een te benchmarken beleidsgebied (producten, diensten, processen) zekere politieke urgentie en prioriteit heeft, bij voorbeeld door benchmarking te verankeren in een bestuurlijk coalitieakkoord, kwaliteitshandvest of programmabegroting. Daardoor stijgt later de kans op agendering in politiek-bestuurlijke gremia.
2. Benchmarking impliceert altijd vergelijking, ranking en streven naar verbetering. Zorg ervoor dat de benchmark ook leidt tot een leesbaar rapport en een analyse van de verschillen in ranking van organisaties. Wie echter teveel wil, bereikt weinig (De Bruijn en Van Helden, 2007: 25). 'Double loop learning' is normaal gesproken op basis van benchmarking uitgesloten. Lichtere verbeteringen in beleidspraktijken treden ook niet vanzelfsprekend op. Soms blijkt een 'best practice' niet zo maar overdraagbaar naar een andere organisatie. Daarom zijn verdere suggesties nodig, zoals tot vorming van fora ter discussiëring van de bevindingen uit benchmarkonderzoek.
3. Betrek potentiële gebruikers tijdig bij de bezinning op bevindingen uit benchmarkonderzoek (Korsten, 1983).

4. Let goed op of sprake is van een organisatie die in een netwerk functioneert. Is dat het geval dan moet ook bezien worden of netwerkpartners vanaf het begin deelnemen aan de benchmark en of ze de netwerkprestaties remmen of bevorderen en waarom (Van Thiel, 2009).
5. Prestatiemeting in een productgeoriënteerde setting (lees: aanbodgerichte benchmarking) blijkt alleen betekenisvol als deze ook elementen bevat uit een procesgerichte benadering. Een zekere betrokkenheid van opdrachtgevers en potentiële gebruikers bij de opzet van benchmarking is dus gewenst. De Bruijn en Van Helden (2007: 25) accentueren de wenselijkheid van 'co-creatie'.
6. Zorg in geval van benchmarking voor de aanwezigheid van normen ter beoordeling van prestaties en kijk niet slechts naar de mogelijkheid dat de organisatie beter kan als ze niet in de kwart van toporganisaties uit de ranking staat. Bezinning hierop is voorafgaand aan een benchmark al nodig, eventueel op basis van reeds uitgevoerde benchmarks.
7. Wees als co-creërende 'partijen' niet sober met het gebruik van indicatoren in benchmarkonderzoek. De presentie van een brede set indicatoren bij de benchmarking van afvalwaterzuivering bleek positief te werken (De Bruijn en Van Helden, 2007: 25; Van Thiel, 2009). Zorg voor veelzeggende indicatoren die een prestatieveld afdekken. De aanwezigheid van via sturing beïnvloedbare variabelen is gewenst.
8. In een benchmarkrapportage met prestatiecijfers komen zonder te weten wat inwerkt op de cijfers, schrikt gebruikers af. Benchmarkonderzoekers moeten altijd op zoek naar de werkelijkheid achter de cijfers (De Bruijn en Van Helden, 2007: 25).
9. Bevorder als overheidsbestuur tijdig de aanwezigheid van een forum dat zich richt op bezinning op uitkomsten van een benchmark.
10. Voorkom in opdrachtgever-opdrachtnemerrelaties dat tal van valkuilen bij benchmarking optreden en bevorder dat 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; Hakvoort & Klaassen, 2010: 133) eerder al eens deed. Serieuzer werk maken van de opdracht voor benchmarkonderzoek, en van de entreebesprekingen in dat verband, 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 worden nagegaan hoeveel en welke relevante gemeenten in de vergelijking meegenomen kunnen of moeten worden (de zgn. homogene referentiegroep),
- moet worden bijgedragen aan het verkrijgen van meerjarencijfers,
- een oorzakenanalyse van verschillen tussen scores onderdeel uit moet maken van de benchmarkrapportage,
- de ontvanger een afzonderlijk rapport voor de eigen organisatie moet vragen met cijfers, interpretaties en 'best practices' die voor de eigen gemeente relevant zijn,
- de opdrachtgever 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 wat zij met de uitkomsten willen gaan doen. Gaat het slechts om 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 geworden tot een kwestie van 'voor een dubbeltje op de eerste rang willen zitten'. Opdrachtgevers krijgen suboptimale benchmarkrapporten omdat ze te weinig betrokken zijn bij de opdracht en ook niet voldoende willen betalen voor kwalitatief hoogstaande benchmarking en daarmee voor 'leren', en 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.

14 Reflectie: enkele stellingen over benchmarking

We geven ten slotte bij wijze van reflectie op de voorgaande hoofdstukken enkele stellingen ter discussie en lichten die toe. Veel stellingen gaan over het ultieme doel van benchmarking: verbeteren. Daarom wordt nogal aandacht besteed aan de 'best practices' en het trachten te verbeteren van de organisatie in de richting 'good governance'. De mythen keren niet in stellingvorm terug omdat deze al behandeld zijn.

Fig.22: Stellingen over benchmarking op een rij

Stelling 1: *Benchmarking is onder voorwaarden toepasbaar in de publieke sector als middel voor prestatiemeting en als hulpmiddel voor prestatieverbetering.*

Stelling 2: *Een benchmarkexercitie is in principe niet ingewikkeld, maar wordt soms ingewikkeld gemaakt.*

Stelling 3: *Door periodiek aan een benchmarkexercitie deel te nemen, heeft een organisatie zicht op haar verbeteringen en zal deze daardoor stap voor stap op een hoger niveau van presteren komen.*

Stelling 4: *Als een benoeming van 'best practices' in een benchmarkrapport niet meer is dan een reeks cijfers, is de kans op onbruikbaarheid van het rapport groot.*

Stelling 5: *Benchmarking (vergelijken, rangordenen, analyseren en benoemen van 'goede praktijken') heeft iets van een inhaalrace.*

Stelling 6: *Als iets evident een transfereerbare 'best practice' is, is het einde van die 'best practice' nabij.*

Stelling 7: *Hoe meer geaggregeerd een 'best practice'- variabele is, hoe groter de kans is op niet overname van die 'best practice'.*

Stelling 8: *Het nastreven en verwerkelijken van 'best practices' die voortkomen uit benchmarking bezorgt een organisatie nog geen superieure en onderscheidende status.*

Stelling 9: *Het nastreven van 'best practices' als gevolg van benchmarking is aantrekkelijk, maar maakt van een organisatie nog geen innovatieve organisatie.*

Stelling 10: *Een organisatie die in alle onderzochte opzichten 'best practices' kent, bestaat in werkelijkheid niet.*

Stelling 11: *In werkelijkheid is het overnemen of nadoen van goede praktijken door organisaties als gevolg van benchmarking zelden iets van 'copy & paste'.*

Stelling 12: *Het analyseren naar 'bad practices' kan uit het oogpunt van leren zeker ook nuttig zijn.*

Stelling 13: *Het over de schutting kijken naar goede praktijken is uit doelmatigheidsoogpunt discutabel als organisaties al 'goed' functioneren maar wel nuttig om te zien 'dat het anders kan'.*

Stelling 14: *Het nastreven van goede praktijken is verdedigbaar, maar onzinnig als een organisatie op heel veel gebieden onder de maat functioneert.*

Stelling 15: *Kijken naar goede praktijken gebeurt meestal bij 'vergelijkbare organisaties' uit dezelfde sector of branche maar kan ook sectoroverstijgend of branchevreemd zijn.*

Stelling 16: *Benchmarking heeft vrijwel nooit primair betrekking op de kwaliteit van het bestuur van en toezicht op het bestuur van een organisatie.*

Stelling 17: *Benchmarking beantwoordt in de gemeentelijke praktijken vaak niet aan de ambities die in de literatuur aan benchmarking in het openbaar bestuur worden toegedicht.*

Stelling 18: *Gemeentelijke benchmarkrapporten komen zelden actief in de openbaarheid en dragen niet bij aan de actieve verantwoording over gemeentelijke prestaties.*

Stelling 19: *De kwaliteit van tal van gemeentelijke benchmarkonderzoeken is verbeterbaar.*

Stelling 20: *Het niet grondig bespreken van bevindingen van benchmarking door colleges van b&w en gemeenteraden moet mede verklaard worden uit de opvatting van collegeleden en raadsleden dat ze benchmarking*

eerder zien als bedrijfsvoering - en daarmee als een zaak van ambtenaren en van (uitvoerings)techniek - dan als een zaak van politieke sturing.

Zoals nog zal blijken, kan niet elke stelling een verdediging doorstaan.

Toelichting

Stelling 1: Benchmarking is onder voorwaarden toepasbaar in de publieke sector als middel voor prestatiemeting en als hulpmiddel voor prestatieverbetering.

Toelichting: Benchmarking heeft veel potentie. Benchmarking is veelvuldig toegepast, in de publieke sector en zeker ook bij gemeentebesturen en hun organisaties. En benchmarking wordt door leidinggevenden van gemeenten en andere organisaties erkend als zijnde van belang. Leidinggevenden, ambtenaren en bestuurders, verstreken immers de opdrachten tot benchmarking. Dat zegt iets. Benchmarking is blijkbaar ook mogelijk als rekening gehouden wordt met afwijkende kenmerken van de publieke sector in vergelijking met de private sector. Benchmarking heeft echter ook beperkingen. Benchmarking van of bij geheime diensten of toezichtorganisaties komt niet voor omdat dergelijke organisaties zich niet lenen voor benchmarking. De producten, diensten of processen laten zich niet makkelijk vangen in expliciet kwantitatief geformuleerde ambities en streefdoelen. Benchmarkonderzoek is vooral toepasbaar in een sector of branche waar organisaties zich qua producten, diensten of processen laten vergelijken en rangordenen, waar essentiële prestaties adequaat meetbaar zijn en waar benchmarking voldoet aan een aantal andere kwaliteiten. En als valkuilen, zoals het optreden van ongewenste effecten – bij voorbeeld in de vorm van cijfermanipulatie -, vermeden worden.

Stelling 2: Een benchmarkexercitie is in principe niet ingewikkeld, maar wordt soms ingewikkeld gemaakt.

Toelichting: Dat een benchmarkonderzoek niet ingewikkeld is maar ingewikkeld wordt gemaakt, is een empirische stelling omdat die verwijst naar hoe het in werkelijkheid gaat. Is hier bewijs voor? De redactie van het weekblad Elsevier publiceert regelmatig overzichten van de beste woongemeenten of veilige gemeenten, de beste opleidingen, de beste ziekenhuizen. Dat zijn op het oog wel analyses met een doel maar het zijn geen erg pretentieuze operaties. De weekbladredactie vergelijkt, rangordent en verduidelijkt welke organisatie als geheel, gegeven de indicatoren, de beste is. Daar kan de lezer het dan mee doen. De redactie toont zich praktisch en geeft recent ook een verantwoording; dat is een pluspunt (Elsevier, 29 juni 2013).

Kan het nog gemakkelijker? Een benchmarkexercitie kan zelfs, zo is wel beweerd, op een flip-over met 'geeltjes' worden uitgevoerd, als de deelnemers de intentie hebben om gezamenlijk informatie uit te wisselen en van elkaar te leren. Als de focus daarentegen op de kwaliteit van de metingen komt te liggen en het gesprek over de definities van de variabelen gaat, staat het proces van benchmarking centraal en niet de uitkomsten. Dan wordt benchmarking al gauw ingewikkelder. Uiteindelijk kan het zelfs gaan om stochastische analyses (meer in De Groot e.a., 2004; Hakvoort en Klaassen, 2004, 2013). Een 'flip over' - aanpak is dan te verkiezen? De stelling (dat benchmarking ingewikkeld wordt gemaakt) is bij nader inzien, gelet op de toelichting, *niet* houdbaar. Wat zijn de tegenargumenten dan tegen het pleidooi voor een 'flip-over' - aanpak?

Het belangrijkste tegenargument is dat benchmarking een vorm van uitsluitend of hoofdzakelijk kwantitatief onderzoek is waarbij de analyse gericht moet plaatsvinden door deskundige onderzoekers, de aanpak verantwoord moet worden en de bevindingen toegelicht moeten worden in een rapport voor de opdrachtgever(s). Zonder deugdelijk onderzoek - dat relevant is, geldig, betrouwbaar, herhaalbaar -, zonder verantwoording, analyse en rapportage is niet van adequate benchmarking sprake. Nadenken over definities en afspraken over wat vergeleken wordt, is dus onvermijdelijk, evenals eenstemmigheid over de indicatoren, het aantal deelnemende organisaties,

de opslag van meerjarige data, etc. Dat is al een voldoende argument om de stelling (dat benchmarking te ingewikkeld wordt gemaakt) sterk te kritiseren of 'af te wijzen'. Benchmarking is nu eenmaal onderzoek en geen methode die om het even welke ambtenaar zo maar kan verrichten. De rapportages wijzen het uit (zie ook De Groot e.a., 2004; Poerstamper e.a., 2007).

Daar komt als tweede tegenargument nog bij dat een discussie 'met geeltjes' alleen *geen* goed zicht kan geven op de onderbouwing van 'best practices'. Wat is de kern van de beste praktijk en waarom is dat zo? Objecten van benchmarking zijn in de publieke sector veelal te complex om eenvoudig te analyseren wat de beste praktijk is.

Een derde tegenargument luidt als volgt. Een discussie met ervaringsdeskundigen, die gebruik maken van hulpmiddelen maar zich niet baseren op een deugdelijk onderzoeksverslag (waarbij een analyse die de verschillen tussen scores verklaart, ontbreekt), kan nuttig zijn maar dat noemen we dan, in lijn met de literatuur geen benchmarking. Voorstelbaar is dat vijf hoofden burgerzaken van vijf gemeenten bij elkaar gaan zitten en daarbij ervaringen uitwisselen over onder meer openingstijden, klantcontacten, reactiesnelheid en balietevredenheid van klanten. Mogelijk gaat het om een zinnige professionele bezinning maar dat is in onze opvatting géén serieuze benchmarking. Een bezinning kan *na afloop* van benchmarking enig nut niet ontzegd worden, maar we noemen deze actie van bij voorbeeld vijf personen op zich geen benchmarking. Een benchmarking van bij voorbeeld pensioenfondsen is echt niet uitsluitend te doen 'met geeltjes'.

Is het benchmarkonderzoek dan altijd deugdelijk en verdedigbaar? Uit bepaalde benchmarkonderzoeken blijkt dat een onderzoeker kan *verdwalen* in de data en dat een ordening van uitkomsten, zoals bleek bij vergelijking en rangordening van politie-organisaties, niet meer te vatten is. Een onderzoeker moet goed nadenken over wat er met de indicatoren aan de hand is als een organisatie in het ene jaar op een heel lage plaats te midden van twintig andere belandt en een jaar later op een heel hoge en daarna weer op een lage plaats. De onderzoekers Cleuver en Blank (2006) liepen hier dan ook tegenaan.

Vanuit de literatuur komt naar voren dat er een aantal methoden van kwantitatieve analyse van prestatiegegevens voor grotere aantallen organisaties bestaat, zoals multivariate analyses. Deze analyse wordt verricht door onderzoekers uit de eredivisie van benchmarkonderzoek. Ze publiceren in bladen als *'Journal of the Royal Statistical Society'* en in het *'Journal of Public Administration, Research and Theory'* (Smith, 2006).

Conclusie : er moet goed nagedacht worden over de gewenste kwaliteit van een benchmark in relatie tot het doel van de vergelijking en het gebruiksdoel in de praktijk.

Stelling 3: Door periodiek aan een benchmarkexercitie deel te nemen, heeft een organisatie zicht op haar verbeteringen en zal deze daardoor stap voor stap op een hoger niveau van presteren komen.

Toelichting: Volgens theoretische inzichten over benchmarking krijgen organisaties die periodiek meedoen aan zelfde benchmarkonderzoeken inderdaad een beter zicht op hun functioneren, met name door dit functioneren te spiegelen aan externe organisaties. Daardoor wordt informatie verzameld die kan bijdragen aan een permanente verbetering in het presteren van de organisatie. So far, so good.

Deze argumentatie is echter nog niet voldoende. Periodiek deelnemen aan een benchmarkonderzoek moet ook zicht geven op het *begrijpen* van verschillen in scores en op '*beste praktijken*' (wat gebeurt waarom elders beter?) om daadwerkelijk te kunnen leiden tot verbeteringen in praktijken.

Een periodieke deelname aan een onderzoek kan impliceren dat een organisatie volstaat met 'de vinger aan de pols houden' en inzichtverschaffing. Maar de vraag is of niet *ook* tot deliberatie over verbetermogelijkheden moet worden overgegaan of zelfs daadwerkelijk verbeteren.

Wat nu te concluderen? De stelling veronderstelt dat organisaties die periodiek benchmarken *vanzelfsprekend* naar een hoger niveau van presteren komen en dat de goede praktijken van elders *transfereerbaar* zijn naar een andere context van mensen en organisaties. Daarvoor is echter door Cleuver en Blank (2006) in hun studie naar politie-organisaties en door Abma en Korsten (2009) in hun studie *'Gemeenten in rapportcijfers'* geen bewijs gevonden. Organisaties die herhaaldelijk deelnamen aan benchmarking zijn vergeleken. Wie als overheidsorganisatie periodiek deelneemt, blijkt niet geleidelijk een hoger niveau van presteren te bereiken. Niet uit te sluiten is overigens dat verder onderzoek naar andere domeinen of onderzoek van andere organisaties een andere conclusie oplevert.

Stelling 4: Als een benoeming van 'best practices' in een benchmarkrapport niet meer is dan een reeks cijfers, is de kans op onbruikbaarheid van het rapport groot.

Toelichting: Deze stelling slaat op een sobere rapportage in menig benchmarkrapport. Wat is de consequentie van die sobere rapportage? Een beste praktijk in de vorm van slechts een score die hoger is dan van andere organisaties zegt nog heel weinig en kan niet beschouwd worden als het verschaffen van helderheid over de beste praktijk(en). Het cijfer geeft nog geen richting voor verbetering aan, hoogstens waar je 'de wijsheid' zou kunnen vinden. Het is immers onbekend *waarom* het cijfer zo hoog is. Is de score verdiend door een bepaalde doelgerichtheid, projectmatigheid of werkwijze of door een bepaalde organisatiecultuur of wat dan ook? De kans is heel groot dat een benchmarkonderzoek, dat geen verklaring(en) bevat voor een beste score, ongebruikt in de la blijft liggen en niet tot bezinning aanzet. Een benchmarkonderzoeker doet er dus goed aan, zo wordt alom erkend (De Groot, 2004), om te onderzoeken waarom een score 'beste praktijk' wordt verkregen.

De paradox is dat, ondanks dit gegeven, de opdrachtgevers hiertoe toch niet vanzelfsprekend over gaan. Dat heeft meerdere 'oorzaken'.

De eerste oorzaak is dat gemeentelijke benchmarkonderzoeken vaak aanbodgedreven zijn omdat het uitvoerend bureau gemeenten vraagt aan te schuiven. Vanwege de geringe budgetten die in het spel zijn, is instemming tot deelname ook betrekkelijk laagdrempelig.

Ten tweede, deelnemers zeggen: 'we willen eerst eens de cijfers zien en dan zien we verder', of 'een dergelijke aanvulling kan altijd nog', of 'een dergelijke analyse is ons te duur'. Het gevolg is dat benchmarkdeelnames een zwakke opdrachtgever-opdrachtnemer-discussie kennen. Zelden vindt op collegeniveau een stevige discussie plaats over de vraag of het gemeentebestuur of managementteam moet streven naar de status van 'lerende organisatie'.

Stelling 5: Benchmarking (vergelijken, rangordenen, analyseren en benoemen van 'goede praktijken') heeft iets van een inhaalrace.

Toelichting: Om een metafoor te gebruiken: als een mannelijke hoogspringer een wereldrecord heeft gevestigd door 2 meter en 48 centimeter hoog te springen, zal gepoogd worden dit record te breken. Zo is het volgens de theorievorming over benchmarking ook. Deze stelling achten we uitgaande van de benchmarkambities verdedigbaar. Tenminste, als daadwerkelijk sprake is van streven naar verbetering bij de leiding van organisaties die op de tweede, derde en volgende plaats in de rangorde van vergeleken organisaties staan.

Of de gemeentelijke praktijk dit ambitieus gedrag feitelijk ook echt laat zien, is een andere vraag.

Wie de stelling als een empirisch bewezen stelling ziet, moeten we voorlopig teleurstellen. Abma en Korsten (2009), Abma (2012) en Schutgens (2014) konden niet waarnemen dat gemeentelijke organisaties de betere organisaties nadrukkelijk gingen kopiëren of imiteren. Representanten van gemeentebesturen gaan wel elders kijken maar *'policy transfer'* is, zo blijkt in het openbaar bestuur,

door politieke, fysieke, economische, sociaal-culturele en andere contextverschillen tussen gemeenten niet vanzelfsprekend mogelijk.

Stelling 6: Als iets evident een transfereerbare 'best practice' is, is het einde van die 'best practice' nabij.

Toelichting: Deze stelling is - volgens de theorie van benchmarking - als ambitie verdedigbaar maar empirisch bewijs is nog niet voorhanden.

De officiële ambitie achter benchmarking luidt: het einde van een 'best practice' is nabij als andere organisaties weet hebben van wat de 'best practice' is. De aanname is dat organisaties met een zwakte die 'beste' praktijk ook willen overnemen. Is een 'best practice' eenmaal overgenomen dan vervalt die 'best practice' als 'best practice'; dan is sprake van een gewone goede praktijk en kan het zoekproces naar een nieuwe 'beste praktijk' volgen (Shapiro, 2012). Anders gesteld, springen vier hoogspringers over 2.46 dan *evenaren* ze het door ons hier veronderstelde record en is 2.47 of hoger de *gewenste nieuwe* best practice. De hoogspringers springen met een sprong van 2.47 'het oude record uit de boeken', zoals dagbladjournalisten melden. De aanname is dat ze dan weer hoger willen springen.

Maar gaat het ook zo bij politie-organisaties, ziekenhuizen, instellingen voor middelbaar beroepsonderwijs of gemeentebesturen? Nee, dergelijke organisaties zijn veel te groot en polycentrisch om, als de leiding het al zou willen, ze te richten op het nastreven van een record. Veel grote organisaties zijn een rijk van eilanden, met een eigen koers. Van zuiver hiërarchische sturing vanuit één punt is in tal van organisaties, ondanks erkenning van het politiek primaat, niet of nauwelijks sprake. En er is sprake van netwerken van van elkaar afhankelijke organisaties (Lammers, 1983, 1993). Een opeenvolging van 'beste praktijken' is dan ook (nog) niet waargenomen.

Stelling 7: Hoe meer geaggregeerd een 'best practice'-variabele is, hoe groter de kans is op niet overname van die 'best practice'.

Toelichting op stelling 7: Een benchmarkonderzoeker kan op basis van rangordening en analyse op zoek gaan naar een *kleine* goede praktijk of een *grote* goede praktijk. Een paar simpele voorbeelden illustreren dit.

Wie de kwaliteit van de gemeentelijke afdeling Burgerzaken onderzoekt, kan tot de bevinding komen dat gemeente X de beste is op het vlak van openingstijden omdat het melden van een verhuizing of aangifte van een geboorte 16 uur per dag kan. Dan is de 'beste praktijk' op het vlak van *een element* van Burgerzaken een simpele (kleine) variabele. Gemakkelijk te begrijpen en vermoedelijk ook wel gemakkelijk toe te passen en dus te kopiëren. Maar een variabele kan ook heel omvattend samengesteld zijn. Stel dat men vaststelt: Nederland is de beste rechtsstaat in West-Europa. Deze 'best practice' is opgebouwd uit misschien wel 100 omvattende deelvariabelen, zoals de aanwezigheid van grondrechten (als het kiesrecht voor mannen en vrouwen), het zich houden aan mensenrechten, de aanwezigheid van veel elementen uit de politieke democratie, een mede op solidariteit gebaseerd belastingstelsel, recht op rechtsbijstand voor zwakkeren, etc. We zeggen hiervan dat sprake is van een hoog aggregatieniveau. Een dergelijke 'best practice' is niet gemakkelijk over te nemen door de regering van een ander land. We weten immers dat het overnemen van het concept van een democratische rechtsstaat kan botsen met eigen culturele tradities en regels in landen als bij voorbeeld Polen, Egypte, Irak of Kenia (Korsten, Bertrand, De Jong en Soeters, 1995; Fukuyama, 2005).

De implicatie hiervan is: 'dromen' mag, maar zoek in het verband van benchmarking niet naar heel grote en complexe 'best practices', als je wilt dat een praktijk ook overgenomen wordt.

Wie onderzoek doet naar de doelmatigheid van gemeenten en tot de vaststelling komt dat gemeente X de meest doelmatige gemeente is in periode 1 heeft voorwaarden geschapen voor het *niet* kopiëren van deze praktijk omdat zelfs doelmatigheid vermoedelijk een te omvattend concept is om de daarmee verbonden onderliggende praktijken volledig over te nemen.

We zien in de praktijk van 'leren' dan ook geen voorbeelden van overname van hele grote concepten, ook al weten we tevens dat juist 'sensitizing concepts', als 'de zakelijke overheid', navolging kunnen vinden (Korsten, 1997). Dergelijke concepten bevinden zich tussen een ideologie en een oneliner. Het bestempelen van een ziekenhuis tot beste ziekenhuis en verwachten dat het succes van dit ziekenhuis dan volledig en in de volle 'breedte en diepte' gekopieerd kan worden, miskent hoe de praktijk werkt. Goede praktijken in bestuur, organisatie, management en werkwijzen van een ziekenhuis kopiëren, is geen eenvoudige zaak (De Jong, Korsten en Van der Made, 1992; Maarse, 2011). Bovendien speelt verschil in context een rol, zoals een andere zorgverzekeraar, andere concurrenten, andere bestuurlijke ambities, andere specialisaties. Het Tilburgs model van prestatiebesturing en contractmanagement bij gemeenten is dan ook nergens door een gemeente letterlijk gekopieerd, hoezeer het concept ook is uitgedragen. Het concept haalde zich overigens in Tilburg zelf in. Het bleek te zeer intern gericht (Korsten en Tops, 1998; Tops en Van Vugt, 1998).

Stelling 8: Het nastreven en verwerklijken van 'best practices' die voortkomen uit benchmarking bezorgt een organisatie nog geen superieure en onderscheidende status.

Toelichting: Deze stelling 8 is juist omdat wie iets kopieert van een ander, niet beter wordt dan die ander.

Stelling 9: Het nastreven van 'best practices' als gevolg van benchmarking is aantrekkelijk maar maakt van een organisatie nog geen innovatieve organisatie.

Toelichting: Nadoen van wat andere organisaties doen, is handig maar niet vanzelfsprekend echt vernieuwend (Shapiro, 2012:111). Kopiëren is immers niet gelijk aan fundamenteel vernieuwen (Loos, 2006). Wie streeft naar een innovatieve organisatie, wil vernieuwen en moet daarvoor condities scheppen en barrières opruimen. Dat impliceert zowel leren als afleren. Een lerende cultuur moet opgenomen zijn in de strategisch, tactische en operationele plannen. Dat houdt in: rekening houden met innovatie in selectie-, opleidings- en doorstromingsbeleid, leerplatformen inrichten en eigen creativiteit mobiliseren.

Er is hier en daar in benchmarks overigens wel aandacht voor innoverend vermogen. Innoverend vermogen wordt wel 'als een cruciale factor voor succes en overleven' beschouwd (Poerstamper e.a., 2007: 84).

Stelling 10: Een organisatie die in alle onderzochte opzichten 'best practices' kent, bestaat in werkelijkheid niet.

Toelichting: Uit benchmarking van politieorganisaties, gemeentebesturen en hun organisaties en universitaire opleidingen blijkt dat een organisatie die in alles (in vergelijkend opzicht) de beste is niet bestaat. Er geen gemeente aangetroffen met uitsluitend scores 'goed' of 'excellent' (Abma en Korsten, 2009; Abma, 2012). Een organisatie bestaande uit uitsluitend 'best practices' is dus een fictie.

Presteert een grote organisatie wel vaak beter dan een kleinere? In veel benchmarks wordt een relatie gevonden tussen de omvang van de instelling en de prestaties (Poerstamper e.a., 2007: 85). Dit is bij voorbeeld gebleken bij woningcorporaties (Van Bortel e.a., 2012). In de sfeer van thuiszorginstellingen, verpleeg- en verzorgingshuizen blijkt dat kleine instellingen, die 'stand alone' functioneren of als onderdeel van een groter geheel zoals een concern, zowel op het vlak van

financiën als op dat van kwaliteit beter presteren. Willen grotere organisaties ook kwaliteit leveren en positief presteren, wat dan te doen? Een oriëntatie op de menselijke maat is gewenst; zorg voor een kleinschalige manier van organiseren. Dit is een bijna universele wetmatigheid.

Een aanvulling: 'Instellingen in verstedelijkte gebieden scoren bijna altijd minder goed dan instellingen op het platteland, op alle bouwstenen'(Poerstamper e.a., 2007: 85)

Stelling 11: In werkelijkheid is het overnemen of nadoen van goede praktijken door organisaties als gevolg van benchmarking zelden iets van 'copy & paste'.

Toelichting: 'Knippen en plakken' veronderstelt de mogelijkheid van mechanisch handelen, waarbij alle overige variabelen geen verandering ondergaan en vanzelfsprekend meebewegen. In veel organisaties gaat die voorwaarde niet op. Machinebureaucratieën komen nog het meest in aanmerking voor het overnemen van goede praktijken.

Stelling 12: Het analyseren naar 'bad practices' kan uit het oogpunt van leren zeker ook nuttig zijn.

Toelichting: Benchmarking is gericht op het vinden van beste praktijken en het begrijpen daarvan, zodat andere organisaties die praktijk kunnen aangrijpen voor het starten van verbeteringen. Moeten het altijd beste praktijken zijn? Soms valt er juist veel te leren van een toegepaste misère-index of van één casus, zeker ook als dat een 'bad practice' is. Een analyse van een slechte praktijk, soms faalcasus genoemd, kan thema's (fouten, zwakten, valkuilen, oorzaken) in beeld brengen die een studie van veel gevallen in de vorm van benchmarking nooit boven tafel brengt. Denk aan de analyse van de totstandkoming van stadhuizen in Amsterdam en Apeldoorn (Otten, 2000), het ontwerp en de aanleg Noord/Zuidlijn in Amsterdam, dus de metro-affaire (Soetenhorst, 2011), en aan andere affaires, zoals de hogesnelheidslijn-Zuid (De Vries en Bordewijk, 2009).

In de sociale huisvestingssector traden vanaf 1990, het telmoment, tal van affaires aan het licht. Het ging om de volgende corporaties, met tussen haakjes de 'saneringsheffingen' van het Centraal Fonds Volkshuisvesting: De opdracht, Lelystad (42 miljoen, 1990); Nieuw Amsterdam (214 milj., 1993); Woningbeheer Limburg, Geleen (948 milj., 1996); ANWS, Amersfoort (63 milj., 1997); Huisvesting Bejaarden, Nieuwegein (10 milj., 1997); De Samenwerking, Capelle a.d. IJssel (51 milj., 1998); Verantwoord Wonen, Nieuwerkerk a.d. IJssel (20 milj., 2001); SGBB, Hoofddorp (35 milj., 2010); Vestia, Rotterdam (675 milj., 2012).

Met het debacle bij woningcorporatie Woningstichting Geertruidenberg (WSG) werd de parlementaire enquêtecommissie woningcorporaties in juli 2013 van een nieuwe casus voorzien. De overige corporaties moeten, omdat ze onderling elkaars waarborg zijn, om en nabij 118 miljoen euro aan noodsteun geven voor deze corporatie die gekenmerkt werd door wanbeleid. Wat is het centrale probleem? De corporatie treft het verwijt van overmoed en zelfoverschatting (NRC, 310713). Door het bestuur werd de schuld voor het echec gelegd bij de in 2012 weggestuurde corporatiedirecteur, die niet alleen werd beschuldigd van verduistering en valsheid in geschrifte maar ook van falend beleid. De directeur, die jarenlang als accountant had gewerkt, vergaloppeerde zich met de aankoop van bouwgrond, investeringen in zorgcomplexen en commercieel vastgoed. De corporatie kwam in problemen toen de vastgoedmarkt instortte. De corporatie boekte wel af, maar haar eigen vermogen bleek niet groot genoeg. Ondanks het debacle vond het openbaar minister in juli 2013 nog onvoldoende harde aanwijzingen voor een strafrechtelijke vervolging.

Deze casus past daarmee in het beeld dat Aedes, de branchevereniging van woningcorporaties, in 2013 naar voren had gebracht in het rapport 'De balans verstoord'. In dat rapport werd de vinger gelegd op een tekort aan een goed werkend collectief moreel kompas en op gebrekkig zelfcorrigerend vermogen. Het zelfreinigend vermogen schoot en schiet zowel bij menige corporatie als in de sector als geheel tekort. Alle benchmarkonderzoeken en visitaties in de sector hebben

blijkbaar *niet voldoende preventief gewerkt* (Van Bortel e.a., 2012). Dat is wel begrijpelijk omdat benchmarkonderzoek en visitaties zich doorgaans niet expliciet richten op de risicovolle porte feuille. Wat de hele sector parten speelde, is de financiële crisis en de pe rikelen op de woning- en vastgoedmarkt die door weinigen zo voorzien zijn dat 'het schip tijdig de veilige haven binnen kon varen'. Conclusie: de parlementaire enquêtecommissie woningcorporaties (commissie- Van Vliet) ging zich richten op 'bad cases' & 'bad practices'. Het zijn vaak dergelijke commissies die zich op slechte praktijken richten. Daar valt blijkbaar ook van te leren.

Laten we een andere sector bij de kop pakken. We kennen benchmarking van ziekenhuizen gericht op het verkrijgen van inzicht in financiële prestaties van algemene ziekenhuizen maar die informatie is slechts gebaseerd op jaarverslagen (BDO, 2012). Weliswaar nuttig om zicht te krijgen op financiële performance maar toen de Inspectie voor de Gezondheidszorg in 2013 kwam met een analyse van het bestuur van en toezicht op het bestuur van het Amsterdamse ziekenhuis VUmc heeft deze inspectie aanbevolen goed te kijken naar deze ene casus. Deze casus was te typeren als '*hommes in het hospitaal*'. Er ging jarenlang veel mis in het besturen van dit ziekenhuis. Bestuurders verloren de aansluiting op primaire processen en konden bepaalde problemen niet oplossen, laat staan voorkomen. De toezichthouders schoten volgens de inspectie ook jarenlang te kort. De raad van toezicht greep niet in toen het bestuur faalde. Later is er wel schoon schip gemaakt. Enkele toezichthouders stapten op en in het bestuur trad verandering op. De inspectie beveelt de casus aan omdat deze zich zomaar elders in Nederland voor zou kunnen doen. Een zeer leerzame casus volgens de inspectie. Ook van andere ziekenhuis casus valt te leren. En van analyses van individuele woningcorporaties (Vestia), mbo-organisatie (Amaranthis), hbo- instellingen (InHolland; rapport van de commissie-Leers resp. de commissie-De Wit).

Kunnen we boven de casuïstiek van het VUmc of Amaranthis uitstijgen en komen tot praktische aanwijzingen waarvoor voldoende bewijs is? We menen van wel. Neem de volgende stellingname. De kans is groot dat boven een bepaalde grootte van een onderwijs- of ziekenhuisorganisatie (qua sectoren, personeelsaantallen en primaire processen) en bij de aanwezigheid van omgevingsdynamiek bij het dagelijks bestuur het zicht op wat zich op de werkvloer afspeelt verdwijnt, sterk vermindert of dat het bestuur af en toe in de mist terecht komt waarna besluiten van zwakke kwaliteit genomen worden. Wie het zover niet wil laten komen, kan maatregelen overwegen. Alternatieven zijn: het bestuur herzielt de eigen werkwijzen door zelf meer 'de vloer op te gaan'; het bestuur intitutionaliseert een andere scherpere signalering ('early warning') en interactie met betrokkenen; de organisatie wordt ontmanteld en opgesplitst in kleinere eenheden (zoals bij Amaranthis); bureaucrativering in de vorm van codes e.d.; specifieke maatregelen als gevolg van bepaalde knelpunten of spanningen; werken aan vertrouwensherstel in de organisatie en omgeving. Een combinative van deze en andere suggesties is denkbaar. De dreiging om te centraliseren of bureaucratiseren (meer regels, meer controle) moet tegengegaan worden.

Voor een raad van toezicht is te overwegen: de kans is groot dat boven een bepaalde grootte van een onderwijs- of ziekenhuisorganisatie, zeker als ook sprake is van omgevingsdynamiek, een raad van toezicht niet tijdig meegroeit in de controle op het dagelijks bestuur en onvoldoende zicht krijgt of houdt op zwakke signalen vanuit de organisatie of de omgeving richting dagelijks bestuur. Nader te overwegen remedies: 'checks and balances' versterken (dus krachtiger tegenspel bieden); vermijden van steeds dezelfde richting op denken, meer advocaten van de duivel aanstellen ('groupthink'; 't Hart, De Jong en Korsten, 1991); professionalisering van toezichthouders; frequenter vergaderen/ intensiever toezicht; toezichtcommissies instellen met een bepaalde opdracht; de ondersteuning versterken.

Stelling 13: Het over de schutting kijken naar goede praktijken is uit doelmatigheidsoogpunt discutabel als organisaties al 'goed' functioneren maar wel nuttig om te zien 'dat het anders kan'.

Toelichting: Wie het op tal van fronten al goed doet, zal nog slechts met grote inspanningen kleine stappen voorwaarts kunnen zetten. Aldus de wet van de afnemende meeropbrengst. Het gaat dan - om met de voormalige voetballer en coach Johan Crujff te spreken - nog slechts om verbetering door te letten op details (Davidse, 2008). Denk aan een voetbalclub met een stabiele bestuurlijke structuur, een gezonde financiële huishouding, een kwalitatief goed stadion en veel langjarige sponsorcontracten, een goede trainer en een goede team dat bijna elk jaar kampioen wordt. Misschien toont het kijken naar teams die in andere competities kampioen worden dat de medische begeleiding van spelers bij die kampioenen anders en beter is, of dat het arsenaal aan inzetbare spelers groter moet zijn dan grofweg 17 op niveau inzetbare spelers om zwakheden als gevolg van blessures die optreden door een uitgebreid wedstrijdschema op te kunnen vangen. Goede clubs kopen spelers maar verhuren ze ook aan clubs die geen concurrent zijn, met de contractclausule dat ze terugroepbaar zijn in geval van 'noodzaak'.

Stelling 14: Het nastreven van goede praktijken is verdedigbaar maar onzinnig als een organisatie op heel veel gebieden onder de maat functioneert.

Toelichting: In een organisatie die zwak functioneert, is een heel fundamentele bezinning nodig op 'hoe uit het slop te raken'. Dan is de kans niet groot dat het kopiëren van praktijken van andere organisaties genoeg zoden aan de dijk zet.

Stelling 15: Kijken naar goede praktijken gebeurt meestal bij 'vergelijkbare organisaties' uit dezelfde sector of branche maar kan ook sectoroverstijgend of branchevreemd zijn.

Toelichting: Het streven naar goede praktijken hoort volgens de theorie van benchmarking sectorgericht te zijn. Een voetbalclub kan derhalve in een bepaald opzicht leren van een goede andere voetbalclub en de leidinggevenden in een ziekenhuis kunnen leren van goede praktijken in andere ziekenhuizen. Vaak zal leren dan 'één cirkel leren' (single loop learning) zijn: het aanbrengen van verbeteringen in een bestaand systeem, model of patroon. Maar je kunt ook buiten de sector of branche stappen? Kan voetbal iets leren van hockey? Of gemeenten van het bedrijfsleven? Blikken werpen op branchevreemde goede praktijken vergroot de kans op een basis voor meer revolutionaire veranderingen in de eigen organisatie. Voorbeeld van buiten een sector kijken: de ketenaanpak bij het omgaan met criminele jeugd met een complexe problematiek in Amsterdam (Gerritsen, 2011).

Stelling 16: Benchmarking heeft vrijwel nooit primair betrekking op de kwaliteit van het bestuur van en toezicht op het bestuur van een organisatie.

Toelichting: Wie alle benchmarkrapporten, die betrekking hebben op publieke organisaties en die in Nederland sinds pakweg 1995 zijn verricht, de revue laat passeren, kan niet zeggen dat er benchmarks zijn die *expliciet* gericht waren op een analyse van de bestuurskwaliteit, de kwaliteit van het toezicht op dagelijks besturen van publieke organisaties en de verhouding hiertussen en de balans ('checks and balances'). Daarmee sluiten we aan op wat Korsten (2010) hierover al in 'Deugdelijk bestuur' stelde. Van vergeleken en gerangordende verpleeghuizen, ziekenhuizen, Riaggs, ROC's, woningcorporaties, gemeenten of politiekorpsen is altijd wat anders in benchmarkonderzoek centraal gesteld, zoals bij voorbeeld de overhead of de financiële prestaties. Vaak zijn er wel analyses te vinden van *individuele* organisaties die te maken kregen met een meerjarig problematisch functioneren van het bestuuren eventueel ook het toezicht of met een reeks crises, calamiteiten of incidenten. De 'bad practices' van bestuur van individuele organisaties zijn dus wel belicht. Denk maar aan de analyses van de hbo-instelling InHolland, de woningcorporatie Vestia, of het ziekenhuis VUmc of het drama van de Amsterdamse metro.

Stelling 17: Benchmarking beantwoordt in de gemeentelijke praktijken vaak niet aan de ambities die in de literatuur aan benchmarking in het openbaar bestuur worden toegedicht.

Toelichting: Het gaat volgens normatieve beschouwingen in de literatuur bij benchmarking om het onderzoekmatig analyseren en vergelijken van organisaties of een onderdeel of aspect daarvan (zoals producten, diensten of processen), het rangordenen van organisaties naar goed en minder goed (of omgekeerd), het analyseren van prestatieverschillen en het begrijpelijk maken hiervan, het formuleren van 'best practices' en het aannemelijk maken waarom een 'best practice' er uitspringt. Benchmarking leidt – normatief gezien - tot een zelfstandig leesbaar rapport dat verslag doet van het benchmarkonderzoek en daarmee van de vraagstelling, de werkwijze en resultaten van prestatievergelijking en verschil-analyses en conclusies, onder meer over de goede praktijken. Aan deze eisen voldoen gemeentelijke benchmarks vaak niet. Er is wel sprake van vergelijking en ordening maar de verschillen in prestaties worden niet tegen een norm gehouden en verklaard, en het ontbreekt aan het benoemen van 'best practices'. Als er sprake is van een rapport is het een rapport voor alle deelnemers, niet een reeks gespecificeerde rapporten voor elke opdrachtgever apart. Voor gemeentelijke insiders is het rapport wel leesbaar maar niet zelden moeilijk leesbaar voor iemand die niet vertrouwd is met het object van onderzoek. Daardoor blijven benchmarkrapporten vaak 'hangen' in een ambtelijke afdeling en bereiken ze zeker niet de gemeenteraden. Dergelijke rapporten worden zelden expliciet in gemeenteraden bediscussieerd. Benchmarking is daarmee dus niet wat het pretendeert te zijn.

Stelling 18: Gemeentelijke benchmarkrapporten komen zelden actief in de openbaarheid en dragen niet bij aan de actieve verantwoording over gemeentelijke prestaties.

Toelichting: De meeste gemeentelijke benchmarkonderzoeken zijn aanbodgedreven. Een aanbieder, meestal een adviesbureau, vraagt gemeenten mee te doen aan de toepassing van een benchmarkinstrument dat qua opzet en toepassing (nagenoeg) vastligt. Van een actieve wisselwerking tussen uitvoerder en gemeentebestuur is slechts in een minderheid van gevallen sprake en derhalve ontbreekt het aan maatwerk. Omdat conform afspraken de gemeentelijke opdrachtgever bepaalt of een benchmarkrapport openbaar wordt, verstrekken uitvoerders van een benchmark in een meerderheid van gevallen rapporten niet aan derden. Het gevolg is dat benchmarkrapporten binnen de gemeentelijke organisaties blijven hangen of liggen. Van actieve openbaarheid op websites is in een meerderheid van gevallen geen sprake. Dergelijke rapporten blijken uiterst moeizaam te achterhalen en beschikbaar te krijgen, zo bleek uit een inventarisatie onder Drenthse gemeenten in de periode 2000 tot en met 2017. Het ontbreekt dan ook heel vaak aan bespreking van benchmarkrapporten in de persmedia. De implicatie hiervan is dat benchmarking in de gemeentelijke praktijk in de meerderheid van gevallen niet de functie vervult van het leveren van een bijdrage aan de actieve verantwoording over gemeentelijke prestaties.

Stelling 19: De kwaliteit van tal van gemeentelijke benchmarkonderzoeken is verbeterbaar.

Toelichting: Na het voorgaande is deze stelling volledig te ondersteunen. De praktijk van gemeentelijke benchmarking blijft achter bij dat wat als normatief wenselijk is gezien in de literatuur. De praktijk is dus evident verbeterbaar. In dit boek zijn tal van argumenten aangevoerd in welke richting de verbetering moet gaan. Als de praktijk van benchmarking een paar stappen richting het normatieve beeld van wat benchmarking hoort te zijn gaat, treedt al een verbetering op. Die verbetering is ook gewenst. Treedt die niet op dan sterft benchmarking een stille dood.

Stelling 20: Het niet grondig bespreken van bevindingen van benchmarking door colleges van B&W en gemeenteraden moet mede verklaard worden uit de opvatting van collegeleden en raadsleden dat ze benchmarking eerder zien als bedrijfsvoering - en daarmee als een zaak van ambtenaren en van (uitvoerings)techniek - dan als een zaak van politieke sturing.

Toelichting:

Het eerste argument is dat colleges bedrijfsvoering hier en daar expliciet als argument gebruiken om na de in ontvangstname van een rapport er verder niet uitvoerig in een collegevergadering op in te gaan en een koers te bepalen. Voor gemeenteraden bevat een benchmarkrapport zelden 'politiek vuur' en is het geen 'lont' dat ontstoken kan worden.

In de tweede plaats heeft het politieke bestuur van doen met opinies over waarden die voorop staan of geschonden dreigen te worden, met doelstellingen en prioritering, en vaak minder met uitvoering van een beleid en techniek. Het tweede – daarop direct aansluitend - argument houdt in dat de uitkomsten van benchmarkrapporten vaak juist wel technisch van aard zijn, en daardoor moeilijk interpreteerbare uitkomsten bevatten. Mede hierdoor geven ze niet direct in politiek-bestuurlijk opzicht een richting aan die een bestuur moet inslaan. Denk aan rapporten over de technische aspecten van rioleringsbuizen in een rioleringsbenchmark. Hier komen bedrijfsvoering en 'technische bevindingen' dus als argument samen. Wil een rapport als richtingbepaler desalniettemin niet opgeborgen worden en politiek bestuurlijk betekenis krijgen, dan hoort er een bestuurlijke relevante (opleg)notitie bij het benchmarkrapport waarin juist een politiek-bestuurlijke duiding van de uitkomsten van de betreffende benchmarking wordt gegeven. En zo'n speciale notitie komt niet vanzelf uit de ambtelijke organisatie.

Dit zijn de kernargumenten. In de werkelijkheid van alle dag komen daar nog een paar argumenten bij.

Het derde argument pro de stelling 'benchmarking is een bedrijfsvoeringskwestie' is dat tal van benchmarkrapporten uitkomsten bevatten van het soort 'we zitten niet in het laagste segment van de rangorde uit het benchmarkrapport'. Van daaruit gezien, is er in de ogen van veel bestuurders geen urgentie of druk om tot bestuurlijk handelen over te gaan, zelfs als een rapport wel politiek bestuurlijk relevant is. 'We doen het immers redelijk goed'.

De verschijning van benchmarkrapporten trekt zich doorgaans weinig aan van politieke timing. Er is bij verschijning geen 'policy window', geen politiek 'handig' moment waarbij het onderwerp toch al op de agenda zou komen. Daarom zeggen bestuurders en ook ambtelijke adviseurs nogal eens: 'voor zover er wel iets met het rapport kan of moet gebeuren, wordt het meegenomen in lopende processen'.

Het vierde argument kan zijn: 'de benchmark sluit aan op een verbeterpad dat we al zijn ingeslagen'. Kortom, het rapport sluit daarop aan en gooit met andere woorden de bestuurlijke 'prioritering' niet omver.

Literatuur

- Aardema, H. en A.F.A. Korsten, Gemeentelijke organisatiemodellen – Hoe integraler het moet, hoe minder je het ziet ..., in: Bekke, H., K. Breed en P. de Jong (red.), *Naar een collegiaal en samenhangend overheidsbestuur – De lokale bestuurspraktijk als wenkend perspectief voor het rijk?*, Sdu Uitgevers, Den Haag, 2009, pp. 209-227.
- 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., *Contextmanagement – Een nieuw bestuurs- en managementconcept*, VNG Uitgeverij, Den Haag, 2000.
- Aardema, H., *Control voor leiders*, BMC, Elsevier, Den Haag, 2007.
- Aardema, H., *Doorwerking van BBI*, BMC, Leusden, 2002.
- 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., *Stille waarden – Een reflectie op overnormering in publiek management*, OU, Heerlen, 2005.
- 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.
- Abma, K., Bestuurskrachtmeting inclusief visitatie, in: Eversdijk, A. en J. Schutgens (red.), *Openbaar bestuur op expeditie – Verkenningen van bestuurlijke vernieuwingspogingen*, Boom Lemma, Den Haag, 2010, pp. 18-40.
- Abma, K. *Kwaliteit van Nederlandse gemeenten volgens hun inwoners*, paper gepresenteerd voor het 13^{de} politicologenetaal, 12 en 13 juni 2014.
- Aedesbestuur, *Het moet beter- Visiedocument*, Aedes, mei 2010.
- Albeda, H. en A.L. Buinink, *Benchmark responsesnelheid 8 gemeenten*, Stichting Rekenschap, Den Haag, 2004.
- Algemene Rekenkamer, *Handreiking meten van doelmatigheid*, Den Haag, 2003.
- Allers, M.A. en C. Hoeben, *Kerngegevens belastingen grote steden 2013*, COELO, Groningen, 2013.
- Allers, M.A., C. Hoeben e.a., *Atlas van de lokale lasten 2013*, COELO, Groningen, 2013.
- Allers, M.A., Decentraalisatie en gemeentelijke opschaling, in: *Liberaal Reveil*, juni 2013, pp. 119-123.
- Allers, M.A., L. Toolsema e.a., *Belastingoverzicht grote gemeenten*, COELO RUG, Groningen, 2010.
- Allison, G.T., Public and Private Management: Are They Fundamentally Alike in All Unimportant Respects, in: Perry, J.L. & K.L. Kraemer (eds.), *Public Management*, Mountain View, Mayfield, 1983.
- Allison, G.T., Public and Private Management: Are They Fundamentally Alike in All Unimportant Respects, in: Shafritz, J.M. & A.C. Hyde (eds.), *Classics of Public Administration*, Harcourt Brace College Publ., Fort Worth, 1997, pp. 383-401.
- 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).
- Andrews, C.W., Best Practices in Local Government, in: Vries, M.S. de, P.S. Reddy & M.S. Hague (eds.), *Improving Local Government – Outcomes of Comparative Research*, Palgrave, Londen, 2008, pp. 170-193.
- 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.

Assen, J.G.M., *Gemeentelijke bevolkingsontwikkelingen en lokaal beleid*, Nidi, Nijmegen, 1988.

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.

Baas, B. en M.S. de Vries, Het gaat goed in de gemeente, in: *Bestuurskunde*, 1988, nr. 2, pp. 50-62.

Banner, G., Strategische Führung und Modernisierung Unternehmens Kommune, in: Jann, W., M. Röber & H. Wollmann (Hrsg.), *Public Management – Grundlagen, Wirkungen, Kritik*, Edition Sigma, Berlin, 2006, pp. 253-270.

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

BDO-Accountants en adviseurs, *Zonder daadkracht geen toekomst – BDO-benchmark ziekenhuizen 2012*, 2012.

Beek, D. van, *Sturen naar een intelligente overheid*, Reed Business, Den Haag, 2007.

Beens, D.J. en K. van Koningsbruggen, Outsourcing van gemeentelijke taken, in: Bossert, H., L. Meuleman, J. Weber en N. Zwennes (red.), *Public governance- Van theorie naar praktijk. Voorbeelden van vernieuwing in publieke organisaties*, Sdu, Den Haag, 2006, pp. 117-137.

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.

Bemelmans-Videc, M.L. en M. de Vries, Ontwikkelingen in de evaluatie van overheidshandelen, in: 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, pp. 127-137.

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

Berg, Y. van den, P. Hanemaayer, Y. Wever en S. Monod de Froideville, *Evaluatie transparantiebenchmark*, Den Haag, 14 nov. 2013.

Blank, J. en H. de Groot, Paspoort duurder 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., M. van der Ende, B. van Hulst & R. Jagtenberg, *Bench Marking in an International Perspective - An International Comparison of the Mechanism and Performance of the Judiciary System*, Rotterdam, 2004.

Blank, J.L.T. (ed.), *Public Provision and Performance*, Emerald Group Publ., Amsterdam, 2000.

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. en P.M. Koot, Zuinig met water – Een empirisch onderzoek naar de efficiëntie van de bedrijven in de drinkwatersector, in: *Tijdschrift voor Openbare Financiën*, jrg. 36, 2004, nr. 2, pp. 51-63.

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., Data Envelopment Analysis: een goede praktijk, op www.benchmarkenindepubliekesector.nl, Kenniskamer Benchmarken Publieke sector, Min. van BZK, 2003.

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

Boelhouwer, J. en R. Gilsing, *De sociale staat van de gemeente*, SCP, Den Haag, 2012.

Boers, I. en C. van Montfort, Raden van toezicht, in: Bovens, M.A.P. en Th. Schillemans (red.), *Handboek publieke verantwoording*, Boom Lemma, Den Haag, 2009, pp. 143-161.

Bogt, H. ter, en S. Tillema, Kunst en zaken – Hoe New Public Management en verzakelijking ook doordringen in de theatersector, in: Bogt, H. ter, P. Jansen en S. Tillema (red.), *Met het oog op de publieke zaak*, RUG, Groningen, 2010, pp. 49-64.

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

Bogt, H.J. ter, Ambities en werkelijkheid van New Public Management: de moeizame managementvernieuwingen bij lagere overheden in Nederland, in: *Maandblad voor Accountancy en Bedrijfseconomie*, april 2006, pp. 170-179.

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 G.J. van Helden, Contractmanagement bij gemeenten, in: Korsten, A.F.A. en P.W. Tops (red.), *Lokaal bestuur in Nederland*, Samsom, Alphen, 1998, pp. 323-335.

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, Financieel management bij woningcorporaties: van boekhouding naar beleid, in: *Maandblad voor Accountancy en Bedrijfseconomie*, maart 2005, pp. 60-69.

Bogt, H.J. ter, G.J. van Helden and B. van der Kolk, Haagse bluf bij ambities over programmabegroting, in: *Tijdschrift voor Public Governance, Audit & Control (TPC)*, jrg. 10, 2012, nr. 1, pp. 4-9.

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, Politicians in search of performance information? Survey research on Dutch aldermen's use of performance information, in: *Financial Accountability and Management*, vol. 20, 2004a, nr. 3, pp. 221-252.

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, Recent and Future Management Changes in Local Government: Continuing Focus on Rationality and Efficiency?, in: *Financial Accountability & Management*, vol. 24, 2008, nr. 1, pp. 31-57.

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 2004b, pp. 52-58.

Bogumil, J. & S. Kuhlman, Wirkungen lokaler Verwaltungsreformen – Möglichkeiten und Probleme der Performanzevaluation, in: Jann, W., M. Röber & H. Wollmann (Hrsg.), *Public Management – Grundlagen, Wirkungen, Kritik*, Edition Sigma, Berlin, 2006, pp. 349-370.

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.

Boogers, M., P.W. Tops en A. Vries, *Effecten van dualisering voor burgers: beweging naar buiten?*, Vernieuwingsimpuls Dualisme en lokale democratie, VNG Uitgeverij, Den Haag, 2005.

Boogers, M.J.G.J.A., Hoe beoordelen inwoners hun gemeentebestuur en waar hangt dat oordeel van af?, in: *Bestuurswetenschappen*, jrg. 67, 2013, nr. 4, pp. 19-41.

Bordewijk, P. en H. Klaassen, *Begroten met beleid – Pleidooi voor planning en control op maat bij de gemeenten*, Sdu, Den Haag, 2011.

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.

Bordewijk, P., Maten voor gemeenten: de maat is vol, in: *Bank & Gemeente*, 15 maart 2013.

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.

Bortel, G. van, J. Hoekstra en M. Elsinga, *Vijf jaar corporatievisitaties - Balanceren tussen verantwoorden en leren*, TU Delft, Delft, 2012.

Bos, R. ten, *Fashion and utopia in management thinking*, KUB, Tilburg, 2000.

Bos, R. ten, *Modes in management*, Boom, Amsterdam, 2005 (2^e druk).

Bouckaert, G. & J. Halligan, Comparing Performance across Public Sectors, in: Dooren, W. van, en S. van de Walle (eds.), *Performance Information in the Public Sector- How it is used*, Palgrave, Londen, 2008, pp. 72-94.

Bouckaert, G. & J. Halligan, *Managing performance. International comparisons*, Routledge, London, 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 Proctivity 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. en T. Auwers, *Prestaties meten in de overheid*, Brugge, 1999.

- Bouckaert, G., A. Hondelghem, J. Voets, S. op de Beeck en E. Cautiaert (red.), *Handboek overheidsmanagement – Overheid in beweging*, Vanden Broele, Brugge, 2011.
- Bouckaert, G., B. De Peuter en W. van Dooren, *Metten 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, Berlin, 2006, pp. 119-133.
- Bouwman, H., Burgers nemen hun gemeente de maat, in: *Binnenlands Bestuur*, 13 maart 2009, pp. 6-7.
- Bovens, M. en Th. Schillemans, Publieke verantwoording: de staat van de verantwoordingsstaat, in: Andeweg, R. en J. Thomassen (red.), *Democratie doorgelicht – Het functioneren van de Nederlandse democratie*, Leiden University Press, Leiden, 2011, pp. 439-457.
- Bovens, M.A.P. en Th. Schillemans (red.), *Handboek publieke verantwoording*, Boom Lemma, Den Haag, 2009.
- Bowerman, M., G. Francis, A. Ball & J. Fry, The Evolution of Benchmarking in UK Local Authorities, in: *Benchmarking: An International Journal*, vol. 9, 2002, nr. 5, pp. 429-449.
- 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, L.J. O’Toole jr & R.M. Walker (eds.), *Public Sector Performance – Perspectives on Measurement and Management*, Cambridge UP, Cambridge, 2006.
- Boyne, G.A., O. James, P. John & N. Petrovsky, Top Management Turnover and Organisational Performance: A Test of a Contingency Model, in: *Public Administration Review*, vol. 71, july/august 2011, nr. 4, pp. 572-582.
- Braak, H. ter, en G. von Grumbkow, *Intensief samenwerken in het BEL-model*, WagenaarHoes/InAxis, Zeist, 2008.
- Braam, A. van (red.), *Actuele sociologie*, Van Gorcum, Assen, 1970.
- 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.
- Bressers, J.Th.A. en A. Hoogerwerf (red.), *Beleidsevaluatie*, Samsom, Alphen a.d. Rijn, 1995.
- 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.
- Brug, I. van der, R. van Dellen, F. Lustenhouwer e.a., *Sturen met benchmarks- Via meten en vergelijken tot betere bedrijfsprestaties komen*, Aedes, juli 2008.
- Bruijn, H. de, Afrekenen op prestatie leidt tot bureaucratie, in: *NRC*, 3 jan. 2003.
- Bruijn, H. de, en J. van Helden, Effectief prestatie management bij professionele organisaties in de publieke sector, in: *Overheidsmanagement*, jan. 2007, nr. 1, pp. 20-26.
- 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.
- Bureau Louter, *Waar willen we wonen 2013*, Delft, 2013.
- CAB, Leven in de leegte, *Leefbaarheid in Noord-Groningen anno 2010*, Groningen, 2011.
- 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.

COELO, *Atlas rijksuitkeringen aan gemeenten 2011*, RUG, Groningen, 2011.

Commissie Benchmarking, *Handreiking beoordelingsproces commissie benchmarking – Leidraad voor benchmarkaanbieders*, Aedes, 2008.

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.

Commissie-Van Rijn (Commissie Kwaliteitsmaatstaf Benchmarking), *Beter meten- Benchmarking de maat genomen*, Aedes, maart 2007.

Coolsma, Chr., *Metten is slechts het begin van weten*, in: Ridder, J. de, P.L. Polhuis en L.W.J.C. Huberts (red.), *Eigenzinnig bestuur*, Wolf Legal Publishers, Nijmegen, 2009, pp. 47-65.

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

Cuperus, R. en R. Lansink, *Overheidje spelen is niet vrijblijvend*, in: *NRC Handelsblad*, 22 aug. 2013, p. 15.

Davidse, H., *'Je moet schieten, anders kun je niet scoren' en andere citaten van Johan Cruijff*, Uitgeverij BZZToH, Den Haag, 2008.

De Lancer Julnes, P., *Performance Measurement Beyond Instrumental Use*, in: Dooren, W. van, en S. van de Walle (eds.), *Performance Information in the Public Sector- How it is used*, Palgrave, Londen, 2008, pp. 58-72.

Deijkers, H. en A. van Leeuwen, *Leeswijzer bij ranglijst van beste gemeenten*, in: *Elsevier*, 29 juni 2013, pp. 54-55.

Denhardt, R.B., J.V. Denhardt & M.P. Aristigueta, *Managing Human Behavior in Public & NonProfit Organizational Settings*, Sage, Londen, 2002.

Denhardt, R.B., *The Pursuit of Significance – Strategies for Managerial Success in Public Organizations*, Wadsworth Publ. Company, Belmont, 1993.

Denters, B. en T. Toonen, *Bestuurskrachtmeting 3.0*, in: Aardema, H., W. Derksen, M. Herweijer en P. de Jong (red.), *Meerwaarde van de bestuurskunde*, Boom Lemma Uitgevers, Den Haag, 2010, pp. 223-233.

Denters, S.A.H., *Prestatiesturing in de theorie en praktijk van het Nederlandse grotestedenbeleid*, in: *Beleidswetenschap*, jrg. 15, 2001, nr. 4, pp. 356-372.

Denters, S.A.H., *Prestatiesturing: pronkstuk of probleem? Prestatiemeting in theorie en praktijk van het Nederlandse grotestedenbeleid*, in: *Beleidswetenschap*, jrg. 15, 2001, nr. 4, pp. 356-371.

Derksen, W., *Kennis en beleid verbinden: Praktijkboek voor beleidsmakers*, Den Haag: Boom Lemma uitgevers, 2011.

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

Dijkstal, H. en J. Mans, *Krimp als structureel probleem – Rapportage topteam krimp voor Groningen*, Groningen, nov. 2009.

Dijkstra, G.S.A. en Th.A.J. Toonen, *De prestaties van de Nederlandse rechterlijke macht vergeleken en gemeten*, in: *Bestuurswetenschappen*, jrg. 58, dec. 2004, nr. 6, pp. 536-538.

Dimaggio, P.J. & W.W. Powell, *The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields*, in: *American Sociological Review*, vol. 48, 1983, nr. 2, pp. 147-160.

Dittrich, K., *Het nieuwe Nederlandse accreditatiestelsel: doestellingen, verwachtingen en valkuilen*, in: *Thema – Tijdschrift voor hoger onderwijs & Management*, 2010, nr. 4, pp. 23-27.

Dixon, R., C. Arndt, M. Mullers, J. Vakkuri, K. Engblom-Pekkala & C. Hood, *A Lever for Improvement or a Magnet for Blame? Press and Political Responses to International Educational Rankings in Four EU Countries*, in: *Public Administration*, vol. 91, 2013, nr. 2, pp. 484-505.

Donaldson, L., *The Contingency Theory of Organizations*, Sage, Londen, 2001.

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.

Doorn, J.A.A. van, P. Meurs en T. Mijs (red.), *Het organisatorisch labyrint*, Het Spectrum, Utrecht, 1993.

Doorn, J.A.A. van, *Wijken stad: reële integratiekaders*, in: Groenman, S. en H. de Jager (red.), *Staatkaart van de Nederlandse sociologie*, Van Gorcum, Assen, 1970, pp. 231-254.

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., *Accreditatie – Een zelfdiagnose*, Sdu, Den Haag, 2004.

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

Downs, A., *Inside Bureaucracy*, Little Brown, Boston, 1967.

Dozy, M.C., J. P. Laurier en W.J. van Noort (red.), *Balans van het organisatiesociologisch denken – Hoogtepunten uit het werk van Cor Lammers*, Het Spectrum, Utrecht, 1993.

Drew, S.A.W., From Knowledge to Action: The Impact of Benchmarking on Organizational Performance, in: *Long Range Planning*, vol. 30, 1997, nr. 3, pp. 427-441.

Dronkers, J., De maatschappelijke en wetenschappelijke waarde van internationale data over onderwijsprestaties, in: *Pedagogische Studiën*, vol. 88, 2011, pp. 122-131.

Duivenboden, H. van, en M. Lips (red.), *Klantgericht werken in de publieke sector – Inrichting van de elektronische overheid*, Lemma, Utrecht, 2001.

Duijvendak, M., *Wijde horizon, scherpe blik- Rurale geschiedenis en Noord-Nederland; Over rurale problematiek, sociale wetenschap en geschiedschrijving*, Groningen, 2002 (oratie).

Dunleavy, P., *Democracy, Bureaucracy and Public Choice*, Hemel Hampstead, 1991.

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

Eijk, D. van, *Een atlas van de Nederlandse steden*, NRC Handelsblad, Rotterdam, 1998.

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

Elsacker, W.J. van, *Roles of performance measurement in local government – Explorative case studies in Dutch local government*, EPM Sneek, 2007 (diss. RUG).

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, G., *Fatale remedies*, Pallas Publ., Amsterdam University Press, 2009.

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

Esch, R. van, 10 Jaar bedrijfsvergelijking zuiveringsbeheer bij de waterschappen, Bogt, H. ter, P. Jansen en S. Tillema (red.), *Met het oog op de publieke zaak*, RUG, Groningen, 2010, pp. 115-123.

Esfahani, E., *De kracht van betekenis*, Ambo, Baarn, 2017.

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.

Euwes, M., De Ziekenhuis Monitor als afgeleide van de Meyer Monitor; de praktische waarde voor het algemene ziekenhuis, in: Bossert, H., L. Meuleman, J. Weber en N. Zwennes (red.), *Public governance- Van theorie naar praktijk. Voorbeelden van vernieuwing in publieke organisaties*, Sdu, Den Haag, 2006, pp. 87-103.

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.

Fleurke, F., *Brief van de Gemeente aan het Rijk. Over de bestuurlijke verhouding tussen het rijk en de gemeente*, VNG, Den Haag, 2004.

Fleurke, F., *Organische bijstand – Analyse van het interventierepertoire voor gemeenten in aanhoudende bestuurscrisis*, Min. van BZK, Den Haag, 2008.

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.

Frederickson, H.G., *The Public Administration theory Primer*, Westview Press, boulder, 2003.

Frey, B.S. & M. Osterloh, Onderzoeksevaluaties: verborgen kosten, twijfelachtige voordelen en betere alternatieven, in: Jansen, Th., G. van den Brinken 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.

Fukuyama, F., *Het bouwen van een staat*, Uitgeverij Contact, Amsterdam, 2005.

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.

Gerritsen, E., *De slimme gemeente nader beschouwd*, Amsterdam University Press, Amsterdam, 2011.

- Goedee, J. en A. Entken, *(Ont)keten- Implementeren van werken met ketens*, Lemma, Den Haag, 2006.
- Goodijk, R., *Falend toezicht in semipublieke organisaties?*, Van Gorcum, Assen, 2012.
- Goodsell, Ch., *The Case for Bureaucracy*, Chatham House Publ., Chatham, 1985.
- Goudsblom, J., *Pasmunt*, De Arbeiderspers, Amsterdam, 1958 (1976).
- Goudsblom, J., *Reserves*, Meulenhoff, Amsterdam, 1998.
- Gow, J.L. & C. Dufour, Is the New Public Management a Paradigm? - Does it matter?, in: *International Review of Administrative Sciences*, vol. 66, 2000, nr. 4, pp. 573-597.
- 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.
- Grinwis, P., F. van der Meché en E. Spaans, *Weer in contact komen – Eindrapport Onderzoekcommissie Toezichtfunctie VUmc*, Amsterdam, 7 december 2012.
- Grit, K., *Economisering als probleem*, Van Gorcum, Deventer, 2000.
- Groenendijk, N.S., Benchmarking in de Europese Unie, in: *Tijdschrift voor Openbare Financiën*, 2004, nr. 3, pp. 139-153.
- Groenman, S. en H. de Jager (red.), *Staalkaart van de Nederlandse sociologie*, Van Gorcum, Assen, 1970.
- Gronouwe, J. en S. Spierings, *Concernsturing gemeente Heerlen – ‘Het is nu of nooit’*. Eindrapport, PwC, 2011.
- Groot, H. de (red.), *Benchmarking in de publieke sector*, Sdu, Den Haag, 2004.
- Groot, H. de, Benchmarking als beleidsinstrument, in: SCP, *Doelmatigheid in de publieke sector in perspectief – Overdenkingen bij het afscheid van Jos Blank*, SCP, Den Haag, 2001, pp. 15-21.
- Groot, H. de, en R. Goudriaan, *De productiviteit van de overheid*, Academic Service, Schoonhoven, 1991.
- Groot, H. de, *Evidence-based Public Management*, UTwente, Enschede, 2010 (oratie).
- Groot, H. de, *Tussen markt en politieke arena*, VU, Amsterdam, 2000 (oratie).
- Groot, J.P., *De leefbaarheid van de dorpen in de gemeente Borger*, Landbouwhogeschool Wageningen, Wageningen, 1969.
- Groot, J.P., *Kleine plattelandskernen in de Nederlandse samenleving – Schaalvergroting en dorpsbinding*, Landbouwhogeschool, Wageningen, 1972.
- Groot, L.C.M. & T. Buddingh, The Influence of New Public Management Practices on Product Costing and Service Pricing Decisions in Dutch municipalities, in: *Financial Accountability & Management*, vol. 20, 2004, nr. 4, pp. 421-443.
- Gunsteren, H. van, *Culturen van besturen*, Boom, Amsterdam, 1994.
- Gunsteren, H. van, en E. van Ruyven (red.), *Bestuur in de ongekende samenleving*, Sdu uitgevers, Den Haag, 1995.
- Hajer, M., Landschappen lezen, in: Bijl, R., R. van den Brinke e.a. (red.), *Opvallend gewoon - Het bijzondere van Nederland*, SCP, Den Haag, 2013, pp. 176-184.
- Hakvoort, J. & H. Klaassen, Benchmarking in non-profit organisaties, in: *Beleidsanalyse*, 1999, nr. 3, pp. 11-20.
- Hakvoort, J. & H. Klaassen, International Benchmarking of Public Organizations: a Critical Approach, in: Pollitt, Chr., S. van Thiel & V. Homburg (eds.), *New Public Management in Europe – Adaptation and Alternatives*, Palgrave, Londen, 2007, pp. 107-135.
- Hakvoort, J.L.M. en H.L. Klaassen, *Bedrijfsvoeringstechnieken voor overheid en non-profitorganisaties*, Sdu, Den Haag, 2004/2013.
- Hart, P. 't, P. de Jong en A.F.A. Korsten (red.), *Groepsdenken in het openbaar bestuur*, Samsom, Alphen, 1993.
- 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.
- Have, S. ten, W.D. ten Have, H. de Jong e.a., *Het managementmodellenboek*, Berenschot/ Elsevier, Den Haag, 1999.
- Hazeu, C., De groeten uit Grollo - De stand van het platteland in Nederland, in: *Tijdschrift voor sociaal-wetenschappelijk onderzoek in de landbouw*, jrg. 19, 2004, nr. 4, pp. 217-223.
- Heer, J. de, M. Heite en J. Wels, *Bestuurlijk innoveren*, Sdu, Den Haag, 1998.
- Helden, G. van, *BBI in de praktijk*, RUG, Groningen, 1998.
- 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, & S. Tillema, In Search of a Benchmarking Theory for the Public Sector, in: *Financial Accountability & Management*, vol. 21, 2005, nr. 3, pp. 337-361.

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: *Maandblad voor Accountancy en Bedrijfseconomie*, jrg. 79, 2005, nr. 11, pp. 574-583.

Helden, G.J. van, en S. Tillemans, Naar een benchmarkingtheorie voor de publieke sector, in: *Tijdschrift voor Openbare Financien*, 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. Tillemans en R. van Esch, Benchmarking in de Publieke Sector. Ervaren met de bedrijfsvergelijking zuiveringsbeheer, in: *Economistenblad*, jrg. 28, 2005, juni, pp.13-16.

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

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.

Herweijer, M. en R. Fraanje (red.), *Samen werken aan bestuurskracht- Intergemeentelijke samenwerking onderzocht*, Kluwer, Deventer, 2011.

Heusinkveld, S., *Surges and sediments - Organization concepts between transience and continuity*, KUN, Nijmegen, 2004 (diss.).

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

Hofstee, E.W., Veranderend platteland, in: Groenman, S. en H. de Jager (red.), *Staalkaart van de Nederlandse sociologie*, Van Gorcum, Assen, 1970, pp. 254-275.

Hood, C. & A. Dunsire, *Bureaometrics*, Gower, Westmead, 1981.

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.

Hoogerwerf, A. en M. Herweijer (red.), *Overheidsbeleid*, Kluwer, Deventer, 2008/2014 (9e druk).

Hoogerwerf, A., *Politiek als evenwichtskunst*, Samsom, Alphen, 1995.

Hoogwout, M., De benchmark BWT: monitor van een sector, in: *Bouwregels in de praktijk*, jrg. 2, mei 2007, nr. 5, pp. 16-18.

Hoogwout, M., *De rationaliteit van de klantgerichte overheid*, Uitgeverij Réunion, Nieuwegein, 2010.

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

Hospers, G.J. en N. Reverda, *Krimp, het nieuwe denken*, Boom Lemma, Amsterdam, 2012.

Houweling, P., *Beschikkingverlening in de beschikkingenfabriek*, RUG, Groningen, 1996.

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.

Inspectie voor de Gezondheidszorg, *Met de patiënt voor ogen*, Amsterdam, 2013.

Jansen, D.W.M., *De leefbaarheid van Borsele – Een onderzoek naar de meningen en opvattingen van de bevolking over een aantal woon- en leefaspecten in de gemeente*, Provinciaal Opbouworgaan Stichting Zeeland, Middelburg, 1972.

Jansen, E.P., New public management perspectives on performance and the use of performance information, in: *Financial Accountability and Management*, vol. 24, 2008, nr. 2, pp. 169-191.

Jansen, E.P., *Use, Needs and Determinants of Performance Information: Case Studies in Local Social Security*, Labyrint, Capelle a.d. IJssel, 2000.

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

Janzen, L. en C. Hoeben, Ontwikkeling gemeentelijke woonlasten 1998-2013, in: *Bank & Gemeente*, juli/aug. 2013, pp. 26-29.

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, P. de, A.F.A. Korsten en H. van der Made (red.), *Ziekenhuizen: besluitvorming en management*, Van Gorcum, Assen, 1992.

Jong, P.O. de, *Beroep op tijd – Een onderzoek naar het tijdsbeslag van beroepsprocedures in eerste aanleg in het bestuursrecht*, Den Haag, 2004.

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.

Kampermann, A. en A.F.A. Korsten, *Competentiemanagement*, WEKA, Amsterdam, 2007.

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.

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.

Kickert, W.J.M. & R.J. Stillman (eds.), *The Modern State and its Study: New Administrative Sciences in a Changing Europe and United States*, Edward Elgar, Cheltenham, 1999.

Kickert, W.J.M. & Th.A.J. Toonen, Public Administration in the Netherlands: Expansion, Diversification and Consolidation, in: *Public Administration*, vol. 64, 2006, nr. 4, pp. 969-987.

Kickert, W.J.M. (ed.), *Public Management and Administrative Reform in Western Europe*, Edward Elgar, Cheltenham, 1997.

Kickert, W.J.M. (red.), *Veranderingen in management en organisatie bij de rijksdienst*, Samsom, Alphen, 1993.

Kickert, W.J.M., E.H. Klijn & J.F.M. Koppenjan (eds.), *Managing Complex Networks*, Sage, Londen, 1997.

Kickert, W.J.M., *Europa en de crisis- Hoe Europese overheden omgingen met de bankencrisis, economische crisis en begrotingscrisis in 2008-2011*, Erasmus Universiteit Rotterdam, 2012.

Kickert, W.J.M., *History of Governance in the Netherlands: Continuity and Exceptions*, Reed-Elsevier, Den Haag, 2005.

Kickert, W.J.M., *Public Management Reforms in The Netherlands*, Eburon, Delft, Delft, 2000.

KING, *Vergelijken, verbazen en verbeteren- Gemeentelijke benchmarkgids 2010*, Den Haag, 2010.

Kingdon, J.W., *Agenda, Alternatives, and Public Policies*, Longman, New York, 2003 (2^e editie) .

Klerk, A. de, *Het geheim van de goede corporatie*, Ondersteboven Advies en Raeflex, Den Haag, 2012.

Klunder, R., Hoe presteert de overheid?, in: *Intermediair*, jrg. 25, 17 november 1989, pp. 47-51.

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

Kolthoff, E., *Ethics and New Public Management*, Boom, Den Haag, 2007.

Konisky, D.M. & C. Reenock, Case Selection in Public Management Research: Problems and Solutions, in: *Journal of Public Administration Research and Theory*, vol. 23, april 2013, nr. 2, pp. 361-395.

Konzendorf, G. & R. Hathaway, Towards benchmarking in Germany, in: Fenna, A. & F. Knüpling (eds.), *Benchmarking in Federal Systems*, Australian Government, Melbourne, 2010, pp. 111-123.

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

Koppenjan, J., *The New Public Governance in public service delivery*, Eleven, Den Haag, 2012.

Korsten, A. e.a., *Samen en toch apart*, Open Universiteit, Heerlen, 2002.

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. en P. de Goede (red.), *Bouwen aan vertrouwen in het openbaar bestuur*, Elsevier, Den Haag, 2006.

Korsten, A., De Nederlandse overheidsmanager en de 'nieuwe zakelijkheid', in: Bouckaert, G., A. Hondgehem 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 bestuursschout?, 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, Van beleidswetenschap naar kennissamenleving - Voortgang, vooruitgang en achteruitgang in de beleidswetenschap, in: *Beleidswetenschap*, jrg. 20, 2006, nr. 4, pp. 34-72.

Korsten, A.F.A. en Th.A.J. Toonen (red.), *Bestuurskunde: hoofdfiguren en kernthema's*, Stenfert Kroese, Leiden, 1988.

- Korsten, A.F.A., 10 Jahre Tilburger Modell – Mekka der Oeffentlichen Verwaltung?, in: Mix, U. & M. Herweijer (Hrsg.), *10 Jahre Tilburger Modell*, Sachbuch Verlag Kellner, Bremen, 1996, pp. 21-37.
- Korsten, A.F.A., A.F.M. Bertrand, P. de Jong en J.M.L.M. Soeters (red.), *Internationaal-vergelijkend onderzoek*, Vuga, 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., Benchmarking, in: *Binnenlands Bestuur*, 7 juni 2002.
- Korsten, A.F.A., Bestuurlijke vernieuwing in gemeenten: een illusie?, in: *Beleidswetenschap*, 1997, nr. 1, pp. 41-54.
- Korsten, A.F.A., *Bestuurskunde als avontuur*, Kluwer, Deventer, 1988.
- Korsten, A.F.A., Das Tilburger Modell oder Tilburg als neues Mekka der Oeffentlichen Verwaltung, in: Kleinfeld, R. & A.F.A. Korsten, *Konzern Stadt- Neue Steuerungsmodelle in den Kommunalverwaltungen*, Krefeld, 1995.
- 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., *Grote klasse! – Op zoek naar excellente ambtenaren en leiderschap*, Beljon + Westerterp, Heerlen, 2005.
- Korsten, A.F.A., K. Abma en J.M.L.R. Schutgens, *Bestuurskracht van gemeenten – Meten, vergelijken en beoordelen*, Eburon, Delft, 2007.
- Korsten, A.F.A., *Lokaal bestuur in verandering*, Open Universiteit, Heerlen, maart 2009a.
- Korsten, A.F.A., M. Schoenmaker, e.a., *Onder burgemeesters*, Boom Lemma, Den Haag, 2013.
- Korsten, A.F.A., Monisme en de uitblijvende struikelpartijen van Vlaamse schepenen, in: *Bestuurswetenschappen*, jrg. 66, 2012, nr. 5, pp. 32-52.
- Korsten, A.F.A., *Orde en veiligheid in het publieke domein*, Open Universiteit, Heerlen, sept. 2009b.
- 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.
- Kruijff, J.A.M. de, en M.J.F. Wouters, Prestatiemeting en vergelijkende kengetallen meten is ook maar meten, in: *Tijdschrift voor Openbare Financiën*, 2004, nr. 3, pp. 121-139.
- Kuhry, B. en F. de Kam, *Waar voor ons belastinggeld? Prijs en kwaliteit van publieke diensten*, SCP, Den Haag, 2012.
- Kuhry, B. en J.J.J. Jonker, *Maten voor gemeenten 2009*, SCP, Den Haag, 2009.
- Kuhry, B. en V. Veldheer, *Maten voor gemeenten 2004*, SCP, Den Haag, 2004.
- Kuhry, B. m.m.v. V. Veldheer en J. Stevens, *Maten voor gemeenten 2005*, SCP, Den Haag, 2005.
- Kuhry, B., E.J. Pommer en C.A. de Kam, Prestaties van de publieke sector in internationaal perspectief, in: *Tijdschrift voor Openbare Financiën*, jrg. 36, 2004, nr. 6, pp. 275-284.
- Kwakman, F. en R. Smeulders (red.), *Groot Innovatie Modellenboek*, Van Duuren Management, 2013.
- Laar, S. van de, *Samen sterker – Samenwerking tussen gemeenten geanalyseerd*, Eburon, Delft, 2010.
- Lammers, C.J., *Organisaties vergelijkenderwijs*, Het Spectrum, Utrecht, 1983.
- Lammers, C.J., *Organiseren van bovenaf en van onderop*, Het Spectrum, Utrecht, 1993.
- 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).
- Lavertu, S. & D.P. Moynihan, The Empirical Implications of Theoretical Models: A Description of the Method and an Application to the Study of Performance Management Implementation, in: *Journal of Public Administration Research and Theory*, vol. 23, april 2013, nr. 2, pp. 333-361.
- Leeuw, F.L., *Gedragmechanismen achter overheidsinterventies en rechtsregels*, UM, Maastricht, 2008 (oratie).
- 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.
- Liedelmeijer, K., G. Marlet, J. van Iersel e.a., *De leefbaarometer – Rapportage instrumentontwikkeling Amsterdam en Utrecht*, RIGO, Amsterdam, 2008.
- Linker, P.J., Verantwoording over prestaties, in: Bovens, M.A.P. en Th. Schillemans (red.), *Handboek publieke verantwoording*, Boom Lemma, Den Haag, 2009, pp. 55-79.

Loos, E., *Innovatie aan het werk – Onderzoek op locatie naar succes- en faalfactoren van innovatieprojecten bij Nederlandse gemeenten vanuit een multi-actor perspectief*, Boom, Den Haag, 2006.

Lustgraaf, R. van de, en J. Haagsma, Waarde van een index, in: Andersson Elffers Felix, *Jaarboek 2002*, Utrecht, 2002, pp. 193-197.

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

Maarse, H., *Markthervorming in de zorg*, Universitaire Pers Maastricht, Maastricht, 2011.

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

Mak, G., *Hoe God verdween in Jorwerd*, Atlas, Amsterdam, 1996.

Marlet, G. e.a., *Atlas voor gemeenten 2012*, Atlas voor gemeenten, Utrecht, 2012.

Marlet, G. e.a., *Atlas voor gemeenten 2013*, Atlas voor gemeenten, Utrecht, 2013.

Marlet, G. en C. van Woerkens, *Atlas van gemeenten 2003*, Nyfer, Breukelen, 2003.

Marlet, G. en C. van Woerkens, *De oriëntatie van de bevolking van de Achterhoek*, Utrecht, 2013.

Marlet, G. en R. Ponds, *Scoren in Spangen – De effectiviteit van tien jaar investeren in de Rotterdamse wijk Spangen*, Atlas voor gemeenten, Utrecht, 2012.

Marlet, G., *De aantrekkelijke stad*, VOC uitg., Nijmegen, 2009.

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

Martin, S.J., J.D. Downe, C.L. Grace and S.M. Nutley, Utility, Validity and Evidence Based Policy Making: The Development and Impact of Performance Improvement Regimes in Local Public Services, in: *Evaluation*, vol. 16, 2010, nr. 1, pp. 31–42.

McLaughlin, K., S. Osborne & E. Ferlie (eds.), *New public management*, Routledge, Londen, 2002.

Meer, A.D. van der, Benchmarking, in: Eversdijk, A. en J. Schutgens (red.), *Openbaar bestuur op expeditie – Verkenningen van bestuurlijke vernieuwingspogingen*, Boom Lemma, Den Haag, 2010, pp. 104-113.

Meier, K.J. & L.J. O’Toole jr., Subjective Organizational Performance and Measurement Error: Common Source and Spurious Relationships Bias, in: *Journal of Public Administration Research and Theory*, vol. 23, april 2013, nr. 2, pp. 429-457.

Meier, K.J. & L.J. O’Toole jr., The Proverbs of New Public Management: Lessons from an Evidence-based Research Agenda, in: *The American Review of Public Administration*, vol. 39, jan. 2009, pp. 4-22.

Meijer, A.J., G.J. Brandsma en S. Grimmelikhuisen, Transparantie als fictieve verantwoording, in: *Bestuurswetenschappen*, jrg. 64, 2010, nr. 4, pp. 8-28.

Minderman, G.D. (red.), *Governance in het onderwijs: een nieuwe balans?*, Sdu, Den Haag, 2007.

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, *Kabinetsstandpunt over doe- democratie*, Den Haag, 2013.

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, *Staat van het bestuur 2008*, Den Haag, 2008.

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, *Staat van het bestuur 2010*, Den Haag, 2010.

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, *Staat van het bestuur 2012*, Den Haag, 2012.

Mintzberg, H., Managing Government , Governing Management, in: *Harvard Business Review*, may-june 1996, pp. 75-83.

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.

Mintzberg, H., *The Structuring of Organizations: A Synthesis of the Research*, Prentice Hall, Englewood Cliffs, 1979.

Moore, M.H., *Creating Public Value*, Harvard University Press, Cambridge, MA, 1995.

Moynihan, D.P., Advocacy and Learning: An Interactive-Dialogue Approach to Performance Information Use, in: Dooren, W. van, en S. van de Walle (eds.), *Performance Information in the Public Sector- How it is used*, Palgrave, Londen, 2008a, pp. 24-42.

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

Munters, Q.J., *De stille revolutie op het agrarisch platteland boeren en openbaar bestuur 1917- 1986*, Van Gorcum, Assen, 1989.

Nederlandse Zorgautoriteit, *Impactanalyse verpleeghuiszorg 2017*, NZa, Den Haag, 2017.

Nelen, H., *Evidence maze; het doolhof van het evaluatieonderzoek*, Universiteit Maastricht, 2008 (oratie).

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

Nelissen, N.J.M., *De stad – Inleiding in de urbane sociologie*, Van Loghum Slaterus, Deventer, 1974.

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

Niskanen, W., *Bureaucracy and Public Economics*, Edward Elgar, Cheltenham, 1971b.

Niskanen, W., *Bureaucracy and Representative Government*, Aldine-Atherton, Chicago, 1971a.

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. en J. Sterrenburg, Publieke professionals en verantwoordingsdruk, in: Bovens, M.A.P. en Th. Schillemans (red.), *Handboek publieke verantwoording*, Boom Lemma, Den Haag, 2009, pp. 231-255.

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

Noordegraaf, M., Prestatiemanagement, in: Noordegraaf, M., K. Geuijen en A. Meijer (red.), *Handboek Publiek Management*, Boom Lemma, Den Haag, 2011, pp. 365-394.

Noordegraaf, M., *Professioneel bestuur*, Utrecht, 2008 (oratie)

Noordhoek, M., *Municipal Benchmarking – Organisational learning and network performance in the public sector*, Eburon, Delft, 2013.

Nutley, S.M., I. Walter & H.T.O. Davies, *Using Evidence - How Research Can Inform Public Services*, The Polity Press, Bristol, 2007.

NVAO, *Beoordelingskaders accreditatiestelsel hoger onderwijs*, Den Haag, 6 dec. 2010.

OECD, *Learning for Tomorrow's World: First Results from PISA 2003*, OECD, Parijs, 2004.

OECD, *PISA 2006 Science Competencies for Tomorrow's World*, OECD, Parijs, 2007.

Oerlemans, L., *Netwerken en innovatie: een perfecte match?*, Universiteit van Tilburg, Tilburg, 2007.

Oevering, F., *Demografische krimp – De nieuwe realiteit in perspectief*, Rabobank, Utrecht, 2010.

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

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

Otten, M., *Verstrikt in grote projecten – Hoe de stadhuizen in Amsterdam en Apeldoorn tot stand kwamen*, VNG, Den Haag, 2000.

Out, N., Bewoners maken zelf de leefbaarheid, in: *Agora Magazine*, 2011, nr. 4, pp. 18-21.

Overmeeren, A. van, *Visitatie van woningcorporaties*, OTB, Delft, 2007.

Pawson, R., *Evidence-based Policy – A Realist Perspective*, Sage, Londen, 2006.

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.

Permentier, M., J. Kullberg en L. van Noije, *Werk aan de wijk - Een quasi-experimentele evaluatie van het krachtwijkenbeleid*, SCP, Den Haag, 2013.

Pestman, P., NPM-instrumenten in de praktijk – Van ideologisch debat naar praktische toepassing, in: *Bestuurskunde*, jrg. 16, 2007, nr. 3, pp. 104-114.

Pestman, P., NPM-instrumenten in de praktijk, in: *Bestuurskunde*, jrg. 16, 2007, nr. 3, pp. 104-114.

Peters, B.G., *The Politics of Bureaucracy*, Routledge, Londen, 2001 (5^e editie).

Pfeffer, J. & R.I. Sutton, *Hard Facts, Dangerous Half-truths & Total Nonsense*, Harvard Business School Press, Boston, 2006.

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

Poerstamp, R.J.C., A. van Mourik en A.C. Veldman, *Benchmarking in de zorg*, Pwc/Elsevier, Maarssen, 2007.

Pollitt, Chr., S. van Thiel & V. Homburg (eds.), *New Public Management in Europe – Adaptation and Alternatives*, Palgrave, Londen, 2007.

Pommer, E. en E. Eggink, *Publieke dienstverlening in perspectief*, SCP, Den Haag, 2010.

Pommer, E. en I. Ooms, *Maten voor gemeenten 2011 – Een analyse van de prestaties van de lokale overheid*, SCP, Den Haag, 2011.

Pommer, E., I. Ooms, A. van der Torre en S. Jansen, *Maten voor gemeenten 2012 – Prestaties en uitgaven van de lokale overheid in de periode 2005-2010*, SCP, Den Haag, nov. 2012.

Ponds, R., C. van Woerkens en G. Marlet, *Atlas van kansen voor de grensregio's Zuid-Limburg, Tongeren, Luik en Aken*, Atlas voor gemeenten, Utrecht, 2012.

Poorthuis, A., A. van der Aa, e.a., *Ketens en netwerken – Een zoektocht naar samenhang*, Lemma, Utrecht, 2003.

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

PricewaterhouseCoopers, *Integratie van milieubeleid: degraderen of excelleren - Lessen uit een benchmark van externe integratie bij vijf grote gemeenten*, Utrecht, 2000.

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, *Bevolkingsdaling – Gevolgen voor bestuur en financiën*, Den Haag, 2008.

Raad voor het openbaar bestuur, *Loslaten in vertrouwen*, Den Haag, 2012.

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

Raad voor Maatschappelijke Ontwikkeling, *Terugtrekken is vooruitzien*, Den Haag, 2013.

- Rainey, H.G., R.W. Backoff & C.H. Levine, Comparing Public and Private Organizations, in: *Public Administration Review*, vol. 36, 1976, nr. 2, pp. 233-244.
- Rainey, H.G., *Understanding and Managing Public Organizations*, Jossey-Bass, San Francisco, 2003.
- Rekenkamer Amsterdam, *Meetbaarheid gemeten – Beoordeling jaarplan 2007*, Amsterdam, 2006.
- Reverda, F., *Over krimp – Een sociologisch perspectief op bevolkingsdaling*, Neimed, Heerlen, 2011.
- Ridder, J. de, *Een goede raad voor toezicht*, Boom, Den Haag, 2004.
- Rieper, O., F.L. Leeuw & T. Ling (eds.), *The Evidence Book*, Transaction Publ., Londen, 2009.
- RIGO en Atlas voor gemeenten, *Onderscheid in leefbaarheid – Ontwikkeling van de leefbaarheid 2010-2012*, Utrecht, 2012.
- Rijn, R. van, en M. van Twist, Verantwoorde vernieuwing? Innovatie van publieke verantwoording, in: Bovens, M.A.P. en Th. Schillemans (red.), *Handboek publieke verantwoording*, Boom Lemma, Den Haag, 2009, pp. 231-255.
- Rogge, N., S. de Jager en G. Naets, Burgertevredenheid met lokaal gemeentebelid: meten is weten, in: *Beleid en Maatschappij*, 2013, jrg. 40, nr. 4, pp. 401-418.
- Rose, A. & A. Lawton (eds.), *Public Services Management*, Prentice Hall, Harlow, 1999.
- Rossi, P.H., M.W. Lipsey & H.E. Freeman, *Evaluation*, Sage, Londen, 2004 (7^e editie).
- Rousseau, D.M. (ed.), *The Oxford Handbook of Evidence-based Management*, Oxford University Press, Oxford, 2012.
- Rousseau, D.M., Is There Such a Thing as 'Evidence-based Management'?, in: *Academy of Management Review*, vol. 31, 2006, nr. 2, pp. 256-269.
- 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.
- Schedler, K., Wie ontwikkelt sich die international Debatte um das New Public Management, in: Jann, W., M. Röber & H. Wollmann (Hrsg.), *Public Management – Grundlagen, Wirkungen, Kritik*, Edition Sigma, Berlijn, 2006, pp. 95-109.
- Schillemans, Th. en M. Bovens, Publieke verantwoording 2.0: sober maar scherp, in: Bovens, M.A.P. en Th. Schillemans (red.), *Handboek publieke verantwoording*, Boom Lemma, Den Haag, 2009, pp. 275-295.
- Schillemans, Th., *Verantwoording in de schaduw van de macht*, Lemma, Utrecht, 2007.
- Schillemans, Th., Visitaties als vorm van verantwoording, in: Bovens, M.A.P. en Th. Schillemans (red.), *Handboek publieke verantwoording*, Boom Lemma, Den Haag, 2009, pp. 161-181.
- 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.
- Schutgens, J.M.L.R., *Met vreemde ogen kijken*, Wolf Legal Publ., Nijmegen, 2014.
- Senge, P., *De vijfde discipline*, Scriptum, Schiedam, 1992.
- 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.
- Shafritz, J.M. & A.C. Hyde (eds.), *Classics of Public Administration*, Harcourt Brace College Publ., Fort Worth, 1997.
- Shapiro, S., *Best practices zijn dom*, Het Spectrum, Utrecht, 2012.
- Shillabeer, A. e.a. (eds.), *Evidence-Based Public Management*, M.E. Sharpe, Londen, 2011.
- Simsek, S.G., De lerende organisatie, in: *Management in overheidsorganisaties*, Samsom, Alphen, april 1998, G2170.
- Smit, A., *Sociale staat van Zeeland- Provinciaal rapport 2009*, Scoop, Middelburg, 2009.
- Smith, P., Outcome-Related Performance Indicators and Organization Control in the Public Sector, in: *British Journal of Management*, vol. 4, 1993, nr. 2, pp. 135-151.
- Smith, P.C., On the Unintended Consequences of Publishing Performance Data in the Public Sector, in: *International Journal of Public Administration*, 1995, nr. 2/3, pp. 277-310.
- Smith, P.C., Quantitative Approaches towards Assessing Organizational Performance, in: Boyne, G.A., K.J. Meier, L.J. O'Toole jr & R.M. Walker (eds.), *Public Sector Performance – Perspectives on Measurement and Management*, Cambridge UP, Cambridge, 2006, pp. 75-92.
- Snellen, I.Th.M., *Grondslagen van de bestuurskunde – Een essay over haar paradigma's*, Lemma, Den Haag, 2007.
- Sociaal en Cultureel Planbureau, *Op weg met de Wmo - Evaluatie Wet maatschappelijke ondersteuning 2007-2009*, SCP, Den Haag, 2010.
- Sociaal en Cultureel Planbureau, *Sturen op geluk*, SCP, Den Haag, 2012.

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

Soetenhorst, B., *Het wonder van de Noord/Zuidlijn – Het drama van de Amsterdamse metro*, Uitgeverij Bert Bakker, Amsterdam, 2011.

Starling, G., *Managing the Public Sector*, Wadsworth Cengage Learning, Boston, 2011 (9e editie).

Steenbekkers, A. en L. Vermeij (red.), *De dorpenmonitor*, SCP, Den Haag, 2013.

Steenbekkers, A., C. Simon en V. Veldheer (red.), *Thuis op het platteland – De leefsituatie van platteland en stad vergeleken*, SCP, Den Haag, 2006.

Steenbekkers, A., C. Simon, L. Vermeij en W. Spreeuwers, *Het platteland van alle Nederlanders*, SCP, Den Haag, 2008.

Stevens, R., *Met open vizier- Auditing als stimulerende interventie*, Tilburg, 2012 (diss.).

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.

Sutcliffe, S. & J. Court, *Evidence-based policy-making: What it is? How does it work? What relevance for developing countries*, Overseas Development Institute, nov. 2005 (paper).

Swanborn, P.G., *Evalueren*, Boom, Amsterdam, 2007.

Talbot, C., Performance Management, in: Ferlie, E, L.L. Lynn jr & C. Pollitt (eds.), *The Oxford Handbook of Public Management*, Oxford UP, Oxford, 2005, pp. 491-521.

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

Teisman, G., *Publiek management op de grens van orde en chaos*, Academic Service/ Sdu, Den Haag, 2007.

Termeer, C.J.A.M., M. van Lieshout e.a., *Politieke besluitvorming over het Landbouwontwikkelingsgebied Witveldweg in de Gemeente Horst aan de Maas*, Wageningen, 2009.

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, *Bestuurskundig onderzoek – Een methodologische inleiding*, Uitgeverij Coutinho, Bussum, 2007.

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.

Thiel, S. van, Prestatiemeting in de publieke sector, in: *Overheidsmanagement*, jan. 2009, nr. 1, pp. 20-23.

Thissen, F., J. Drooglever Fortuijn en P. van Konijnenburg, *Kwetsbaar of succesvol? – De woonsituatie van ouderen in de dorpen in de Kop van Noord-Holland*, Universiteit van Amsterdam, Amsterdam, 2003.

Thissen, F., Wat houdt een (krimp) dorp leefbaar en vitaal?, in: *Real Estate Magazine*, 2010, nr. 13, pp. 23-26.

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, Benchmarken, 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.

Titulaer, R., K. Lammers e.a., *Kwaliteit bij burgerzaken: stap voor stap op weg naar rekenschap – Rekenschapsonderzoek naar resultaten en verklaringen van de kwaliteit van de dienstverlening van afdelingen burgerzaken*, mei 2001.

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

Tonkens, E. en Tj. Swierstra, Meetlust: de wil tot meten, in: Swierstra, Tj. en E. Tonkens (red.), *De beste de baas?*, Amsterdam University Press, Amsterdam, 2008, pp. 119-149.

Tonkens, E., *Mondige burgers, getemde professionals*, NIZW, Utrecht, 2004.

Tops, P., *Regimeverandering in Rotterdam*, Atlas, Amsterdam, 2007.

Tops, P.W. en G.W.M. van Vugt (red.), *Zoeken naar een modern bestuur – Het Tilburgs model en de logica van de burger*, Samsom, Alphen a.d. Rijn, 1998.

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

Triantafyllou, P., Benchmarking in the Public Sector – A Critical Conceptual Framework, in: *Public Administration*, vol. 85, 2007, nr. 3, pp. 829-846.

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.

Veenswijk, M. (ed.), *Organizing Innovation*, IOS Press, Amsterdam, 2005.

Veldheer, V., *Kantelend bestuur*, SCP, Rijswijk, 1994.

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.).

Velthuys, J., Neem het heft in eigen hand!, in: *Elsevier*, 24 aug. 2013, pp. 16-21.

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).

Visser, M., *Schoolkeuze procedures, kwaliteit en kwaliteitszorg: een project onderwijsvernieuwing Amsterdam 1994-2001*, RUG, Groningen, 2003.

VNG, *Het beeld van de burger*, Den Haag, juni 2005.

VNG, *Gemeentelijke kwaliteitshandvesten op maat*, Den Haag, november 2006.

Voerknecht, H. en M. van Bokhorst, *Benchmark actieve, gezonde, duurzame mobiliteit gemeenten*, CE Delft, 2018a.

Voerknecht, H. en M. van Bokhorst, *Uitkomsten benchmark duurzame, actieve, gezonde mobiliteit – Gemeente Enschede*, CE Delft, 2018b.

Vogelzang, P., Efficiency politie is moeilijk in cijfers te vangen, in: *De Volkskrant*, 6 april 2002.

Vollaard, B.A., *Performance indicators for police forces*, CPB Document nr. 31, Den Haag, 2003.

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, J. de, en M. van Dam, *Politiek-bestuurlijk management*, Samsom, Alphen a.d. Rijn, 1998.

Vries, J. de, en P. Bordewijk (red.), *Rijdende treinen en gepasseerde stations – Over Srebrenica, de kredietcrisis en andere beleidsfiasco's*, Van Gennep, Amsterdam, 2009.

Vries, M.S. de, *De ontmythologisering van het openbaar bestuur*, Nijmegen, 2005.

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.

Vries, M.S. de, P.S. Reddy & M.S. Hague (eds.), *Improving Local Government – Outcomes of Comparative Research*, Palgrave, Londen, 2008.

Vries, M.S. de, Performance measurement and the search for best practices, in: *International Review of Administrative Sciences*, vol. 76, 2010, nr. 2, pp. 313-330.

Vught, F.A. van, & D.F. Westerheijden, Multi-dimensional ranking: a new transparency tool for higher education and research: in: *Higher Education and Policy*, vol. 22, 2010, nr. 3, pp. 1- 25.

Vught, F.A. van, & F. Fiegele (eds.), *U-Multirank – Design and Testing the Feasibility of a Multidimensional Global University Ranking – Final Report*, CHERPS- Network, European Commission, Brussel, 2011.

Waalwijk, 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.

Whitford, A.B. & K.J. Meier, Methodological Innovations in Public Administration: A Symposium, in: *Journal of Public Administration Research and Theory*, vol. 23, april 2013, nr. 2, pp. 331-333.

Wijnen, G., M. Weggeman en R. Kor, *Verbeteren en vernieuwen van organisaties*, Samsom, Alphen, 1999.

Wijnen, N., Pionnen op het bordspel – De beïnvloeding van de doorwerking van beleidsevaluaties in de gemeentelijke praktijk, in: Bossert, H., L. Meuleman, J. Weber en N. Zwennes (red.), *Public governance- Van theorie naar praktijk. Voorbeelden van vernieuwing in publieke organisaties*, Sdu, Den Haag, 2006, pp. 155-168.

Willemse, R., Eigenzinnige wetenschap tegenover eigenwijs bestuur, in: Ridder, J. de, P.L. Polhuis en L.W.J.C. Huberts (red.), *Eigenzinnig bestuur*, Wolf Legal Publishers, Nijmegen, 2009, pp. 23-35.

Willemse, R., *Het bestaansrecht van de Nederlandse gemeente*, Eburon, Delft, 2001.

Wilson, J.Q., *Bureaucracy- What Government Agencies Do and Why They Do it*, Basic Books, New York, 1989.

Wingerden, D. van, P. van Eekeren, P. van Dijk en G. van der Pijl, Op weg naar het benchmarken van ICT, in: *Informatie*, okt. 2007, pp. 52-57.

Wittebrood, K., De prestatieparadox van politie en justitie, in: SCP, *Averechtse effecten*, Den Haag, 2002, pp. 63-68.

Woude, W. van der, *Financiële controle in het gemeenterecht – Een juridisch onderzoek naar de 'dualisering van de financiële functie'*, Kluwer, Deventer, 2011.

Wouw, D. van der, *Leefbaarheid - Themaport sociale staat van Zeeland*, Scoop, Middelburg, 2011.

WRR, *Bewijzen van goede dienstverlening*, WRR, Den Haag, 2004.

WRR, *Publieke zaken in de marktsamenleving*, Amsterdam University Press, Amsterdam, 2012.
Zaat, I., De dynamiek van het Tilburgs model: het vernieuwde Tilburgse model, in: *Management in overheidsorganisaties*, Samsom, Alphen, dec. 2000, B1140.
Zee, H. van der (red.), *Op weg naar een effectieve overheid*, Sdu uitgevers, 2005.
Zifcak, S., *New Managerialism*, Open University Press, Buckingham, 1994.

Bijlage 1 - KING: het Kwaliteits Instituut Nederlandse Gemeenten

KING is het Kwaliteits Instituut Nederlandse Gemeenten. KING is beheerder van de Standaard voor gemeentelijke benchmarks.

KING onderscheidt, zo lezen we letterlijk in een publicatie '*Gemeentelijke benchmarkgids 2010*', drie groepen benchmarks. 'Zo is er de groep 'aangemeld maar niet-gecertificeerd'. Niet-gecertificeerd zijn kan allerlei redenen hebben. Bijvoorbeeld dat een benchmark het afgelopen jaar niet is uitgevoerd of dat de benchmark nieuw is opgezet en nog geen hele cyclus heeft doorlopen. De benchmarks met het certificaat 'Benchmark' vormen een volgende groep. Deze benchmarks voldoen aan de 'KING-standaard'. Tot slot is er een categorie met benchmarks die extra energie stoppen in het ontwikkelproces voor de deelnemers. Deze benchmarks krijgen het aanvullende certificaat "Benchlearning".'

1 Certificaat KING-standaard

Wat houdt de KING-standaard in? 'Met de Standaard beschrijft KING de kwaliteitsprincipes voor gemeentelijke benchmarks. KING certificeert benchmarks op basis van deze Standaard. Benchmarks die hieraan voldoen, mogen gedurende één jaar het KING Benchmark Certificaat voeren. Ieder jaar vindt een nieuwe certificeringsronde plaats. Om in aanmerking te komen voor een KING Benchmark Certificaat moet een benchmark aan eisen voldoen.' De KING-publicatie '*Gemeentelijke benchmarkgids 2010*' noemt de volgende. We volgen de letterlijke tekst.

Voor en van gemeenten

'De benchmark wordt in samenspraak met de gemeenten opgezet en verder doorontwikkeld. Het eigenaarschap van de door de gemeente aangeleverde cijfers blijft in eerste instantie bij de gemeente. In het benchmarkproces wordt aandacht besteed aan het bestuurlijk comitment van de gemeente.' Aldus de brochure.

Kwaliteit van het onderzoek

'Binnen het kwantitatieve onderzoek wordt gebruik gemaakt van robuuste, duidelijke en heldere indicatoren. In het onderzoek wordt gebruik gemaakt van een bondige set van indicatoren. Het onderzoek is gebaseerd op een relevante populatie. De statistische gegevens worden gecontroleerd op validiteit en betrouwbaarheid.'

Leer en ontwikkelingsinstrument

De brochure: 'De benchmarkcyclus vindt plaats in een vertrouwelijke leeromgeving, de gegevens worden niet direct openbaar gesteld. Er vinden fysieke ontmoetingsmomenten plaats waar de deelnemers elkaar kunnen treffen om de inhoud en uitkomsten van de benchmark met elkaar te bespreken. Er wordt aandacht besteed aan het verhaal achter de cijfers, er wordt samen met de deelnemers een interpretatie en analyse van de uitkomsten gemaakt. In de benchmarkcyclus wordt expliciet aandacht geschonken aan de verbetermogelijkheden van de deelnemende gemeente. Deelnemende gemeenten worden in staat gesteld met regelmaat deel te nemen aan een benchmark; de benchmark wordt gedurende meerdere jaren gecontinueerd.'

Aansluiting bij Waarstaatjegemeente.nl

De indicatorenset van de benchmarks zal in de nabije toekomst worden afgestemd op de indicatoren van *Waarstaatjegemeente.nl*. Zo wordt dubbele uitvraag van gegevens bij gemeenten voorkomen.

2 Certificaat Benchlearning

De brochure: 'Het is van groot belang dat benchmarks zijn gericht op leren en ontwikkelen. De benchmarks met een certificaat van KING voldoen hieraan. De wijze waarop het leer- en ontwikkelingsproces is ingericht, kan per benchmark verschillen. Er zijn benchmarks die het leer- en ontwikkelingsproces bij de deelnemende gemeenten op een actieve en methodologische manier faciliteren. Deze benchmarks komen in aanmerking voor het KING Benchlearn Certificaat. Dit is een aanvulling op het Benchmark Certificaat. Om in aanmerking te komen voor dit aanvullende certificaat, moet de benchmark dus in elk geval voldoen aan alle eisen van het Benchmark Certificaat.' Daarnaast moet aan de volgende aanvullende eisen worden voldaan.

Leren van binnenuit en van buitenaf

De brochure: 'Op basis van de benchmark wordt een leer- en ontwikkelingsproces ingegaan, dat op een methodologisch manier is vormgegeven. In dit leerprogramma wordt onderscheid gemaakt tussen het leren van binnenuit en van buitenaf.'

Van binnenuit: 'het leren in een benchmarkcyclus is een groepsaangelegenheid. Het gaat hier om het leren van elkaar, binnenuit de groep. Dit op basis van een methodologie.'

Van buitenaf: 'op de onderdelen waarin de benchmarkgroep niet in de eigen leerbehoefte kan voldoen, wordt expertise van buitenaf gehaald. Bijvoorbeeld door iemand uit een geheel andere sector een gastcollege te laten verzorgen of door een nieuwe technologie te introduceren.' Aldus de brochure.'

3 Besluitvorming

De brochure: 'KING voert jaarlijks de toetsing uit. Producenten kunnen hun benchmark hiervoor aanmelden. De toetsing is kosteloos. De beschrijving van de Standaard voor gemeentelijke benchmarks, het Toetsingskader en de Toetsingsprocedure staan op de website van KING en op Waarstaatjegemeente.nl.'

'De uitvoering van de toetsing wordt gedaan door de afdeling Benchmark, ontwikkeling en innovatie (BOI) van KING. De afdeling wordt hierin geadviseerd door het Adviescollege Benchmarking. Het Adviescollege bestaat uit vertegenwoordigers van o.a. de VNG en de Vereniging van Gemeentesecretarissen (VGS). De directie van KING neemt de uiteindelijke beslissing over het toekennen van een certificaat.'

Bijlage 2: Overzicht van een aantal gemeentelijke benchmarks

Rubriek		Titel benchmark
1		<i>Algemeen bestuur en dienstverlening</i>
	1.1.	Benchmark publiekszaken
	1.2	Benchmark juridische functie
	1.3	Klantenmonitor vergunningen
	1.4	Benchmark Minder regels, meer service
	1.5	Benchmark Shared service
	1.6	Benchmark Staf en ondersteuning
	1.7	Benchmark Telefonische bereikbaarheid
2		<i>Onderwijs en jeugd</i>
	2.1	Benchmark Jeugdbeleid
	2.2.	Benchmark RMC Functie
3		<i>Maatschappelijke ondersteuning</i>
	3.1	Benchmark Re-integratie
	3.2	Benchmark Wmo-Resultaten
	3.3	Benchmark Wmo
	3.4	Benchmark Blik op werk
	3.5	Benchmark Omvang overheid welzijnsinstellingen
	3.6	Benchmark Wmo Kernindicatoren
	3.7	Benchmark WWB
4		<i>Publieke gezondheid</i>
	4.1	Benchmark GGD'en
5		<i>Milieu</i>
	5.1	Gemeentelijke benchmark afvalscheiding
	5.2	Benchmark Afvalinzameling
	5.3	Rioleringszorg
6		<i>Ruimtelijke ordening en volkshuisvesting</i>
	6.1	Bouw- en woningtoezicht
	6.2	RO Digitaal
	6.3	Benchmark Gemeentelijk va stgoed
	6.4	Benchmark Grondbedrijfstakingen
	6.5	Benchmark Geo
7		<i>Financiën en sport</i>
	7.1	Benchmark Belastingoverleg grote gemeenten
	7.2	Benchmark Pauw
	7.3	Benchmark Financiële functie
	7.4	Benchmark Uitvoering WOZ
8		<i>Bedrijfsvoering</i>
	8.1	ICT Benchmark
	8.2	Benchmark Omvang ambtelijk apparaat
	8.3	IT Benchmark Lokale overheid
	8.4	Benchmark Vip-com
	8.5	Benchmark Beloning en formatie
	8.6	Benchmark Beheer basisregistraties
	8.7	Benchmark Doelmatigheid subsidieverwerving
	8.8	Benchmark Business and IT Alignment
	8.9	Benchmark Informatiebeveiliging
	8.10	Benchmark Programma- en projectrisico
	8.11	Benchmark Meetmethodiek kwaliteit facilitaire activiteiten
	8.12	Benchmark Facilitaire kengetallen gemeenten
	8.13	Internetspiegel
Bron: KING, 2010		

**Bijlage 3: Scores van burgers op het item 'goed bestuur'
van de gemeente in 'De Staat' (Bron: gegevens Abma/Korsten, 2009)**

	Score op de vraag naar 'goed bestuur' van een gemeente in 'De staat'	gem. grootteklasse	inw.	goed be- stuur
1.	Opmeer	<25	11212	6,5
2.	Hatterem	<25	11630	6,4
3.	Strijen	<25	9195	6,3
4.	Ten Boer	<25	7164	6,3
5.	Bunschoten	<25	19475	6,2
6.	Heemskerk	25-50	38006	6,2
7.	Leek	<25	19004	6,2
8.	Staphorst	<25	15899	6,1
9.	Dalfsen	25-50	26541	6,1
10.	Vlaardingen	>50	72553	6,1
11.	Bloemendaal	<25	17028	6,0
12.	Korendijk	<25	10850	6,0
13.	Ermelo	25-50	26305	6,0
14.	Zoeterwoude	<25	8400	6,0
15.	Meerssen	<25	19762	6,0
16.	Ermelo	25-50	26251	6,0
17.	Krimpen a.d. IJssel	25-50	28783	6,0
18.	Oss	>50	76416	6,0
19.	Wierden	<25	23553	6,0
20.	Heemstede	>25	25678	6,0
21.	Haaksbergen	<25	24348	5,9
22.	Oud-Beijerland	<25	24019	5,9
23.	Deurne	25-50	31841	5,9
24.	Heerhugowaard	25-50	49800	5,9
25.	Wijchen	25-50	39479	5,9
26.	Capelle a.d. IJssel	>50	65374	5,9
27.	Vlaardingen	>50	71461	5,9
28.	Dantumadeel	<25	19461	5,9
29.	Meerlo-Wansum	<25	7660	5,9
30.	Hoorn	>50	68174	5,9
31.	Zwolle	>50	114635	5,9
32.	Nunspeet	25-50	26583	5,9
33.	Veere	<25	21947	5,9
34.	Dongen	>25	25586	5,9
35.	Alkmaar	>50	94158	5,9
36.	Nunspeet	25-50	26629	5,9
37.	Bellingwedde	<25	9514	5,8
38.	Borne	<25	20521	5,8
39.	Sevenum	<25	7625	5,8
40.	Woudenberg	<25	11403	5,8
41.	Landerd	<25	14741	5,8
42.	Wormerland	<25	15856	5,8
43.	Borger-Odoorn	25-50	26297	5,8
44.	Alkmaar	>50	94174	5,8
45.	Borne	<25	20531	5,8
46.	Schinnen	<25	13437	5,8

47.	Heerenveen	>25	42732	5,8
48.	Aa en Hunze	25-50	25212	5,8
49.	Someren	<25	18134	5,7
50.	Duiven	25-50	25605	5,7
51.	Teylingen	25-50	35008	5,7
52.	Harenkarspel	<25	15910	5,7
53.	Oosterhout	>50	53295	5,7
54.	Groesbeek	<25	18826	5,7
55.	Heumen	<25	16827	5,7
56.	Hardenberg	>50	57909	5,7
57.	Hengelo (O)	>50	81299	5,7
58.	Aalburg	<25	12221	5,7
59.	Wijchen	>25	39495	5,7
60.	Onderbanken	<25	8391	5,6
61.	Leudal	25-50	36818	5,6
62.	Hoogeveen	>50	54383	5,6
63.	Lelystad	>50	72252	5,6
64.	Westland	>50	98869	5,6
65.	Goedereede	<25	11631	5,6
66.	Haren (Gron)	<25	19107	5,6
67.	Oud-Beijerland	<25	23899	5,6
68.	Ouder-Amstel	<25	13142	5,6
69.	Waalre	<25	16487	5,6
70.	Bernheze	25-50	29544	5,6
71.	Heerhugowaard	25-50	49444	5,6
72.	Rijswijk	25-50	47152	5,6
73.	Teylingen	25-50	34683	5,6
74.	Wassenaar	25-50	25622	5,6
75.	Giessenlanden	<25	14456	5,5
76.	Kessel (L)	<25	4267	5,5
77.	Sint Anthonis	<25	11766	5,5
78.	Kampen	25-50	49359	5,5
79.	Oude IJsselstreek	25-50	40068	5,5
80.	Doetinchem	>50	56238	5,5
81.	Druten	<25	18114	5,5
82.	Graft-De Rijk	<25	6591	5,5
83.	Schijndel	<25	22912	5,5
84.	Boxtel	25-50	30001	5,5
85.	Castricum	25-50	34863	5,5
86.	Dirksland	<25	8316	5,5
87.	Onderbanken	<25	8366	5,5
88.	Zwartewaterland	<25	21872	5,5
89.	Achtkarspelen	25-50	28223	5,5
90.	Bronckhorst	25-50	37700	5,5
91.	Geldrop-Mierlo	25-50	37738	5,5
92.	Midden-Drenthe	25-50	33282	5,5
93.	Veendam	25-50	28185	5,5
94.	Lelystad	>50	71447	5,5
95.	Bolsward	<25	9515	5,5
96.	Nederweert	<25	16441	5,5
97.	Rhenen	<25	18354	5,5
98.	Oude IJsselstreek	>25	40240	5,5
99.	Zederik	<25	13553	5,5
100.	Sliedrecht	<25	23877	5,5
101.	Wijchen	25-50	39885	5,5
102.	Harlingen	<25	15465	5,4

103. Lopik	<25	13999	5,4
104. Wassenaar	25-50	25608	5,4
105. Littenseradiel	<25	10850	5,4
106. Oirschot	<25	17852	5,4
107. Coevorden	25-50	36040	5,4
108. Medemblik	25-50	26569	5,4
109. Oldebroek	<25	22880	5,4
110. Putten	<25	23173	5,4
111. Twenterand	25-50	33438	5,4
112. Hilvarenbeek	<25	15115	5,4
113. Oosterhout	>25	52853	5,4
114. Graafstroom	<25	9685	5,4
115. Bennebroek	<25	5156	5,3
116. Bergen (L)	<25	13561	5,3
117. Lemsterland	<25	13434	5,3
118. Menaldumadeel	<25	13809	5,3
119. Wijk bij Duurstede	<25	23366	5,3
120. Brielle	<25	15990	5,3
121. Grave	<25	12868	5,3
122. Loon op Zand	<25	22807	5,3
123. Ubbergen	<25	9401	5,3
124. Roosendaal	>50	77703	5,3
125. Noordwijkerhout	<25	15170	5,3
126. Waddinxveen	>25	26042	5,3
127. Weststellingwerf	>25	25651	5,3
128. Zwolle	>50	111806	5,3
129. Dronten	25-50	38182	5,2
130. Ridderkerk	25-50	44679	5,2
131. Gouda	>50	70953	5,2
132. Rijnwaarden	<25	11066	5,2
133. De Ronde Venen	25-50	34353	5,2
134. Ridderkerk	25-50	44757	5,2
135. Emmen	>50	108589	5,2
136. Echt-Susteren	>25	32325	5,2
137. Neder-Betuwe	<25	22309	5,1
138. Noordwijk	<25	24700	5,1
139. Langedijk	25-50	25931	5,1
140. Oegstgeest	<25	21887	5,1
141. Sluis	<25	24326	5,1
142. Oisterwijk	>25	25905	5,1
143. Renkum	>25	31640	5,1
144. Giessenlanden	<25	14436	5,1
145. Liesveld	<25	9718	5,1
146. Oosterhout	>50	52870	5,1
147. Elburg	<25	22216	5,0
148. Maasdriel	<25	23514	5,0
149. Noordwijk	<25	24673	5,0
150. Woudrichem	<25	14390	5,0
151. Waalwijk	25-50	45762	5,0
152. Gorinchem	25-50	34352	5,0
153. Zandvoort	<25	16599	4,9
154. Waddinxveen	25-50	25896	4,9
155. Schouwen-Duiveland	25-50	34435	4,9
156. Oostflakkee	<25	10242	4,9
157. Ooststellingwerf	>25	26150	4,9
158. Boxtel	25-50	29488	4,9

159. Vught	25-50	25239	4,8
160. Barneveld	>50	50448	4,8
161. Alphen-Chaam	<25	9497	4,8
162. Goirle	<25	22129	4,8
163. Boxtel	>25	29650	4,8
164. Brunssum	>25	29641	4,8
165. Nieuw-Lekkerland	<25	9577	4,7
166. Vlissingen	25-50	45073	4,6
167. Bergen op Zoom	>50	66325	4,6
168. Reusel-De Mierden	<25	12288	4,6
169. Zundert	<25	20797	4,5
170. Zaltbommel	25-50	26191	4,5
171. Steenbergen	<25	23301	4,4
172. Schiedam	>50	75120	4,2
173. Den Helder	>50	58957	3,9
174. Hardinxveld-Gie	<25	17799	3,5
Gemiddeld score			5,5

Over de auteurs

Prof.dr. A. (Arno) F.A.Korsten is sinds september 1986 hoogleraar Bedrijfs- en bestuurswetenschappen, in het bijzonder bestuurskunde aan de Open Universiteit (faculteit managementwetenschappen) en sinds 1991 hoogleraar Bestuurskunde van de lagere overheden aan de Universiteit Maastricht (faculteit rechtsgeleerdheid). Vanaf eind 2010 is hij emeritus aan de OU en honorair hoogleraar aan de Universiteit Maastricht. Tevens free lance consultant. Is Lid van Verdienste van de Vereniging voor Bestuurskunde en werd in 2005 Officier in de Orde van Oranje-Nassau vanwege bijdragen aan de bestuurskunde en het openbaar bestuur. Zie www.arnokorsten.nl.

Dr. K. (Klaas) Abma studeerde aan de Open Universiteit. Promoveerde in 2012 aan de Open Universiteit op het proefschrift *'Beoordeling van gemeenten'* (Wolf Legal Publ.). Hij was eerder adjunct-gemeentesecretaris in de gemeente Littenseradiel. Is thans werkzaam bij de gemeente Súdwest-Fryslân. Publiceerde eerder met Korsten o.a. *'Gemeenten in rapportcijfers'* (Eburon, 2009).

Drs. A.D. (Anne-Douwe) van der Meer studeerde aan de Open Universiteit. Was eerder onder meer werkzaam in een gemeente. Was als controller werkzaam bij het Bureau Jeugdzorg van de Agglomeratie Amsterdam. Thans werkzaam bij de Belastingdienst.