

Peter Noordhoek

Benchmarking en publieke verantwoording

Drs. P. Noordhoek is directeur van Northedge BV (www.northedge.nl) en oud-voorzitter van de zogenaamde rijkstrede benchmark.

In deze bijdrage wordt ingegaan op recente ontwikkelingen ten aanzien van het gebruik van het instrument benchmarking. Wat zijn mogelijkheden en valkuilen en hoe werkt het gebruik van dit soort instrumenten in de praktijk? Zijn ze alleen of voornamelijk relevant voor de eigen bedrijfsvoering of ook in het kader van publieke verantwoording en toezicht? De auteur gaat in op recente ervaringen met een grootschalige benchmark van uitvoeringsorganisaties.

Van Hoofdlijnenakkoord tot Andere Overheid, van Burgerzaken tot Belastingdienst; overal komt de term benchmarking terug. De term heeft een logica gekregen die onweerlegbaar lijkt. Wie kan er tegen zijn? Waarom is het er eigenlijk nog zo weinig? Toch is dat vreemd, want benchmarking is van alle kwaliteitsbenaderingen de meest schaarse en meest moeizame, zeker in overheidsland. Terwijl de hype rond kwaliteitsbenaderingen als ISO en INK ruim over haar hoogtepunt heen is, wordt wel gewerkt aan wat waarschijnlijk het moeilijkste arsenaal in het kwaliteitsinstrumentarium is: het *leren van* het systematisch vergelijken van prestaties, benchmarking. Daarom is het de moeite waard om eerst met verwondering te schrijven over de opkomst van het fenomeen benchmarking, voordat op de inhoud wordt ingegaan. Bij benchmarking gaat het om twee elementen: 1) het vergelijken van prestaties, en 2) het leren van die vergelijking. De praktijk laat zien dat het vergelijken van prestaties, het eerste element van benchmarken, vooral gebeurt als er

sprake is van goed vergelijkbare producten en diensten, bij de beschikbaarheid van grote aantallen cijfers met een hoog routinematig karakter en bij aanwezigheid van een aantal randvoorwaarden. Bij het laatste gaat het om vooral de beschikbaarheid van goed gedefinieerde en gevalideerde definities van indicatoren over een langere periode. In dat laatste geval is er sprake van een stabiele basis voor vergelijking.

Het leren van elkaar, nodig voor de volledige definitie van benchmarking, kent andere randvoorwaarden, zoals vertrouwen en openheid. Deze elementen zijn niet bij voorbaat aanwezig. In de praktijk is dat een proces van elkaar leren kennen en vervolgens ervaren dat van de situatie geen misbruik wordt gemaakt. In het bedrijfsleven blijkt dit lastig, omdat het al snel om het vergelijken van prestaties tussen potentiële concurrenten gaat. In de overheid spelen factoren een rol als de gevoeligheid voor politieke incidenten.

Samengevat zijn de mogelijkheden voor 'echte' benchmarking afhankelijk van twee factoren. Een kwantitatieve,

bepaald door de beschikbaarheid van betrouwbare en gevalideerde gegevens, en een kwalitatieve, bepaald door de mate waarin de betrokken partijen een stabiele basis voor vergelijking weten te bereiken. Dit laatste wordt eerder door vertrouwen bepaald dan door formele aspecten. Alleen in de combinatie van het beschikbaar zijn van betrouwbare (gevalideerde) gegevens in een omgeving die een stabiele basis biedt voor vergelijking, is er een kans op benchmarking in de klassieke zin: het leren van het vergelijken van prestaties. In alle andere gevallen moet er eerst een proces op gang komen.

Soms is dat wel degelijk het geval. In een aantal omgevingen is benchmarken een koud kunstje. Als het doel is om bijvoorbeeld te weten hoe het eigen verzekeringsbedrijf het doet ten opzichte van andere, dan is er veel en goed gevalideerd materiaal beschikbaar en is er ook een houding die vergelijken logisch maakt. Binnen de overheid zijn daar ook wel voorbeelden van te vinden. Zo zijn meerdere uitvoeringsorganisaties bezig hun informatiseringsprocessen te vergelijken op basis van de aanpak van Gartner (www.gartner.com). Semi-overheidsorganisaties als het ABP en PGGM doen volop mee in de benchmarks van de pensioen- en verzekeringswereld.

Op gemeentelijk terrein is het opvallend hoe goed de benchmarks zijn op gebieden als riolering, maar ook op onderdelen van het werk van sociale diensten (Benchmark Divosa, www.divosa.nl). Daar zijn goede cijfers beschikbaar, is het product bekend en bestaan er ook een wens en noodzaak om meer samen op te trekken.

Het voorbeeld dat centraal staat in dit artikel is de Rijksbrede Benchmark (RBB). De RBB omvat een groot aantal uitvoeringsorganisaties, werkzaam in zeer uiteenlopende sectoren en werk-

zaam in omgevingen (politie, sociale zekerheid) die allerm minst stabiel zijn. De meeste ervan halen met enige regelmaat de krantenkolommen. Met andere woorden: de RBB is een schoolvoorbeeld van een benchmark waarin het meer gaat om het proces van het vergelijken dan om het vergelijken zelf. Kan zo'n benchmark dan toch een rol spelen in publieke verantwoording?

Rijksbrede Benchmark

De RBB is om een aantal redenen opmerkelijk. Allereerst vanwege de omvang. De elf organisaties verrichten activiteiten (kasstromen plus apparaatkosten) die samen circa 31 procent van de rijksuitgaven omvatten (meer dan 51 miljard euro in 2001, exclusief ABP) en waarbinnen meer dan 23.000 mensen werkzaam zijn (cijfers op basis van de samenvattende rapportages van 2002 en 2003, zie voor 2002 onder andere www.rdw.nl).

Tot de deelnemers uit de eerste ronde (2002-2003) behoorden onder andere het Algemeen Burgerlijk Pensioenfonds (ABP), het Centrum Opvang Asielzoekers (COA), de Informatie Beheer Groep (IB-Groep), het agentschap Centrale Financiën Instelling (Cfl), de Rijksdienst voor het Wegverkeer (RDW), de Sociale Verzekeringsbank (SVB) en de regiopolitie Amstelland. In een tweede ronde (2003-2004) zijn daar organisaties als de Dienst Justitiële Inrichtingen (DJI), het Uitvoeringsinstituut Werknemersverzekeringen (UWV) en de Belastingdienst aan toegevoegd. Daarmee is deze benchmark veruit de grootste binnen de publieke sector. Inmiddels zijn er voldoende nieuwe organisaties om in 2005 een derde ronde van start te laten gaan. Daarnaast zullen sommige organisaties uit de eerste ronde opnieuw de vergelijking ingaan.

De aanleiding voor het ontstaan van de RBB ligt volledig in de sfeer van publie-

ke verantwoording. Enkele jaren geleden worstelde zowel de IB-Groep als het ministerie van OCenW met een grote overschrijding in de kosten van het zogenaamd 'herontwerp' van de bedrijfsvoering. Het ging om forse bedragen, maar hoe fors is fors? Het ministerie en de uitvoeringsorganisatie vonden elkaar in de behoefte aan een vergelijking met andere organisaties. Het was bekend dat ook veel andere organisaties, inclusief veel bedrijven, te maken hadden met tegenvallers bij automatiseringsprojecten. Het ministerie was geïnteresseerd in de vraag hoe de tegenvaller zich verhiel tot andere. De IB-Groep was dat ook, maar had daarnaast nog een andere frustratie. Daar vond men dat de tegenvaller bij het herontwerp te veel het zicht ontnam op de prestaties die er ondertussen wel degelijk door de IB-Groep werden geleverd. Via een benchmark zou men richting de Tweede Kamer en anderen kunnen laten zien wat er vergelijkenderwijs aan de hand is. Uiteindelijk werden er vijf organisaties bereid gevonden om aan de benchmark deel te nemen. Op basis van een methode geïnspireerd op het INK-model werd een eerste voorzichtige vergelijking gemaakt. Met het nodige voorbehoud resulteerde dit in een eerste benchmark. Het rapport hierover werd aan de Tweede Kamer gezonden, met de cijfers van de IB-Groep openbaar en die van de andere organisaties anoniem. Daar werd het rapport positief, maar verder geruisloos ontvangen.

De ervaring smaakte naar meer. Het verantwoordingsaspect bleek niet zo spannend als in eerste instantie gedacht, terwijl de wisselwerking met andere organisaties inspirerend bleek. Toch was het niet eenvoudig om de deelnemende organisaties te binden aan een vervolg. Door een golf van reorganisaties bleken organisaties niet meer te bestaan of

werd de vergelijkingsbasis onbetrouwbaar. Toen is vanuit de zijde van het departement het initiatief genomen om de benchmark rijksbreed te trekken. De vraag is aan het SG-beraad voorgelegd om vanuit elk departement te participeren en daar is een goede respons op gekomen.

De eerste echte ronde kon van start gaan, onder leiding van een onafhankelijke voorzitter en een secretariaat waarin de ministeries van OCenW, BZK en Financiën participeerden. De feitelijke werkzaamheden werden voor een belangrijk deel door KPMG gedaan. De resultaten van de benchmark werden op de eerste 'Dag van de uitvoering' gepresenteerd, opnieuw zonder veel aandacht in de buitenwereld. Goed nieuws is geen nieuws.

In de tweede ronde is de werkwijze ruwweg hetzelfde gebleven, maar de benadering van de benchmark is anders geworden en heeft een nieuwe lading gekregen. Veel meer dan in de eerste ronde zijn het nu de deelnemende organisaties zelf die de gang van zaken bepalen. De database met informatie is bij hen in beheer, samen bepalen ze de koers van de groep. In goed overleg is de rol van de externen terugdrongen en is het accent verschoven naar het zelf doen. Daarmee is het accent ook aan het verschuiven van het verzamelen en interpreteren van de informatie naar het beoordelen en leren. In de tweede ronde wordt de departementale bemoeienis minder, want de partners in de benchmark hebben de ondersteuning nu zelf ter hand genomen. Departementale interesse blijft er wel: bijvoorbeeld het ministerie van Sociale Zaken stimuleert alle aan haar gelieerde organisaties om deel te nemen, reden waarom organisaties als UWV, CWI, SVB (opnieuw) en de sociale diensten van enkele grote gemeenten participeren.

'Brenge

De benchmark is opmerkelijk vanwege de werkwijze die is gevolgd. Deze is nadrukkelijk gebaseerd op het INK-model. Dit model kent een onderscheid tussen vijf aandachtsgebieden die de inzet bepalen van een organisatie en vier resultaatgebieden die de resultaten aangeven van die organisatie in de ogen van belanghebbenden. Inzet en resultaat worden met elkaar verbonden op basis van een regelkring, de zogenaamde Deming-cirkel. De mate waarin men in staat is die regelkring te laten werken - te leren - bepaalt de mate van ontwikkeling van de organisatie (zie onder andere de site www.ink.nl en voor artikelen met commentaar op het model www.northedge.nl).

De werkwijze is uitgemond in een rapportage voor de betrokken organisatie zelf en een samenvattende rapportage. In het eigen rapport kan de organisatie de eigen prestaties vergelijken met de beste, gemiddelde en slechtste van de scores van de andere organisaties. Gericht wordt aangegeven op welke punten de organisatie kan 'halen' en waar zij kan 'brengen' naar andere partners in de benchmark. Dan blijkt dat er meer dan voldoende aspecten zijn waarop de deelnemers van elkaar kunnen leren.

In de tweede ronde van de benchmark (2003-2004) valt te zien dat er minder tijd nodig is voor allerlei vragen rondom definities en methoden. Er gaat meer tijd zitten in de eigenlijke prestatievergelijking en de cijfers worden betrouwbaarder - al moet er nog altijd veel rekening worden gehouden met het feit dat het om zeer verschillende organisaties gaat, werkzaam in zeer verschillende contexten. Het proces richting een stabiele basis voor de vergelijking gaat ook door. Dat blijkt zowel uit de methodeontwikkeling als uit de wijze van organiseren. Er wordt voor gekozen minder aan het

externe bureau over te laten. De validering van de resultaatgebieden gebeurt nu in de vorm van audits. Audits waarin op basis van een vaste methodiek vertegenwoordigers van de partners in de benchmarkgroep elkaar bezoeken en bevragen. Daarmee wordt ook daadwerkelijk invulling gegeven aan het leerelement van benchmarken. Andere leerelementen, zoals het overnemen van best practices, komt moeizamer van de grond, maar formeel en informeel weet men elkaar steeds beter te vinden. Dat weerspiegelt zich ook in de organisatie. Waren in de eerste ronde de spelers nog voor een belangrijk deel vreemden voor elkaar, nu leert men elkaar echt kennen en waarderen. De departementen en de onafhankelijke voorzitter worden bedankt. Men doet het verder zelf; een bestuur en secretariaat worden ingericht en er komt een database in eigen beheer. Allemaal stappen in de richting van een stabiele basis.

Met die stabiele basis voor ogen worden er nog twee belangrijke stappen gezet. De eerste is het voornemen om de bestuurlijke structuur licht te houden en helemaal toe te wijden aan het proces van benchmarken. Er is dus geen sprake van andere doelstellingen dan benchmarken en in het verlengde daarvan de verbetering van de bedrijfsvoering. De tweede stap bestaat uit strakkere afspraken ten aanzien van de externe publicatie van de rapportages. Een dergelijke externe publicatie is niet de bedoeling. De informatie is er voor de organisaties zelf. De buitenwereld krijgt de beschikking over een samenvattende rapportage, met anoniem gemaakte informatie. Het gevolg van deze stappen is per saldo een versterking van het beeld van de benchmark. Terwijl de leiding van de benchmark de resultaten zelf goed weet te relativieren, ontstaat naar binnen en naar buiten toe het beeld van een goed lopend proces. Dit betaalt zich op ver-

schillende manieren uit. Zo is het geen probleem om deelnemers te vinden voor een derde ronde en worden de onderlinge banden merkbaar hechter. Ook op een andere manier komt er waardering. Vanuit projecten als 'Andere Overheid' is er duidelijk waardering voor deze benchmark. Maar in deze lof zit nu juist een probleem.

Geen 'beauty contest'

De belangstelling voor de benchmark is relatief gering buiten een kleine kring van deskundigen. Goed nieuws is geen nieuws. Tegelijkertijd is er veel kritiek op uitvoeringsorganisaties. Het recente IBO-rapport (IBO, 2004) over verzelfstandiging stelt de status van ZBO's ondubbelzinnig ter discussie. In het verlenge daarvan mogen ook diensten en agentschappen verwachten dat ze nog strakker aan de ministeriële verantwoordelijkheid worden gebonden. En dan is het bijna 'gelukkig' dat de belangstelling voor de benchmark relatief gering is. De rijksbrede benchmark is nog in opbouw. Een paar belangrijke slagen kunnen nog worden gemaakt, zoals de verbreding richting kleinere uitvoeringsorganisaties en het aangaan van de vergelijking met het bedrijfsleven. Bovenal moeten de deelnemers zelf er nog meer uithalen dan nu al het geval is. Met dat in het achterhoofd, is het niet zonder risico's om de benchmark te betrekken in discussies over publieke verantwoording. In de discussie over de relatie tussen benchmark en verantwoording overheersen twee benaderingen: enerzijds zijn er degenen die zeggen dat het niet verstandig is om benchmarking in de sfeer van publieke verantwoording te trekken. Anderzijds zijn er degenen die stellen dat publieke verantwoording juist het doel zou moeten zijn van benchmarking. De vertegenwoordigers achter de eerste benadering scharen zich doorgaans achter het rapport van de

Raad voor het openbaar bestuur uit oktober 2002 (ROB, 2002). De raad wijst nadrukkelijk op de mogelijkheid van onbedoelde en ongewenste ('perverse') effecten wanneer benchmarking als verantwoordingsinstrument wordt ingezet: 'zowel in verticale als in horizontale zin'. Het beoogde leergedrag zal dan onder druk komen of zelfs teniet worden gedaan. Zoals de Raad concludeert: 'Benchmarking heeft een leerfunctie, en niet primair een informatiefunctie, laat staan een toezichtfunctie.' De trekkers van de RBB sluiten zich aan bij dit standpunt. Niet omdat men tegen openheid zou zijn, integendeel, maar vanwege een inschatting van de risico's en een overtuiging dat er op andere manieren al uitgebreid verantwoording wordt afgelegd.

Bij het verkeerd betrekken van de benchmark bij de publieke verantwoording valt onmiskenbaar veel te verliezen. Als de deelnemende partijen worden gedwongen tot een soort 'beauty contest' zullen ze allemaal afhaken. Terecht, niet in het minst omdat door de verschillen in taak en inrichting van de organisaties het nooit een eerlijke vergelijking kan worden. Het competitieelement dat in zo'n vergelijking gaat zitten kan al snel een vertekend zicht op de prestaties opleveren. Bijvoorbeeld: een organisatie die in de sfeer van veiligheid werkzaam is, laat zich wel op aspecten vergelijken met een organisatie die bijvoorbeeld in de sfeer van de sociale zekerheid werkzaam is, maar door de verschillen in context is het noch zinvol noch wenselijk om de ene organisatie tot 'beter' uit te roepen dan de andere.

Het risico van deze benadering is het defensieve karakter ervan. De Raad voor het openbaar bestuur stelt in hetzelfde rapport ook dat er een proactieve houding wordt verwacht van overheidspartijen richting consumenten en dergelij-

ke. Daar is geen sprake van als het proces en de resultaten van de benchmark niet of nauwelijks wordt gedeeld met derden.

De andere benadering, gericht op openbaarheid, wint mede daardoor aan kracht. Twee argumenten worden daarbij regelmatig genoemd. Het eerste argument gaat uit van transparantie als norm voor overheidsorganisaties en betwijfelt of er ooit goede redenen zijn om niet transparant te zijn. Angst is een slechte raadgever, moed wordt beloond. Het tweede argument sluit aan bij de maatschappelijke druk die er is rondom het thema publieke verantwoording. Er is duidelijk behoefte aan een sluitend stelsel van publieke verantwoording. Een verantwoording op prestatieniveau en niet alleen op inspanningsniveau. Ministers en anderen komen niet meer weg met een mooi verhaal. Elke bron van prestatiegegevens wordt onder die omstandigheid aantrekkelijk. Als die bron van prestatiegegevens via benchmarking kan worden aangeboord, dan is dat alleen maar des te beter.

Het argument kent overigens een interessante nuancering. Het inzicht groeit dat publieke verantwoording zich moet ontwikkelen richting 'meervoudige verantwoording' (zie ook de inleiding bij dit nummer). Verantwoording wordt een kwestie van meerdere vormen van verantwoording richting meerdere partijen. De rapportage over de benchmarking is dan een vorm van verantwoording die in ieder geval richting de eigen stakeholders (management, medewerkers, raad van bestuur/toezicht) moet worden gebruikt, maar niet noodzakelijk richting de minister hoeft te gaan. Dan gaat het in feite om een beperkte openbaarmaking. Maar werkt dat in de praktijk? Het risico van de benadering schuilt in de hiërarchie die in de praktijk impliciet en expliciet wordt gehanteerd als het om publieke verantwoording gaat. Ver-

antwoording aan minister en Kamer weegt zwaarder dan welke andere vorm van verantwoording dan ook. Is de rapportage eenmaal openbaar, dan is er in feite geen barrière meer tegen een volledige openbaarmaking en ligt de rapportage vol in de publieke arena. Helaas mag er niet vanuit worden gegaan dat openbaarmaking altijd ten goede komt aan degene die tot openbaarmaking overgaat.

Tot slot

De deelnemende uitvoeringsorganisaties zullen zelf moeten oppassen dat ze niet verkramppt omgaan met de uitkomsten van de benchmark. De benchmark is niet voor niets juist ontstaan uit een behoefte aan publieke verantwoording. Die behoefte is impliciet en expliciet nog steeds zo groot dat daar een vorm voor moet worden gevonden. Zolang de methode integer blijft en niet alle cijfers direct op straat komen te liggen, moet het daarom mogelijk zijn voor individuele organisaties om hun rapport te publiceren. Daarin staan dan alleen de eigen data, afgezet ten opzichte van de anoniem gehouden overige rapportages. De politiek mag haar zegeningen tellen. Bij de start van de benchmark hebben de diverse fracties zich voorbeeldig gedragen. Ze hebben om de benchmark gevraagd, zich laten informeren over de uitkomsten ervan voorzover het om een specifieke organisatie ging en daarna alleen maar gezegd er vooral mee door te gaan. Bij alle vijandbeelden die er de ronde doen tussen politiek en uitvoeringsorganisaties mag dit ook wel eens worden verteld. Laat men vanuit de politiek deze houding vooral volhouden. Meer moet er worden opgepast door instellingen als de Algemene Rekenkamer en degenen die zich beleidsmatig bezighouden met de inrichting van de overheid. Er is bij de uitvoeringsorganisaties een terechte zorg dat de kennis

ontbreekt over de spelregels die gelden bij benaderingen als benchmarking. Het lijkt wel of de ene helft van de ambtenaren het cynisch afwijst als de zoveelste truc om greep te krijgen op iets wat men toch niet kan pakken, en de andere helft het dood wil drukken in het verlangen om er mee te scoren naar de buitenwereld toe. Wellicht moet er een scherper onderscheid worden gemaakt tussen prestatievergelijking en benchmarken. Prestatievergelijking gaat om in beginsel openbaar beschikbare kengedaten die met elkaar in verband worden gebracht. Benchmarken is het mooie experiment dat je vooral niet moet verstoren.

Noot

1. Zie het rapport van de MDW-werkgroep Benchmark gemeentelijke dienstverlening - Lokaal loket, Den Haag, 21 mei 2002. Inmiddels krijgt dit een ruim vervolg, onder andere in het kader van de werkgroep benchmarken openbare sector, waarin Rijk, IPO, VNG en consumentenbond vertegenwoordigd zijn en in het platform kwaliteit van de VNG, zie www.vng.nl. Een Handreiking prestatievergelijking ten behoeve van benchmarks en een Gids prestatievergelijking en Benchmarks binnen de publieke sector zijn binnenkort beschikbaar.

Literatuur

- Raad voor het Openbaar Bestuur, Presteren door leren. Benchmarken in het binnenlands bestuur, Den Haag, 2002.
- Rapport van de werkgroep Verzelfstandigde Organisaties op Rijksniveau, Een herkenbare staat. Investeren in de overheid. Interdepartementaal Beleidsonderzoek 2003-2004, nr. 1, Den Haag, 2004 (IBO, 2004).