

Wat ervan terecht komt

Zicht op beleidsevaluatie

Prof.dr. Arno F.A. Korsten

08 04 2013

VOORWOORD

In de loop van 2012 bereikte mij het verzoek van het provinciebestuur van Noord-Brabant om mee te denken met de verdere ontwikkeling van beleidsevaluatie door en voor de provinciale organisatie en het bestuur. Ik zou in 2013 een inleiding dienen te houden ten behoeve van topambtenaren en een aantal functionele relaties van het provinciebestuur uit het maatschappelijk middenveld. Hierover en over de discussie naar aanleiding daarvan zou dan een brochure verschijnen in het licht van 'leren door evaluaeren'.

Op dat verzoek wou ik graag ingaan omdat ik over het onderwerp beleidsevaluatie en over de benutting van onderzoeksbevindingen in de loop der jaren gepubliceerd heb. Ik heb in de loop van de tijd namelijk zelf als opdrachtgever van evaluatieonderzoek gefungeerd (bij wetsevaluatie), zelf beleidsevaluaties verricht (onder meer op het vlak van innovatiebeleid) en ook als voorzitter of lid van een visitatiecommissie tal van visitaties van organisaties of organisatie-eenheden uitgevoerd in Nederland en België, onder meer op het vlak van bestuurskracht. Dat leverde mij een schat aan inzichten op. Daarom besloot ik niet alleen in te gaan op een verzoek om een inleiding te houden maar ook een achtergronddocument op te stellen. Deze beschouwing is te beschouwen als dat achtergronddocument.

Dit document geeft een overzicht van wat ik beschouw als de stand van zaken ten aanzien van beleidsevaluatie bij overheden.

De verhandeling laat zich zelfstandig lezen. Ze is mogelijk ook van nut voor vele andere lezers, waaronder studenten.

Arno Korsten, 7 april 2013

Inhoudsopgave

1 Inleiding

- Veel beleid, toenemend aantal beleidsevaluaties
- Kanttekening
- Aantal evaluaties per jaar
- Een specialisme ontstaat
- Fase in de beleidscyclus
- Casus 1: Evaluatie van trage standpuntbepaling – een tragedie
- Casus 2: Evaluatie als schoenlepel - de Floriadesof
- Drie lessen

Deel I: Algemeen

2 Evalueren: een introductie

- Wat evalueren is
- Wat beleidsevaluatie is
- Wat een gebruiksgerichte evaluatie is
- Terugblik op de casus
- Wat belicht wordt
- Klassieke beleidsevaluatie
- Voorwaarden voor effectiviteitsevaluatie
- Tot welke bevindingen leidt de effectiviteitsevaluatie?
- Dilemma: achteraf discussie vanuit andere criteria

3 Waarom evaluaties starten: de argumenten pro

- Argumenten
- De aanleiding voor evaluatie: over sunsets en meer

4 Functies en nevenfuncties van beleidsevaluatie

- Functies van beleidsevaluatie
- Nevenfuncties van evaluatie

5 Een stappenplan voor een klassieke beleidsevaluatie

- Stappenplan voor evaluatie

Deel II: Typen beleidsevaluatie

6 De klassieke beleidsevaluatie: effectiviteitsanalyse

- Vertrek vanuit een beleidstheorie
- Formele evaluatie

7 De opzet van een klassieke beleidsevaluatie: enkele dilemma's

- Kan het zoals het moet of moet het maar zoals het kan?
- Wet van de evaluatieve gebreken

8 Het te vroege evaluatiemoment: early warning evaluation als oplossing

- Wat een early warning evaluation is
- Het waarom van early warning evaluation
- Voorwaarden
- Werkzaamheden
- Conclusie

9 De veelheid aan maatstaven: de normenbank en selectiviteit als oplossing

- Maatstavenvloed

Twee gevolgen: normenbank en code
Maatstaven in een netwerkcontext

Deel III: De keuze van een evaluatie-opzet en de evaluatieladder

10 Er is meer dan de klassieke evaluatie: typen beleidsevaluatie op een rij

Doelstellingsvrije beleidsevaluatie ex post
Indelingen
Toelichting: ex ante en ex post evaluaties
Inputevaluatie als vorm van ex ante evaluatie
Efficiency-evaluatie als ex post of ex ante evaluatie
Responsieve evaluatie
Retrospectieve en prospectieve evaluatie ex post
Van welke aard is het meeste beleidsevaluatieonderzoek?

11 De evaluatieladder: kiezen voor een lichte of zware evaluatie?

Geen evaluatie nodig of slechts een sobere
Zware evaluatie
Pseudo-evaluatie

Deel IV: Ex post evaluatie nader bezien

12 Monitoring: laag op de evaluatieladder

Wat het is
Conclusies

13 Benchmarking: hoger op de evaluatieladder

Benchmarking, wat het is
Plaats op de evaluatieladder
Beoordelingskader
Functies van benchmarking
Voorwaarden
Beperkingen en grenzen van vergelijking
Waardering van benchmarking

14 Zelfevaluatie en visitatiecommissies: hoog op de evaluatieladder

Plaats op de evaluatieladder
Populariteit
Conditie en inrichtingsbeslissingen
Beginselen van behoorlijk visiteren
Geijkte praktijk
Retrospectief en ook prospectief?
Conclusie

Deel V: Ex ante evaluatie nader bezien

15 Ex ante evaluatie

16 Een subtype van ex ante evaluatie: effectrapportages

Deel VI: Leren van evalueren

17 De evaluatoren: wat ze kunnen leren van kritiek

Te weinig bezinning op de opdracht
Begrip van doelbereiking schiet tekort

De keuze van de centrale maatstaf
De reikwijdte van het onderzoek
Veldkennis te gering?
Te weinig nieuws?
Te weinig prospectief
Wat is daar nog als reactie op mogelijk?
Alternatief 1 - Responsieve evaluatie: rekening houden met belanghebbenden
Alternatief 2 - Evaluatie als argumentatie
Alternatief 3 - Netwerkgericht evalueren
Alternatief 4 - Lerend evalueren: leren in een beleidsnetwerk
Conclusie

18 De ambtenaren en bestuurders: wat ze doen met evaluaties

Drie manieren om gebruik te analyseren
Gebruiksfuncties van beleidsevaluaties
Aanzienlijk gebruik
Welke kennis meest gebruikt: doelbereikingsanalyses of procesevaluatie?
Wat werkt in op gebruik of benutting?
Naar een lerende cultuur

19 De politici: wat ze doen met evaluaties

Stelling
Bij welke type volksvertegenwoordigers evaluatie past
Omgaan met een opdracht
De beleidsevaluatie ligt er. En dan?
De aard van de kwestie doet ertoe
Adviezen
Conclusies

20 Leren door evalueren en wat leren belemmert

Omschrijving van beleidsleren
Aspecten van leren
Leergradaties
Hoe het leren plaatsvindt
Belemmeringen voor leren
Conclusie

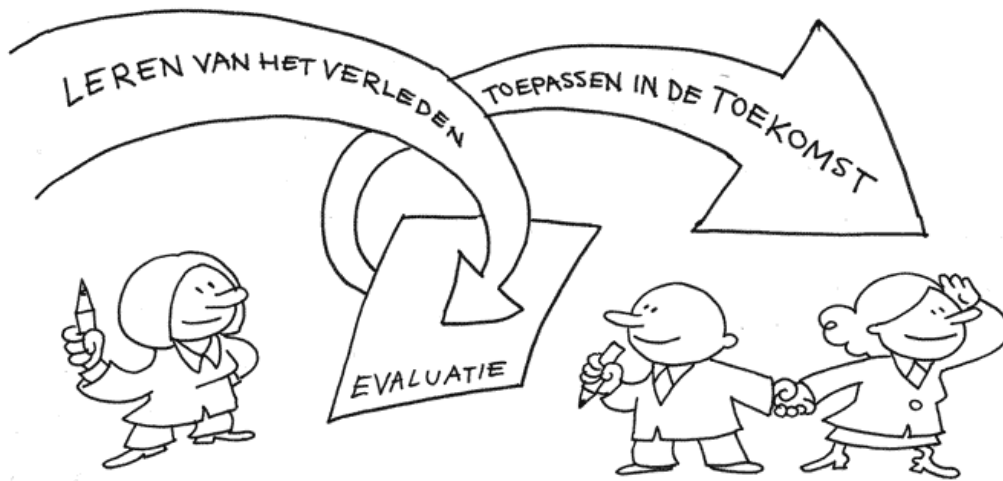
Deel VII: Capita selecta en samenvatting

21 Denkpistes: beleidsevaluatie in relatie tot benaderingen van beleid

22 Samenvatting

Literatuur

WAARDE TOEVOEGEN



1 INLEIDING

Om maar direct met een stelling te beginnen: *waar veel beleid ontstaat en vastgesteld wordt, is er behoefte aan beleidsevaluatie*. Hiervoor bestaan diverse argumenten.

Het argument van onzekerheid. Veel beleid zal niet het doel bereiken omdat het een sprong in het duister is (Bovens en 't Hart, 1996; 2001). De vraag wordt dan waarom het lukt om succes te hebben met beleid of niet lukt? En politiek wordt een zaak van repareren, zoals Hans Wansink (20 juli 2002) terecht stelt. We kunnen wel een voedsel- en warenautoriteit oprichten maar toch bleek sprake van gekke koeienziekte (BSE), van verpakte producten waarin paardenvlees zat dat niet op het etiket voorkwam, van verkoop van waterige tomaten en plofkippen, van producten met het ehecategorie-gevaar, etc. Evaluatie en maatregelen gevraagd.

Argument: ontrafel het fiasco. We kennen natuurlijk bij zo veel beleid ook beleidsmiskleunen. Er kunnen schandalen volgen en crises, beleidsaffaires en beleidsfiasco's. Het aantal daarvan neemt historisch gezien toe.

Enkele afkortingen en namen laten dit indicatief zien: RSV-affaire (scheepvaart), paspoort-affaire, CTSV-affaire, Kalkar, GBA (bevolkingsadministratie niet op orde), WBL (woningcorporatie), Ceteco (ambtenaar maakt geld van geld), klimaatverandering en broeikaseffect (cie-Van Middelkoop rapporteerde er over), integratiebeleid (cie-Blok maakte een analyse), tbs-stelsel, grote infrastructuurprojecten (cie-Duivesteijn ging in op de overschrijdingen in tijd en budget en op andere aspecten), onderwijsvernieuwingen (cie-Dijsselbloem over succes en falen), zorguitgaven (cie-Mosterd), privatisering en verzelfstandiging (cie-Kuipers over succes en falen), Icesave, onderzoek naar het financieel stelsel (cie-De Wit over bankencrisis), onderhoud en innovatie op het spoor (waarom zoveel vertragingen?), de Fyra-kwestie (ondeugdelijke trein?).

Argument van beleidsverstremgeling. Beleid roept als vanzelf de noodzaak van evaluatie op. Daarmee komt een volgend argument pro evaluatie naar voren. Waar veel beleid is, gaat dit verstremgeld raken en stapelen en als een beleidsvoerder dan wil ingrijpen, wil hij of zij een 'cognitieve basis' daarvoor hebben. Die behoefte kan onder meer vervuld worden door een beoordeling van beleidsprocessen op effecten. Ik licht dit toe.

Veel beleid, toenemend aantal beleidsevaluaties

De overheden zijn in de tweede helft van de vorige eeuw beleid op tal van fronten gaan maken in het kader van de ontwikkeling van de welvaartsstaat en de doorontwikkeling naar een verzorgingsstaat. Er kwam verdere onderwijs- en arbeidsmarktpolitiek van de grond, en er kwamen tal van wetten op het gebied van sociale zekerheid; om een vangnet te scheppen voor werklozen en voor gehandicapten, etc. . De beleidsproductie is onder invloed van andere omstandigheden, zoals economische crises en andere crises, doorgegaan. Er trad ook volop beleidsverandering op. Veel beleid is tegenwoordig zelfs een kwestie van '*policy change*'.

Hoe de beleidsverandering te typeren?

1. De eerste vorm heet *beleidsinnovatie*. Een beleidsprogramma is nieuw en raakt niet aan bestaand beleid. Er wordt een nieuw gebied betreden. Humaan omgaan met op het strand aangespoelde potvissen? Een protocol voor 'hoe hiermee om te gaan?' Het onderwerp kwam in het parlementair jaar 2012-2013 in de Tweede Kamer aan de orde.
2. De tweede vorm van beleidsverandering is de *integrale of partiële beleidsbeëindiging*: beleid stoppen of budgettair inkrimpen of qua doelgroep beperken.

3. En de derde vorm noemen we *beleidsopvolging*. Het meeste nieuwe beleid is beleidsopvolging. Het ene beleid vloeit als het ware over in een nieuw beleid.

Deze *'policy succession'* kan tal van vormen aannemen.

1. Een vorm van beleidsopvolging is *'replacing'*: het voorbeeld van 'oude schoenen maken plaats voor nieuwe schoenen' is wat bedoeld wordt. Denk aan de Bescherming Bevolking (BB) die in de jaren tachtig van de vorige eeuw is ingeruild voor de Rampenbestrijding en civiele verdediging.
2. Een andere vorm van beleidsopvolging is *'policy split'*, een uitsplitsing van bestaand beleid in nieuw 'afzonderlijk' beleid. Denk bij voorbeeld aan podiumkunstenbeleid dat uiteengaat in beleid voor ballet, beleid voor orkesten, voor toneel, voor cabaret, etc. Denk aan de ontwikkeling van milieubeleid tot specifiek beleid voor de bodem, de lucht, het oppervlaktewater, de zee, etc.
3. Een derde vorm van beleidsopvolging is *'policy integration'*. Een aantal wetten wordt bijvoorbeeld na enige tijd samengevoegd tot een nieuwe wet, de opvolger. Zo zijn milieuwetten afgeschaft en is het onderwerp van die vorige wetten opgegaan in een - integrerende - Wet milieubeheer.

Soms komen combinaties van gedeeltelijke beëindiging en opvolging voor. Beleid wordt bij voorbeeld gedecentraliseerd (heroriëntatie) onder gelijktijdig inkrimping van de middelen (krimp). Zoals in 2012/2013 is voorgesteld door minister Plasterk in samenspraak met vakministers ten aanzien van de decentralisaties op het vlak van jeugd, zorg (AWBZ) en werk (Participatiewet). Dan verandert de institutionele setting en is sprake van beleidsopvolging in combinatie met partiële beëindiging.

Kanttekening

Een bijzonder fenomeen is het volgende. Het aantal verwijzingen naar wetten neemt sinds pakweg 1980 toe. Dat heet *beleidsverstrengeling* of 'Vernetzung'. Speelt beleidsevaluatie hierin nog een rol? Aan een gerichte beleidsverandering ligt vaak een beleidsevaluatie ten grondslag en niet zelden ook aan de uitvoering van beleid.

Beleidsvaluatie speelt zeker een rol bij het stoppen van beleid. Gebleken is dat wie beleid wil stoppen aan drie voorwaarden moet voldoen om betrekkelijk snel succes te hebben:

1. het ideologisch tij moet gunstig gezind zijn,
2. er moet een ombuigingsnoodzaak bestaan en
3. er moet een beleidsevaluatie ten grondslag liggen aan een nieuw beleidsvoorstel.

Drie voorwaarden dus, die bij voorkeur gelijktijdig moeten voorkomen.

Beleidsvaluatie is ook wenselijk om twee bijzondere redenen. Goed beleid vraagt om meer beleid en slecht beleid vraagt om een correctie. De rechtvaardigingsgrond voor dergelijke aanpassing zal vaak in een beleidsevaluatie te vinden zijn.

Aantal evaluaties per jaar

In welke orde van grootte moeten we ons die beleidsevaluatie op rijksniveau voorstellen? Te denken is aan een pakweg *300 evaluaties per jaar* op rijksniveau (Bongers en Den Hertog, 2007: 30).

Met de toename van beleid en de toenemende verstrengeling van beleid en verandering in de aard van het beleid (meer beleidsopvolging) is de beleidsevaluatie toegenomen. Logisch, want

uit veel evaluaties zal komen dat beleid maar gedeeltelijk een succes is en zich leent voor ombuigingen of heroverwegingen. Ook de omvang van de wetenschappelijke literatuur over evalueren is gegroeid.

Een specialisme ontstaat

Beleidsvaluatie groeide na de jaren zestig in de vorige eeuw uit tot een vakgebied, een specialisme. Drie indicaties:

- Er kwamen nieuwe tijdschriften van de grond met de naam ‘*Evaluation*’.
- Er verschenen internationaal boektitels met koppen als ‘*Public Policy Evaluation*’ en ‘*Evaluierungsforschung*’.
- Het ministerie van Financiën ging via de afdeling Beleidsanalyse beleidsvaluatie pushen.

Het gebied groeide ook onder invloed van ontwikkelingen in overheidsland zelf. Er is al het een en ander over gesteld. Meer beleid, meer verstrengeling van beleid: wordt het beoogde dan bereikt?

Fase in beleidscyclus

Tegenwoordig is beleidsvaluatie een regulier onderdeel geworden van wat heet de beleidscyclus. De beleidscyclus bestaat uit probleemanalyse, beleid ontwerpen, besluitvorming, beleid uitvoeren, beleid beoordelen op effectiviteit, beleid stoppen (als het probleem verdwenen is of opgelost is) of beleid aanpassen. Die cyclus zit weliswaar in de hoofden van veel topambtenaren van de rijksoverheid, provincies, gemeenten, waterschappen maar de praktijk is een andere. Een provinciale topambtenaar zei me in 2013:

‘Toen ik nog op een departement werkte, lieten we wel ex ante evaluaties verrichten maar hier op de provincie is dat duidelijk veel minder gebruikelijk. Ook ex post evaluatie hangt er wat bij. Dat zou niet zo moeten zijn want in de beleidscyclus volgt vanzelfsprekend na enige tijd een bezinning op beleid. Dan is beleid sleets, of niet sleets maar het beleid is uitgewerkt, of er is van alles een beetje sprake: de noodzaak tot ombuiging, en knelpunten in de beleidsuitvoering. En of het beleid nu opnieuw zou worden uitgevonden, als die vraag vandaag gesteld zou worden, is dan misschien ook zeer de vraag’.

Deze ambtenaar zei me ook:

‘Er is eigenlijk meer beleidsvaluatie nodig. We moeten af en toe halt houden. En dan is het handig als er een speciale afdeling is voor beleidsvaluatie. Die afdeling kan af en toe eens langs komen en vragen naar wat echt aan evaluatie onderworpen moet worden en die kan zo een evaluatie ook zelf uitvoeren’.

Ik repliceerde:

‘Ik begrijp dat. Maar dan vinden jullie zo’n afdeling weer lastig. Dan parkeren jullie de beleidsvaluatie daar.’

Zijn reactie:

‘Ja, dat zou kunnen maar dat is natuurlijk niet de bedoeling. Hoe dan ook moeten er meer evaluaties plaatsvinden. En het evalueren is vakwerk. Dat kunnen de meeste provincieambtenaren niet want die hebben uiteenlopende professionele achtergronden. Velen zijn ongeschikt voor dit onderzoekswerk. En een slager moet het eigen vlees niet keuren. Dus een afdeling moet niet zowel het beleid ontwikkelen als

beleiden als ook nog eens evalueren. Dan ontstaat de neiging om beleid goed te beoordelen omdat dit een herbevestiging geeft van het nut van de eigen afdeling. Kijk eens hoe goed wij het doen.'

Ik weer: Er was toch al een beleidsevaluatiecommissie:

'Ja, ik meen van wel maar dat was een commissie van Provinciale Staten. Niet verkeerd. De politiek heeft het primaat. Maar zo een commissie zorgt niet voor een gangbare opname van beleidsevaluatie in de beleidscyclus waar ambtenaren bij betrokken zijn. Evalueren moet gericht zijn op betrokkenheid van alle relevante actoren: ambtenaren, GS en PS en niet te vergeten ook nog van degenen die meewerken aan het beleid, zoals andere overheden.'

De conclusie is duidelijk. Beleidsevaluatie verdient evident meer prioriteit, zoals ook de commissaris des konings van Noord-Brabant, de bestuurskundige Wim van de Donk, tegen me zei.

Kent beleidsevaluatie zelf niet ook een cyclische ontwikkeling?

Inderdaad. Na enige tijd werd de afdeling Beleidsanalyse van Financiën opgeheven. Beleidsevaluatie was gemeengoed geworden en het ministerie van Financiën moest zelf ook bezuinigen. Dus verdween ook het tijdschrift Beleidsanalyse, dat door Financiën werd uitgegeven. Daarentegen ging de Algemene Rekenkamer juist meer beleidsevaluaties verrichten, iets was voor 1970 niet gebeurde. En deze rekenkamer hield stug vol.

Conclusie: Beleidsevaluatie is inmiddels gewoon geworden maar kan wat injectie in provincies en gemeenten wel gebruiken.

Zijn er ook overtuigende argumenten pro evaluatie als je enkele casus beziet? Twee amusante casus worden kort belicht.

Casus 1: Evaluatie van trage standpuntbepaling - een tragedie

In de jaren tachtig kreeg ik – schrijver dezes - de opdracht van enkele departementen om met enkele collega's onderzoek te doen naar het in de steigers gezette nationaal technologisch innovatiebeleid. Er was juist een nationale (rijksbrede) innovatienota verschenen, de eerste. Enkele topambtenaren vonden dat als je een nota opgetuigd had met veel ambities en instrumenten er eigenlijk ook een op evaluatie gericht *onderzoeksprogramma* aan gekoppeld moest worden om te zien of je met de nota op de goede weg was of moest bijstellen. Een van de geprioriteerde projecten betrof het onderwerp '*de opstelling van een onderzoeksprogramma naar biotechnologie*'.

De programmacommissie onder leiding van prof. Schilperoort had snel werk verricht om tot zo'n programma te komen. In een maand of drie, vier lag er een kloek programma dat evenwichtig was en steun bij universitaire vakgroepen die op het gebied van biotechnologie onderzoek deden, genoot. De standpuntbepaling van de regering vergde echter ruim een jaar, vier keer zo lang als het opstellen van het programma zelf. Dat leidde tot verwondering en verbazing, zeker omdat de oorzaak daarvoor niet leek te liggen in de inhoud van het programma. Waarom duurde het proces van departementale standpuntbepaling zo verschrikkelijk lang, ook volgens tal van ambtenaren? Dat moest ik onderzoeken, samen met iemand die thuis was in de biotechnologie. We bekeken het rapport- Schilperoort, analyseerden vergaderstukken van de commissie die het regeringsstandpunt moets voorbereiden en hielden interviews met departementsambtenaren die deelnamen aan de voorbereiding van het

standpunt. Ons rapport verscheen. De begeleidingscommissie toonde zich tevreden met ons rapport.

Een van de vermelde inzichten was dat de interdepartementale samenwerking niet goed was verlopen. Er waren zeker drie departementen betrokken bij de standpuntbepaling. Relevant bleek vooral: de voor het programmadocument eerstverantwoordelijke minister voor Wetenschapsbeleid zou bij de volgende verkiezingen voor de Tweede Kamer vermoedelijk verdwijnen, zo was de inschatting in politiek en ambtelijk Den Haag, en daarmee zou het onder zijn leiding ressorterende programma verschuiven, dat wil zeggen elders 'opgehangen worden'. Dan zou het beleid ten aanzien van biotechnologisch onderzoek verschuiven naar het ministerie van Economische Zaken (EZ) en daarom liepen de betrokken EZ-ambtenaren niet hard om nog voor de Kamer-verkiezingen een standpunt te hebben. Ze traineerden het proces. Het ging hen niet alleen om macht maar ook om een standpunt dat paste in het komende EZ-beleid. Dat was in de concepten nog niet het geval want het ministerie van Landbouw bemoeide zich er ook mee en dat van Onderwijs; die 'hingen tegen'. Een casus vol *bureaupolitieke strijd*, zogezegd. Het was een bijzondere ontdekkingstocht voor mij. Een illustratie van wat in de boeken over bureaucratietheorie te lezen was. Ook een illustratie van een inzicht uit de pastorale psychologie: *'verpleegsters lopen niet hard meer als ze weten dat de patiënt dood gaat'*.

Was dit eigenlijk wel een evaluatieonderzoek? Blijken zal van wel. Het is immers een evaluatie van een beleidsproces, namelijk een *totstandkomingsproces* van beleid. Het antwoord wordt verderop nader toegelicht. In feite was het een bijzondere evaluatie.

Casus 2: Evaluatie als schoenlepel- de Floriadesof

De meest gangbare evaluaties zijn evaluaties van beleid in uitvoering. Dan gaat een evaluator op zoek naar de vraag of prestaties en daarop volgende effecten sporen met de voorgenomen beleidsdoelstellingen en wat een discrepantie verklaart. Daarvan is ook een voorbeeld te geven.

In het begin van deze eeuw viel na een krachtige lobby uit Noord-Limburg het besluit om de Floriade in Noord-Limburg (Venlo) te laten plaatsvinden. De tuinbouwtentoonstelling zou goed passen bij een oriëntatie op nabij gelegen agrobusiness in Greenport, logistieke dienstverlening en een grote tuinbouwveiling en er zelfs een injectie voor zijn. Het gebied rond Venlo, in feite geheel Noord-Limburg, zou nog beter op de kaart komen te staan. Het organiseren van de tentoonstelling zou ook energie vrij maken voor andere initiatieven en de regio zelfs zelfbewuster maken, volgens sommige bestuurders. De voorbereidingen startten en er kwam er een Floriade-bv met een directie, een raad van commissarissen en de gemeenten in de rol van aandeelhouders. In het begin van 2012 liep de Floriade op het oog ook redelijk goed. Tal van internationale hoogwaardigheidsbekleders van overheidszijde, onder andere uit Turkije, vereerden de tentoonstelling met een bezoek. De media berichtten. Er werden veel kaartjes aan bezoekers verkocht. De weersomstandigheden waren redelijk en Noord-Limburg liep er ook voor uit.

Was het genoeg? Werden de gewenste aantallen bezoekers gehaald? Was de grote tuinbouwtentoonstelling ook een financieel succes? Dat was altijd wel de bedoeling geweest. Het provinciebestuur had geholpen met budget dat oorspronkelijk 'een herindelingspremie voor Noord-Limburg was'; het geld kreeg in feite een nieuwe bestemming. Maar dat bedrag was helemaal niet genoeg. Hoewel de eindafrekening - toen dit geschreven werd - nog moest volgen, barstte eind 2012 veel kritiek los van de zijde van gemeenten, die zich garant hadden gesteld om tekorten te dekken. Er waren

wel vrij veel bezoekers geweest maar een aantal bezoekers was abonneementhouder. Sommige bezoekers hadden de Floriade wel een keer of twintig bezocht. Ze werden wel als bezoeker geteld maar betaalden niet per keer. En het was bepaalde weken vrij fris geweest, wat de bezoekersaantallen had getemperd. Algemeen directeur Becks liet positieve geluiden horen ('niet te vroeg jammeren'; 'wacht op de eindafrekening') maar er waren ook tegengeluiden. Hij zag de bui echter hangen, vertrok eerder naar een nieuwe baan in Rusland en zei af te willen zien van een prestatiebonus, als die al zou worden toegekend. Hij had die bonus niet hoeven te weigeren want het zag er toch al naar uit dat hij die nooit gekregen had van de aandeelhouders, gezien de Floriadestorm die opstak. Het evenement had miljoenen aan verliezen geleden. 'Wie betaalt, bepaalt', maar ook het omgekeerde geldt: *'wie bepaalt, krijgt ook de verliezen op zijn bord'*.

Waarom werden de hooggestemde verwachtingen niet waargemaakt? En door wie zijn op welk momenten verwijtbare fouten gemaakt die de financiële tekorten hebben veroorzaakt?

De gemeenten die hebben meebetaald en de tekorten op hun bord krijgen, wilden weten *hoe de zaak uit de hand kon lopen*. Hoewel een onderzoek de tekorten niet wegwerkt, en mosterd na de maaltijd is, moest dat onderzoek er wel komen: een onderzoek naar doelbereiking en waarom de doelstellingen niet bereikt werden? Lag het verlies aan zaken waarop de directie invloed had kunnen uitoefenen (o.a. de opzet van de tentoonstelling zelf, het imago; de marketing in binnen- en buitenland; de aard, omvang en prijsstelling van shuttles en consumpties; de omlijsting met evenementen op het terrein; parkeerfaciliteiten; parkeermogelijkheden) of waarop geen invloed mogelijk was (temperatuur; weersomstandigheden in termen van zonne-uren, regen, wind; etc.). Had de directie voldoende en tijdig de vinger aan de pols gehad en had ze signalen ontvangen om nog tijdig bij te sturen en ook echt bijgestuurd? Had de raad van commissarissen adequaat geopereerd naar de directie?

Drie lessen

Wat leert deze casus? Ik geef drie inzichten.

Het eerste inzicht slaat op *multi-gebruik van de evaluatie in het Floriadenetwerk*. De evaluatie is niet slechts voor één actor bestemd maar voor tal van betrokken partijen in het Floriadenetwerk. Die moeten er lering uit trekken. De evaluatie is bestemd voor de Floriadirectie, de raad van commissarissen en de aandeelhouders. En in feite voor alle betrokken gemeentebesturen want de vertegenwoordigers van de gemeenten in de aandeelhoudersvergadering moeten met hun boodschap ('opdraaien voor verliesaandeel') naar de gemeenteraden. En die raden wens en tekst en uitleg. Die kunnen individuele wethouders en burgemeesters niet geven. Daarom is een onafhankelijk onderzoek nodig. Het onderzoek moet *'de achterlanden in het gareel brengen'*.

Het tweede inzicht slaat op het *schoenlepel-effect*: dat er een rekening in een schoen geschoven moet worden en dat er geen voet te vinden is die de schoen wil aantrekken. De evaluatie moet de schoenlepel zijn die er voor zorgt dat de schoen toch wordt aangetrokken. We noemen het officieel neutraler, namelijk de verantwoordingsfunctie van een evaluatieonderzoek.

Een dwaas onderzoek? Nee, want er was 'een schoenlepel nodig' om de gemeenteraden zo ver te krijgen dat ze de rekening willen betalen in tijden van financiële krapte. Maar rijker zijn in feite die evaluaties die ingebed zijn in de beleidscyclus en waarbij het evaluatieonderzoek nog kan leiden tot aanpassingen of vernieuwing in beleid of desnoods het (gedeeltelijk) stoppen van beleid (dus te komen tot 'policy change').

Het derde inzicht luidt: bestuurders willen leren 'waarom het mis ging'. Nuttig? Ja maar de kans om het een volgende keer beter te doen, komt in dit geval niet. De tuinbouwtentoonstelling was namelijk eenmalig, althans in Noord-Limburg. Het is dus *'leren in het luchtledige'*. Bij de organisatie van een volgende Floriade, als die er komt, zal het dan aantredende organisatienetwerk proberen te leren van de Floriade die in Noord-Limburg georganiseerd was.

DEEL I: ALGEMEEN

NIET ZO:



MAAR ZO:



2 EVALUEREN: INTRODUCTIE

Wat evalueren is

Hoe zouden we 'evalueren' omschrijven?

Evalueren is beoordelen of in andere woorden iets op waarde beoordelen of schatten. Beoordelen doe je van iets dat is waargenomen, dus als er sprake is van een voorstelling van een bepaald verschijnsel. Wie beoordeelt, heeft bovendien maatstaven of criteria voor beoordeling nodig.

Daarmee zijn *drie elementen van alle evaluatie* duidelijk:

- een onderwerp (een waargenomen verschijnsel),
- de waarneming zelf,
- maatstaven of criteria ter beoordeling van een waargenomen verschijnsel (Bressers en Hoogerwerf, 1995: 21).

Evalueren doen mensen op een bijna vanzelfsprekende manier, zonder er altijd bij stil te staan. Wie in een sloot kijkt, neemt wellicht waar dat sprake is van 'bruin uitziend troebel en niet stromend water'. Deze waarnemer hanteert als veronderstelling dat vervuiling meestal leidt tot 'bruin water'. Zijn oordeel zal luiden 'vervuild' en 'ongezond'. Hij zal, als hij een boer met vee is er daarom van afzien om paarden of koeien te laten drinken uit deze sloot.

Ander voorbeeld. Evalueren van Q-koorts, hoe gaat dat? Dan moet er eerst iets zijn waargenomen dat als Q-koorts geldt, vervolgens zijn criteria nodig over of en wanneer de koorts goed of slecht is (lees: een kritische waarde overschrijdt) en wat de gevolgen zijn. Eerst dan valt er te oordelen of bij voorbeeld Q-koorts vanzelf overgaat of actief bestreden moet worden, en of wellicht preventieve maatregelen nodig zijn. De elementen onderwerp, waarneming, criteria voor beoordeling komen ook hier dus weer terug.

Wat beleidsevaluatie is

Het voorgaande ging over evaluatie maar wat is *beleidsevaluatie* bij of door een overheid? Beleidsevaluatie bij een overheid slaat op het op een *systematische manier beoordelen van overheidsbeleid*. Het woord 'systematisch' wordt hier toegevoegd om daarmee een onderscheid aan te brengen met impressionistische of impulsieve evaluatie. Impulsief uitgevoerde beleidsevaluatie of 'zo maar een greep' voeren we daarmee direct af.

De literatuur geeft verder aan:

beleidsevaluatie van overheidsbeleid is *het beoordelen van beleidsprocessen* – zoals bij voorbeeld de probleemanalyse of de totstandkoming van beleid -, de beleidsinhoud, de organisatie van het beleid of effecten aan de hand van bepaalde maatstaven of criteria (cf Bressers en Hoogerwerf, 1995: 21). De lezer ziet hier de onderdelen uit de beleidscyclus terugkeren. Elke fase of onderdeel leent zich voor een beleidsevaluatie.

Tot beleidsprocessen behoren daarmee:

- het proces van beoordeling van de agendering van een beleidsonderwerp,
- de beoordeling van een specifieke probleemanalyse,
- de beoordeling van de totstandkoming van beleid en het ontwerpproces,
- de beoordeling van de besluitvorming,
- de beoordeling van de inhoud en organisatie van de beleidsinvoering en beleidsuitvoering,

- en de beoordeling van de beleidsdynamiek in de vorm van gedeeltelijk of geheel stoppen van beleid of beleidsaanpassing of beleidsinnovatie.

In geval van het beoordelen van overheidsbeleid kan een beoordelaar dus een aantal kanten op. Er zijn keuzen te maken en accenten te zetten. Overheidsbeleid is bij voorbeeld te bezien op de aansluiting op een probleem dat aanleiding was voor de totstandkoming van een beleid: is het beleid een effectief antwoord op het probleem doordat het maatschappelijk probleem (vervuild water?; Q-koorts?) verdwijnt of kleiner wordt?

Het overzien van de evaluatieliteratuur toont dat evaluaties op bijna alle terreinen van overheidsbeleid voorkomen, zoals bij voorbeeld bij begrotingsbeleid, jeugdzorg, detentie en resocialisatie, rechtshandhaving, alcoholcontroles, verkeersveiligheidsbeleid, gezondheidszorg, integratiebeleid, de wet maatschappelijke ontwikkeling, ouderenbeleid, natuurbeleid, megastallenproblematiek, verplaatsing van agrarische bedrijven, mestwetgeving, betaald voetbal. Noem maar op.

Wat een gebruiksgerichte evaluatie is

Beleidsevaluatie kan bog twee kanten op: op een wetenschappelijk kennisdoel gericht zijn of gebruiksgericht zijn. Hier gaat het om *gebruiksgerichte beleidsevaluatie*.

Gebruiksgerichte beleidsevaluaties zijn evaluaties waartoe op de een of andere manier opdracht is verleend of die al bij de start door een overheidsbestuur zijn gelegitimeerd en vaak ook door een overheid gefinancierd worden. Hieronder vallen in principe beleidsevaluaties van de Algemene Rekenkamer en evaluaties van andere rekenkamers of rekenkamercommissies. Patton (1997) spreekt van '*utilisation-focused evaluation*'.

Terugblik op de casus

Hoe nu naar de eerder gegeven casus te kijken? De studie naar de totstandkoming van een ministerieel standpunt over het biotechnologieprogramma, hoewel als *opdrachtresearch* verricht vanuit een universitaire vakgroep bestuurskunde, was een voorbeeld van een gebruiksgerichte studie, die vooral gericht was op verheldering en de boodschap 'zo nooit meer'. Concrete andere gevolgen kon die studie verder niet meer hebben. De nevenfunctie van deze evaluatie was dat de opdrachtgever direct liet zien dat de evaluaties van de Innovatienota gestart waren en dat er durf werd opgebracht om kritisch naar het innovatiebeleid te kijken.

Is in geval van de casus over biotechnologie sprake van beleidsevaluatie? Zeker, het betrof een analyse van een totstandkomingproces van een regeringsstandpunt. Deze totstandkoming werd onder meer bezien op het criterium *voortgang* en interdepartementale coördinatie en coöperativiteit en de verklaring van gebrekkige voortgang.

Deze beschouwing handelt verder over *gebruiksgerichte* beleidsevaluatie, met name richt ik me op de beoordeling van gewenste en ongewenste effecten van beleid, dat in uitvoering is of kan komen. Dat in de wetenschap dat evaluatie betrekking *kan* hebben op andere beleidsprocessen, beleidsinhoud, de beleidsorganisatie en de beleidseffecten.

Wat belicht wordt

Drie aspecten van beleidsevaluatie komen aan bod. Beleidsevaluatie is te zien als *product*, *proces en procedure* (Pröpper, 2001: 146). Ik licht dat toe.

Beleidsevaluatie als *product* verwijst naar een rapportage die het resultaat is van de evaluatie: de beoordeling van beleid, de toelichting en argumentatie.

Beleidsvaluatie als *proces* verwijst naar de handelingen die leidden tot een bepaald resultaat, het product. Hier kan het gaan om de formulering van de evaluatievraag, de keuze van de beoordelingsmaatstaven of evaluatiecriteria, de interactie tussen evaluatoren en opdrachtgever, de relatie tussen evaluatoren en betrokkenen. Hoe uitgebreid is het forum dat betrokken wordt bij de beoordeling.

Beleidsvaluatie als *procedure* verwijst naar de methodische opzet.

Klassieke beleidsvaluatie

De klassieke beleidsvaluatie is een beoordeling achteraf van de resultaten en effecten van een in uitvoering zijnde beleidsprogramma en een antwoord op de vraag of die effecten bereikt worden door het beleidsprogramma zelf of door gebeurtenissen, omstandigheden of oriëntaties daarbuiten. Doorgaans zullen de gewenste effecten onderzocht worden. Dan vormen de *officiële* doelstellingen dus het vertrekpunt, want officiële doelstellingen zijn gewenste of beoogde effecten. Deze doelstellingen zijn al voorafgaand aan de democratische besluitvorming geformuleerd. Ze staan bij voorbeeld in een Memorie van Toelichting op een wet en passeren dus de Tweede en Eerste Kamer.

Deze klassieke beleidsvaluatie noemen we een *effectiviteitsvaluatie*. Ze omvat

- de *doelbereikingsanalyse* (ook wel 'goal attainment analysis' genoemd) plus
- de '*oorzakelijkheidsanalyse*' (ligt het bereiken van doelstellingen aan beleid?).

Voorwaarden voor effectiviteitsvaluatie

Om een dergelijke klassieke beleidsvaluatie te kunnen verrichten zouden idealiter de volgende *voorwaarden* vervuld moeten zijn.

1. Er moet sprake zijn van een *before-aftermeting*, zodat verschillen te zien zijn die door het betreffende beleid veroorzaakt worden.
2. De before-aftermeting wordt verricht bij een *controlegroep* en een of meer *experimentele groepen*. De interventie vindt plaats in de experimentele groep en in de controlegroep is die afwezig. Het beleid geldt dus niet voor alle burgers.
3. Achter het te evalueren beleid gaan daadwerkelijk *doelstellingen* zijn.
4. Deze doelstellingen zijn ook *de echte*. En er komen er gaande de uitvoering geen bij.
5. Deze doelstellingen zijn *stabiel*. Ze verschuiven niet in de tijd. Je kunt je erop verlaten.
6. De doelstellingen zijn '*smart*' geformuleerd: Specifiek, Meetbaar, Aanvaardbaar, Realistisch, Tijdgebonden. Een onderzoeker moet immers bij voorkeur over een operationalisatie van doelstellingen kunnen beschikken.
7. De overheid is een *unicentrische overheid*. De overheid die beleid uitvoert 'zit zelf volledig aan de knoppen'; ze beschikt over alle taken, bevoegdheden en middelen. Van een beleidsnetwerk, waarbij de betreffende overheid niet over alle doorzettignsmacht beschikt, is geen sprake. Er zijn geen afhankelijkheden. Niemand kan vetomacht uitspreken en beleid dwarsbomen of vertragen.
8. Onderzoekers zijn *ideale professionals*: ze opereren onafhankelijk, ze beschikken over zowel veldkennis als kennis van overheidsbestuur en vaardigheid van methodologie en onderzoeksmethoden, ze krijgen de volle medewerking van alle betrokkenen bij de overheid en daarbuiten, ze beschikken over een ruim budget om het onderzoek op te

zetten en uit te voeren en ze formuleren de resultaten, al of niet in een apart leesbare samenvatting en een modnelinge toelichting, op een gebruiksgerichte manier.

9. *Gebruikers* kijken naar de bevindingen van de beleidsevaluatie uit. Ze gaan actief op zoek naar het rapport en bestuderen de resultaten zodra het rapport verschenen is individueel en bespreken het in een relevante groep.

Tot welke bevindingen leidt de effectiviteitsevaluatie?

Ik kom vaak de volgende bevindingen van klassieke beleidsevaluaties of effectiviteitsanalyses tegen: een beleidssucces, een beleidsfalen, een gedeeltelijk succes of falen, het resultaat is 'satisficing' (politici aanvaardden de bevindingen, ze leggen zich erbij neer), of 'we weten het niet precies'.

Bijna nooit is sprake van een evident succes. Vooral het omgaan met '*wicked problems*', dus met *hardnekkige problemen* ('brandende kwesties', 'controverses'), is problematisch. Denk aan de bestrijding van voetbalvandalisme.

De oorzaken voor misleunen liggen volgens de effectiviteitsanalyses *bijna altijd bij de overheid zelf* (Bovens & 't Hart, 1996, 2001). Logisch, want er is vanuit een *unicentrisch* overheidsperspectief vertrokken; niet uitgegaan van een beleidsnetwerk.

De effectiviteitsevaluaties leiden nogal eens tot de volgende waarnemingen: weinig doeltreffend beleid, teveel ambitie en daarmee hooggegrepen doelstellingen, overoptimisme over kostenbeheersing of aantallen bezoekers, aan bepaalde doelstellingen geen middel verbonden, mismanagement, ondoelmatig uitgevoerd beleid, duur, de veronderstellingen over het kunnen bereiken van doelstellingen bleken onjuist, diffuse verantwoordelijkheden (Bovens & 't Hart, 1996, 2001; Leeuw, 2008).

Is het werkelijk zo dat de overheid bij beleidsevaluatie overal de schuld van heeft of krijgt? Licht het aan de politici, aan de ambtenaren, aan het beleid, aan de burgers? Op dit vlak is een aantal stellingen te formuleren.

- Het ligt bij het niet bereiken van beleidssucces aan *de gestegen verwachtingen van burgers* ('rising expectations'). Burgers zien dat de welvaart toeneemt, de middelen voor informatievoorziening (ICT), voor nadenken (scenario's etc.), voor deliberatie. Waar zoveel meer kan dan vroeger verwachten (sommige) burgers voor elk probleem een oplossing. Misschien zijn beleidsdoelstellingen met het beleid wel in enigerlei mate bereikt maar aan de verwachtingen is niet (helemaal) voldaan. Daarmee sluipen dus ook 'verwachtingen' als criterium de evaluatie of nabespreking binnen.
- Het ligt aan politici: *die beloven teveel*. De beleidsdoelstellingen blijken te hoog gegrepen. Terugschroeven?
- Het ligt aan *ambtenaren*: die willen hun afdeling aan het werk houden, blijven budget 'opmaken' en denken vanuit hun functie (padafhankelijkheid). Dus als beleid geen succes is, zullen ze niet pleiten voor nieuw beleid (wat dat inhoudt is ongewis) maar bestaand beleid willen voortzetten met meer en/of betere middelen.
- Het ligt aan *de toerusting*. Het had met het beleid kunnen lukken als er meer aandacht was geweest voorde beleidsontwerp kwaliteit, en voor meer competentieen meer budget.....
- Het ligt aan de *adviesorganen en wetenschappers*: het aantal maatstaven om beleid te beoordelen is de laatste decennia toegenomen. Er is altijd wel een maatstaf waar een beleid niet aan voldoet. Zo gezien, kan beleid bijna niet een succes worden. Hoe dit ook gezien wordt, we worden samen steeds kritischer.

- De tijdfactor. Wat eerst nog als voldoende werd beoordeeld, daar wordt later geen genoegen meer mee genomen. De *tolerantie voor falend overheidsbeleid* neemt af (Bovens & 't Hart, 1996).
- Het ligt aan *de overheid*: teveel hooi op de vork.
- Het ligt aan *de gebrekkige interactie*: als beleidsvoerders eens beter zouden luisteren naar de werkers in de frontlinie of de burgers die de regeling moeten uitvoeren.....dan ...
- Het ligt aan *de omstandigheden*: gewoon pech gehad.
- Het ligt aan de wenselijkheid van *meer toezicht* en *beter kijken*: eens goed kijken want het lukt ze (de gezagsdragers die minder gezag verdienen) vast niet.

Misschien is dat allemaal waar. Dan is politiek een kwestie van repareren.

Met dat kijken, stoten onderzoekers ook op bijvangsten. Ze stoten op zaken die niet kloppen maar niet het hoofdonderwerp van onderzoek waren. Zo bleek bij een analyse van SNS dat ook sprake was geweest van fraude.

Dan ontstaan er natuurlijk ook rationalisaties voor beleidsfalen, zekere ontkenningen en mystificaties.

Risico's zijn onvermijdelijk. Je kunt geen omelet bakken zonder eieren te breken. Waar gehakt wordt, vallen spaanders. Wat op korte termijn een miskleun lijkt, kan op langere termijn anders uitpakken. Het had niet moeten maar de basis is gelegd onder mijn voorganger. Beleid maken is een kwestie van vallen en opstaan. De gemeenteraad is volop geïnformeerd; we dragen met zijn allen schuld. Als je nu maar lering trekt uit deze zaak, dan zal het de volgende keer beter gaan.

Dilemma: achteraf discussie vanuit andere criteria

Als de doelstellingen bereikt worden, is te spreken van een beleidssucces. Maar vaak is het resultaat een beetje 'mistig'. En dan wordt de discussie achteraf ook nogal eens in andere termen gevoerd dan voorafgaand aan de besluitvorming. Een voorbeeld.

Casus Megabanenmarkt

Een aantal jaren terug meldt De Volkskrant (281002) over het fenomeen *Megabanenmarkt*. Deze markt was in Amsterdam gehouden. De Banenmarkt moest de sociale dienst uit het slop trekken door tienduizend bijstandsontvangers om te toveren tot werkzoekenden. Het werden er 4100, waarvan dertienhonderd een baan vonden. Negentig werkgevers hadden meegewerkt en vijfhonderd ambtenaren hadden zich ingezet om de operatie te laten slagen.

De voorstanders van de Banenmarkt vonden het resultaat een succes: beter een half gevulde hand dan een lege hand, bij wijze van spreken. De tegenstanders riepen dat de markt mislukt was. Op basis van dezelfde cijfers.

De tegenstanders:

1. De verwachtingen waren te hooggestemd geweest. Vier op de tien bijstandstrekken waren bij nader inzien geheel of gedeeltelijk arbeidsongeschikt. Uiteindelijk gingen er 4100 bijstandstrekken in procedure.
2. Ongeveer 2500 mensen kwamen helemaal niet opdagen op de markt, wat illustreerde dat sommige mensen ten onrechte in de kaartenbakken hadden gezeten.
3. Kosten van het houden van de markt: 70 miljoen euro.

Voorstanders:

1. Door de markt zijn 7000 uitkeringen beëindigd. Dat doorrekenend naar uitkeringen per jaar betekent dit 73 miljoen. De kosten zijn met een bedrag van 70 miljoen lager dan het totaal van uitkeringen.
2. De markt heeft meer inzicht opgeleverd wat we met deze bijstandstrekkers 'kunnen'.
3. De dossiers zijn over het algemeen ook opgeschoond.
4. De sociale dienst kon een impuls gebruiken. Dat is gelukt.

Is het eerlijk dat de discussie na afloop in andere termen gevoerd wordt? Een evaluator zal aarzelen want hij moest uitgaan van de officiële doelstellingen. Politieke vertegenwoordigers en bestuurders denken echter meestal in andere termen over een beleidsprogramma dan in termen van de officiële doelstellingen zoals die ooit geformuleerd werden en door een gemeenteraad geaccordeerd zijn. Waarom?

1) De tijdgeest kan verschoven zijn.

2) Een probleem is in omvang toe- of afgenomen.

3) De electorale cyclus duurt vier jaar en deze actoren hebben misschien niet aan de wieg van een beleidsprogramma zoals de banenmarkt gestaan. Ze kennen de doelstellingen niet eens en zijn opportunistisch. Ze kiezen die criteria die hun politiek uitkomen.

Zie daar een vraagstuk waar een evaluator mee moet omgaan. Advies: evalueer niet uitsluitend in termen van oorspronkelijke doelstellingen maar betrek andere criteria erbij. Wat betekent het evaluatieresultata in termen van andere perspectieven?

3 WAAROM EVALUATIES STARTEN: DE ARGUMENTEN

Voor evaluatie kunnen vaak veel argumenten opgevoerd worden. Waaraan valt te denken?

Evaluatie is nuttig als gevolg van de financiële gang van zaken in overheidsorganisaties. Daarom kende het ministerie van Financiën decennia terug al een afdeling die zich bezighield met beleidsanalyse. Daarin was een prominente plaats ingeruimd voor evaluatiestudies (en ook voor heroverwegingen). Maar ook om andere redenen is evaluatie nuttig. Ik geef een aantal redenen voor evaluatie.

1. *Het heroverwegingsargument:* Evaluaties kun je als afd. financiën laten verrichten omdat je dat beleid wilt opsporen dat je ingekrompen wilt zien als er ineens bezuinigd moet worden. Het is handig voor een minister van Financiën of een wethouder Financiën om op elk moment over een lijst te beschikken van falend beleid of geslaagd beleid en uitgewerkt beleid. De evaluatie verschaft de verantwoordelijke bestuurder dan mogelijkheden om in begrotingsonderhandelingen suggesties te doen.
2. *Het begrotingsargument:* Tegenwoordig bevatten programmabegrotingen van gemeenten en provincies een antwoord op de drie W's: wat willen we, wat doen we ervoor en wat mag het kosten? Dan wordt evaluatie een logische vier component die antwoord geeft op: wat levert het op? Het antwoord daarop is weer van invloed op de eerste W: wat willen we? En zelfs voor de tweede W: wat doen we ervoor?
3. *Het verantwoordingsargument:* Waar gaat het geld voor het beleid eigenlijk heen? Wordt het goed besteed? Wordt het geld wel gebruikt waarvoor het bedoeld was? Als een project financieel uit de hand gelopen is, moeten partijen bereid gevonden worden om de rekening te betalen. De evaluatie moet *de schoenlepel* zijn om de schoen aan te trekken, lees het verlies te dragen: *het schoenlepel-effect van beleidsevaluatie*.
4. *Het argument om beleidshomeopathie te bestrijden:* Achter beleid gaan vaak ongetoetste *veronderstellingen* over doel-middelrelaties schuil. Bijvoorbeeld: als we iemand een straf geven na een vergrijp of misdrijf zal dat heilzaam werken; na een straf neemt de kans op recidive af. Dat is een veronderstelling. Het kan nuttig zijn om te evalueren of veronderstellingen achteraf gezien juist zijn geweest. Zijn ze niet juist dan is het beleid mogelijk een vorm van charlatanerie. Beleidsevaluatie is een hulpmiddel om *beleidshomeopathie* te bestrijden (Leeuw, 2008).
5. *Het meerwaarde-argument:* Van nieuw beleid is veelal niet duidelijk wat de meerwaarde is bovenop het beleid dat al bestaat of wat maatschappelijke processen al doen zonder overheidsinterventie. Een evaluatie van bij voorbeeld een subsidie kan dan duidelijk maken of organisaties een subsidie eigenlijk wel nodig hadden of deze slechts aanwenden om zo het budget dat hiervoor voorzien was aan iets anders te besteden. Beleidsevaluatie kan 'substitutie-effecten' aan het licht brengen.
6. *Het 'ik wil stoppen-argument':* Evaluatie is ook om geheel andere reden nuttig. Wie na verloop van tijd nadenkt over beleid stoppen of niet stoppen, bezuinigen op beleid of anderszins beleid denkt te gaan aanpassen (doelgroep inkrimpen of uitbreiden, middelen wijzigen, etc.) zal er niet omheen kunnen om te beschikken over onderzoek naar de beleidsuitvoering en beleidseffectiviteit. Evalueren dus (Brewer & DeLeon, 1983). Zo werd na evaluatie de Beeldende KunstenaarsRegeling (BKR) afgeschaft en de Wet Investeringsrekening.

7. *Het 'ik ben klaar'-argument*: Beleid evalueren is ook nuttig om te achterhalen of beleid effectief is en uitgewerkt (bv. een dam is aangelegd, een maanprogramma in de ruimtevaart is gerealiseerd, een ziekte is bestreden) of nooit effectief zal worden. In beide gevallen is bezinning en een 'verhaal' nodig: gaan we het beleid stoppen of gaan we toch door?
8. *Het argument van het gelijke speelveld*: Een comparatieve evaluatiestudie naar een bepaald (zelfde) beleid in twee of drie landen kan aan het licht brengen welke instrumenten andere landen inzetten en daarmee bereiken en hier niet worden ingezet. In verband met een discussie over 'level playing fields' is het nuttig beleidshiaten op te sporen. Dit argument speelde bij voorbeeld in de jaren tachtig van de vorige eeuw bij het technologisch innovatiebeleid van de rijksoverheid. Ik heb het zelf gehoord.
9. *Het argument van zicht krijgen op onbedoelde voor- of nadelen*: Beleid kent vaak neveneffecten, aldus de rechtssocioloog Kees Schuyt. Als die neveneffecten gunstig zijn, kunnen ze een argument vormen om tegenvallende resultaten op het vlak van de hoofddoelstelling te 'compenseren'. In geval van neveneffecten zul je wel een bepaalde aanpak moeten kiezen anders kom je niet achter ongewenste effecten. Zie de casus Megabanenmarkt. Opschonen van bestanden van werkzoekende cliënten bleek een grotendeels bereikt positief nevendoeel.
10. *Het 'ik wil bijsturen'-argument*: Wie beleid ontwikkelt, wil bijsturen als dreigt dat de uitvoering niet de goede gewenste kant opgaat. Daarvoor kan een *early warning-evaluation* zijn hebben.
11. *Het show-argument*: Veel beleidsevaluatie is bedoeld om de resultaten ervan zo mogelijk direct en op korte termijn te benutten. De argumenten pro evaluatie laten dat zien. Het komt echter wel voor dat dit niet het geval is. Evaluaties worden ook wel eens gestart omdat ze een onderdeel van een breder beleid betreffen waarvan verondersteld wordt dat het juist een succes is. Beleidsvoerders kunnen belang hebben bij een *show-effect*, omdat beleidssucces dan afstraalt op een breed programma en een verantwoordelijke bestuurder. Anders gesteld: beleidsvoerders hebben helemaal niet het voornemen om een beleid aan te passen of te stoppen of uit te breiden. Leerdoelstellingen zijn er niet echt. Ze willen alleen laten zien hoe goed het beleid is.

De aanleiding voor evaluatie: over sunsets en meer

In de Verenigde Staten gelden andere rigoreuze argumenten. Bijna alle beleid wordt stelselmatig aan evaluatie onderworpen omdat beleid een 'sunset' of horizon kent. Beleid stopt na enkele jaren automatisch ('policy termination'). Het loopt 'altijd af'; de valbijl valt. Maar voorafgaand aan de *valbijlwerking* is er een verplichte evaluatie. Die evaluatie kan reden zijn om het beleid toch voort te zetten. De evaluatie is namelijk de opstap naar een politieke discussie.

In veel staten wordt een dakpansconstructie gevolgd. Elk jaar komt als het ware vanzelf, vanwege sunset legislation, een deel van het totaal aan beleidsprogramma's in aanmerking voor evaluatie.

In Nederland kennen we de constructie van sunset legislation niet zo, maar wel komen vaker evaluatiebepalingen voor. Denk nationaal maar aan wetten als de Wob, WUB en Wabm die zo'n artikel kenden.

Afgezien daarvan, hoe staat het met provinciaal of gemeentelijk beleid? Er zijn in Nederland geen decentrale overheden die alle beleid evalueren, ook niet volgens een dakpansgewijze aanpak.

4 FUNCTIES EN NEVENFUNCTIES VAN BELEIDSEVALUATIE

Er is na het voorgaande iets bekend over wat evalueren is, wat een klassieke beleidsevaluatie is, wat de condities zijn voor het succesvol evalueren, vanuit welke criteria te evalueren valt en wat zoal de argumenten pro evalueren zijn. Er is ook al een dilemma genoemd bij evalueren. Wat zijn de functies van evalueren?

Functies van beleidsevaluatie

Is het verstandig om na te denken over functies van beleidsevaluatie? Zeker. Ik geef een voorbeeld.

Er is onderzoek gedaan naar om en nabij 500 vooral buitenlandse evaluaties van beleidsprogramma's gericht op geweldspreventie en het tegengaan van geweld in de openbare ruimte. Van die programma's waren er *slechts 9 effectief* (Van der Knaap e.a., 2006, 2008; Leeuw, 2008). Wat doelgericht bleek was *'hot spots policing'*: doelgericht politieoptreden. En 'leren' bleek nuttig: *kijk zelf hoe dingen uitpakken en doe het dan beter*.

Veel beleidsevaluaties laten zien dat beleidsinterventie amper of geen doel treft (Leeuw, 2008: 26). Gebruiksgerichte beleidsevaluatie is dus zo gezien urgent. Beleidsevaluatie kan natuurlijk om allerlei redenen verricht worden en functies hebben.

- 1) *Het kwaliteitsargument*. Of een beleidsinterventie werkt of niet werkt, weet een overheid vaak pas achteraf, zeker als er voorafgaand aan een beleidsinterventie geen experiment heeft plaatsgevonden om uit te vinden of en hoe de interventie werkt (Leeuw, 2008: 25). Als het onbekend is of iets werkt, is het in het algemeen verstandig uit een oogpunt van *'kwaliteit van politiek bestuur'* om te achterhalen of beleid werkt of niet. Als allerlei beleid niet werkt zoals beoogd is, moet terugkoppeling plaatsvinden omdat het schadelijk is, in geval van schaarste, om door te gaan met dwaze voornemens die niet werken. Beleid dat niet beantwoordt heeft aan doelstellingen en verwachtingen moet gestopt worden. Doorgaan werkt verspillend en verloederend.
- 2) *De leerfunctie*. Evaluatie heeft in potentie een leerfunctie. Leeuw, Sonnichsen en Rist bepleiten in *'Can governments learn?'* meer aandacht voor evaluatie als middel voor 'leren'. Leren betekent dat na evaluatie ook een grondige bezinning plaatsvindt in ambtelijk-bestuurlijke kringen op het geëvalueerde vraagstuk en pogingen worden gedaan tot het stoppen van beleid, tot opvolging of een andere beleidsverandering ('policy change'; zie Hogwood & Peters).
- 3) *De verantwoordingsfunctie*. Beleidsevaluatie kan een middel zijn dat dient voor verantwoording (van ingezette budgetten) naar de samenleving. Immers, er mag verwacht worden dat het budget dat gaat naar tal van beleidsprogramma's ook de moeite waard was, is en zal zijn. Veelal gaat het immers om gebruik van belastinggeld van burgers.
- 4) *De politieke massagefunctie in een netwerk*. Een evaluatie is soms niet wat het lijkt. De evaluatie van de Floriadesof moest de handreiking zijn om te rechtvaardigen dat een gemeenteraad bijdroeg aan het verlies en het dagelijks bestuur van elke gemeente 'uit de wind bleef' (aan ons lag het niet). In feite moeten de politieke fracties met het onderzoek gemasseerd worden om de beurs te trekken voor een verlies.
- 5) *De functie van 'feedback loop'*. Ter wille van feedback in een politiek systeem is discussie van belang over de vraag of een beleidsinterventie iets uitmaakt of dat de

beleidsinterventie juist sterker moet worden doorgezet of beleid net zo goed of zelfs beter kan worden afgeschaft. In een politieke democratie moet bezinning plaatsvinden. Beleidsevaluaties leveren dus denkstof op voor opstellers van nieuwe verkiezingsprogramma's.

Nevenfuncties van evaluatie

Kunnen evaluaties, zoals van onafhankelijke onderzoekers of visitatiecommissies, ook onbedoelde en ongewenste effecten hebben? Inderdaad. In de literatuur over prestatie management wordt hier betrekkelijk veel aandacht aan besteed (bijv. De Waal en Kercklaan, 2004: 79-88).

Ik volsta hier met enkele mogelijke effecten te noemen:

- het opluchtingseffect,
- het ossificatie-effect,
- de prestatieparadox,
- het proceduralisme-effect en
- het McDonaldization-effect.

1) Het opluchtingseffect.

Wat zich bij evaluatie wel voordoet is dat er officieel wordt gesteld dat evaluatie gericht is op verantwoording en op leren maar dat de werkelijkheid een stuk cynischer is. Mensen lopen soms niet warm voor de evaluatie omdat men '*de wet van de tolerantie voor suboptimaliteit*' kent. Die wet houdt in dat elke organisatie in bepaalde onderdelen of opzichten niet helemaal voldoet aan de gewenste standaarden maar dat wordt door de vingers gezien 'zolang het maar niet uit de hand loopt'.

Die wet werkt indirect door in evaluaties. Soms is evaluatie met een visitatiecommissie een zware gang die bestuurders en anderen moeten maken, waarbij leren voor hen geen hoofdzaak is (immers, men worstelt al met suboptimaliteit). De deelname aan een visitatie is soms onvermijdelijk omdat juist het niet deelnemen aan een benchmarking, monitoring of visitatieproces vragen oproept. Maar men is vaak weer blij als de evaluatie erop zit en 'men weer gewoon verder kan'. Immers bestuurders kennen de sterkten en zwakten van een organisatie of elementen of aspecten vaak wel, maar men beseft ook 'je moet wel de goede mensen vinden' (vacatures?), en 'een team moet een team zijn maar dat valt niet altijd af te dwingen' (een organisatie zit vol menselijke zwakten) of er kan zich plotseling een incident voordoen met impact en dat kun je niet "plannen" '.

Kortom, het streven naar leren kan concurreren met de dagelijkse praktijk van 'hopen'. Dat impliceert dat evaluaties soms wel als een opluchting worden gezien, onder het motto '*he,he, dat hebben we ook weer gehad.*'

2) Het ossificatie-effect: beleidsevaluatie kan de mogelijkheid tot innoveren remmen of tot angst om te innoveren leiden (Leeuw, 2001: 77).

De leerfunctie van beleidsevaluatie is heel belangrijk maar bezinning en het zich los maken van het bestaande is niet eenvoudig? Er is namelijk veel *padafhankelijkheid* ('path dependency'; Werle, 2007). Beleid is doorgaans geen pennestreek op een leeg vel: *waarheen je beleidsmatig heengaat, wordt sterk bepaald door waar je vandaan komt*. Dat wijst op padafhankelijkheid.

Anders gesteld: beleidsprocessen worden voorgestructureerd door historisch gegroeide gewoonten, tradities, heersende normen en waarden, bestaande regels en werkwijzen. En

bedenk: er is vaak al een beleid waardoor een beleidsvoorstel een uitbouw van het bestaande beleid of een aanpassing ('beleidsopvolging') of terminatie is, of een 'replacing'(nieuwe schoenen voor oude). Deze eerdere groei en deze vroegere keuzen beperken 'het andere' te zien en leggen beperkingen op om zich los te maken van een eenmaal ingeslagen weg. Beleidsevaluaties zijn vaak evaluaties van voortgang op *hetzelfde* pad en brengen geen andere route in beeld.

Een voorbeeld van wat padafhankelijkheid is in de praktijk. Dat de schoolstrijd van begin twintigste eeuw opleverde dat bijzondere scholen mogelijk werden, impliceerde dat een weg gebaad werd naar verzuiling, dus naar onderscheid tussen algemeen en bijzonder onderwijs op katholieke of protestants-christelijke grondslag, die tot in de 21^{ste} eeuw doorloopt. Dat heeft onder andere negatieve effecten gehad: doordat wat de een kreeg, de ander niet onthouden 'mocht' worden, is het onderwijs duurder geweest dan in geval er alleen openbaar onderwijs was geweest (zie het proefschrift van Koelman uit 1987; promotor was de hoogleraar en latere onderwijsminister Jo Ritzen).

Is dit ossificatie-effect concreet aangetroffen? Rapportages over verricht benchmarkonderzoeken leiden zelden of nooit tot het volledig op zijn kop zetten van het beleid van de organisatie (Abma en Korsten, 2009). Deze rapportages bevatten doorgaans ook geen aanbevelingen. Vaak omvat een rapport de vergelijking van 10 of 15 organisaties. Wil men de vergelijking begrijpen dan moeten gebruikers een eigen verhaal maken over hun eigen organisatie, dus een vertaalslag. Dat bevordert doorwerking van benchmarking niet. Als deze al tot aanpassingen van beleid leiden is dat doorgaans een zaak van 'stukjes en beetjes'.

3) De *performance paradox* kan optreden (Van Thiel en Leeuw, 2003).

Wie voortdurend prestaties evalueert, denkt te kunnen verbeteren maar kan op een bepaald moment niet verder verbeteren. Er komt als het ware een steeds *verdere afvlakking van de 'meeropbrengst'*. En organisaties die grossieren in evaluaties zijn na verloop van tijd niet meer vanzelfsprekend de beste of meest effectieve organisaties. Waarom? Om drie redenen.

Ten eerste, de betrokkenen bij evaluatie leren zich instellen op evaluaties (Leeuw, 2001: 76). Denk aan de politie die beter moest presteren. Dan gaat men verkeerscontroles houden en vliegen de prestaties in termen van aantallen bonnen met sprongen omhoog. De politie gaat *makkelijke vangsten* organiseren. Een pervers effect van 'performance'-meting.

Ten tweede, wie presteert en evalueert en zich verbetert, kan op een gegeven moment nauwelijks nog verder verbeteren. En organisaties die zwak presteren kunnen, vanuit de wet van de stimulerende achterstand en evaluatief geïnspireerde verbeterinitiatieven, ineens sprongen vooruit maken en de al langer goed presterende organisatie bijna inhalen. Het gevolg kan zijn dat de zich steeds evaluerende organisatie zich bijna niet meer kan onderscheiden van de organisatie met minder prestaties (die na enkele evaluaties ineens grote sprongen maakte).

Een derde mogelijkheid is dat de organisatie die aan evaluatie wordt onderworpen de onderdelen waarop ze slecht scoort af gaat snijden. Scholen doen het volgende: beter leerlingen op vier vwo een jaar over laten doen dan laten zaken voor een eindexamen. Waarom? Ze zouden anders de examenresultaten in het vijfde jaar toch maar bederven. Universiteiten neigen ertoe vakgroepen met weinig onderzoeksoutput op te heffen of samen te voegen met een andere groep.

Wat constateer ik op grond van de analyse? De performanceparadox is in verband met beleidsevaluatie nog niet (empirisch) waargenomen. Dan zouden evaluaties een zeer grote invloed gehad moeten hebben. Waarom om andere redenen dan wel die performanceparadox

kan of zal optreden? Omdat 'new public management' heeft geleid tot prestatieafspraken. Overall waar prestatieafspraken over aantallen ontstaan, komt de neiging tot *perverse effects* op. Als een overheid onderwijsinstellingen financiert op diploma's bevordert dat de tendens om toch vooral diploma's te verstrekken ook in die gevallen dat dit niet gerechtvaardigd is.

4) *Het proceduralisme-effect.*

Er is nog een vierde neveneffect van evaluatie te noemen. Het is mogelijk dat voortdurende evaluatie van beleid leidt tot steeds meer verbeteringen die niet betrekking hebben op de veronderstellingen achter een bepaald beleid maar tot 'next best' - verbeteringen in termen van processen, zoals: meer verantwoording, meer 'dash boards' om de voortgang van een beleid te volgen, (nog) meer monitoring, meer protocollen (Leeuw, 2001: 79). De evaluator vlucht door middel van aanbevelingen in een *cumulerend procesfetisjisme* (of in andere woorden een proceduralisme-effect). Dat kan onwenselijk zijn omdat zo een verdere bureaucratisering ontstaat.

Wat constateer ik ten aanzien van het proceduralisme-effect? Dit effect komt zeker voor naar aanleiding van evaluaties (Leeuw, 2001). Hoe dat kan? Omdat procedures veranderen makkelijker gaat dan de beleidsinhoud veranderen.

5) *Het McDonaldization-effect.*

Ik kom bij een vijfde neveneffect. De McDonalds-restaurants worden gekenmerkt door dat ze op alle plaatsen in de wereld hetzelfde leveren: niet meer, niet minder, niet anders. Een BigMac is overall dezelfde BigMac. Zo kan het ook gaan in de evaluatiewereld. Evaluatie kan uitwijzen dat een bepaald doelstelling niet gehaald is. Het gevolg daarvan kan dan zijn dat via organisatievergelijking toegewerkt wordt naar steeds hetzelfde resultaat dat wel haalbaar is. Een voorbeeld is te vinden bij de politie. Een politie-agent die op straat werkt, moet per dag gemiddeld zus en zoveel bonnen uitschrijven. Ook voor andere 'street-level bureaucrats' kan deze kwantificeerbare aanpak ontstaan. Volgens Ritzer is hier sprake van de '*irrationality of rationality*'.

Het effect van McDonaldization kan optreden bij een organisatie met lopende band-processen (bij een machinebureaucratie om de indeling van Mintzberg in organisatietypen te volgen) maar niet bij een geheime dienst of een sterk op stimulering, coördinatie en samenwerking gerichte provinciale organisatie.

Samenvattend, we moeten de kans op het optreden van deze neveneffecten niet overdrijven. Het effect van de performanceparadox ligt niet steeds op de loer bij alle typen organisaties. Maar de kans is wel groter dat padafhankelijkheid optreedt en leerprocessen remt die juist 'out of the box'-denken vragen (Werle, 2007). Proceduralisme acht ik wel een dreigend effect. De Algemene Rekenkamer is zich hiervan bewust geworden, zo meldt de voormalige directeur Frans Leeuw (2001) ons.

5 EEN STAPPENPLAN VOOR EEN KLASSIEKE BELEIDSEVALUATIE

Hoe zou een stappenplan voor beleidsevaluatie eruit kunnen zien? Ter eerste oriëntatie een houtskoolschets.

Een provinciebestuur heeft een medewerkster aangenomen die tot taak krijgt om een provinciaal beleidsprogramma op het gebied van (uitvoering van) sociaal-economisch beleid te evalueren.

Vraag 1: Hoe kan ze te werk gaan?

Wie weinig van beleidsevaluatie weet, doet er goed aan enkele boeken ter hand te nemen en tijdschriften te raadplegen als *Evaluation*, en *Beleidswetenschap*. Qua boeken is te denken aan de volgende inleidende werken:

A *'Overheidsbeleid'* onder redactie van A. Hoogerwerf & M. Herweijer, Samsom, Alphen, 2008 (8^e druk). Deze bundel bevat hoofdstukken als: De inhoud van beleid, Beleidsinstrumenten, De uitvoering van beleid, De evaluatie van beleid.

B *'Beleidsevaluatie'* onder redactie van J. Bressers & A. Hoogerwerf (red.), Samsom, Alphen, 1995 (3^e druk). Dat boek bevat hoofdstukken als: Het beleid en het beleidsveld, Het vaststellen van doelbereiking, De bijdrage van het beleid aan doelbereiking, De verklaring van de effectiviteit van beleid, Presentatie en gebruik van evaluatie-onderzoek.

Vraag 2: Welke evaluatieve aanpak in de vorm van stappenplan kan de medewerkster kiezen? Wij dachten met haar mee en stelden de volgende checklist in de vorm van een stappenplan op. Bepaalde stappen kunnen wellicht worden overgeslagen en andere toegevoegd, in concrete gevallen.

Stappenplan voor evaluatie

Omdat er keuzen bij evalueren mogelijk zijn, heeft een stappenplan zin om die keuze te verhelderen.

Stap 1: Inventariseer de bestaande informatie

- Bestaat er een *dossier* van het beleidsprogramma en van in- en uitvoeringsmaatregelen? Zo ja, vul het aan. Zo nee, leg het aan.
- Welke *afdeling* is verantwoordelijk voor het provinciaal beleid op het gebied van sociaal-economisch beleid? Wie heeft destijds het beleid ontwikkeld? Was het een betrekkelijk makkelijk ontwerpproces resp. wat waren de barrières?
- Bestaat er een *verbinding* met Europees en/of nationaal beleid? Welke?
- *Perspublicaties*: zijn die er geweest? Is er een knipseloverzicht? Wat is het grote toegekende voordeel van beleid en wie draagt dat uit? Wat is de hoofdkritiek en wie heeft die vooral?

Stap 2: Breng de bestaande informatie over het beleidsprogramma nader in kaart

- Wat is de *beleidsinhoud* van het te evalueren beleid in termen van doelstellingen, doelboomstructuur, in te zetten middelen, tijdspad, verwachtingen over doelbereiking? Is sprake van doelstellingen zonder middelen om ze te realiseren of middelen zonder doelstelling over de inzet?
- Welke *veronderstellingen* gaan achter het beleid schuil over gedrag?
- Valt een *beleidstheorie* (het geheel van veronderstellingen over oorzaken en gevolgen, over doelen en middelen) te reconstrueren? Zo ja, hoe ziet die theorie eruit?

- Wat voor een *geschiedenis* kent dit beleid?
- Wat toont de *totstandkominganalyse* van het beleid qua beoogde beleidsdoelstellingen, een eventuele doelstellingenhiërarchie, veronderstellingen over gedrag van verschillende actoren en eventueel verwachte nevendoelestellingen van beleid?
- Bestaat er een *invoeringsprogramma*?
- Welke zijn de *doelgroepen van het beleid*? Hebben zij informatie over de reactie op de beleidsuitvoering in de vorm van een jaarverslag of anderszins?
- Is sprake geweest van een *nulmeting* of niet?
- Heeft er al ooit *eerder* een evaluatie van het betreffende beleid plaats gevonden? Zo ja, wat hield die in en wat waren de bevindingen?
- Heeft er al ooit eerder evaluatie van dergelijk beleid plaats in *een andere provincie* gevonden? Wat houdt die evaluatie in en wat kwam eruit qua resultaten/uitkomsten?

Stap 3: Voer gesprekken met relevante actoren om zo een verdere schets van het beleidprogramma en de uitvoering te krijgen (verder inzoomen)

- Wat zijn de *doelstellingen* van dat beleidsprogramma in de woorden van de meest betrokken medewerker? Wat waren de *verwachtingen* per doelstelling? Welke veronderstellingen zoal?
- Welk *sturingsconcept* ging schuil achter het beleid? Welke instrumenten?
- *Uitvoering*: wie voert het programma uit: de provincie zelf, gemeenten of derden? Waarom?
- *Politiek*: welke gedeputeerde had/ heeft dit beleid in portefeuille? Geërfd van voorganger? Was het politiek omstreden? Is het beleidsprogramma destijds makkelijk door de commissie (welke?) en Provinciale Staten geloosd? Kan het verslag te voorschijn komen?
- Is het beleid sinds het in uitvoering is *onderwerp van discussie* geweest in GS of Provinciale Staten? Zijn er vanuit de samenleving of het maatschappelijk middenveld reacties geweest op het beleid en/of de uitvoering en/of de doelrealisatie?
- Welke *belangen* zijn in het spel? Welke actoren uit de samenleving hebben voor- of nadeel van het beleid gehad? Wie is aanspreekbaar buiten de provinciale organisatie? Wat is de kijk van werkgever en werknemersorganisaties op dit beleid?

Stap 4: De evaluatieopdracht en een eventueel evaluatieplan

- *Van wie* komt de wenselijkheid of eis van een evaluatie?
- Is de evaluatieopdracht *duidelijk*? Voor wie is die bestemd? Hoe precies moet de evaluatie zijn?
- *Aan wie moet voor wanneer hoe* uitgebreid of beperkt worden gerapporteerd?
- Maakt de evaluatieopdracht deel uit van een *breder geheel van evaluaties*? Zo ja, zijn er aanwijzingen of indicaties over omvang en richting van de evaluatie?
- Zijn er andere opmerkingen te vinden over *de wijze van evalueren*, bijvoorbeeld in een Statencommissie of een motie?
- Moet/mag de evaluatie *intern* plaatsvinden of moet/mag die extern worden uitgevoerd? Als extern mogelijk is: bestaat er een gremium dat zich bemoeit met de vraag: intern of extern? Hoe verhoudt dit zich tot het tijdpad van rapportage?
- Is een *evaluatieplan* gewenst en zo ja, moet een orgaan of ambtelijk team dit plan *goedkeuren*?
- Bestaat er een link naar de actuele provincie-, lands- of Europese politiek? Wil iemand het *bestaande beleid veranderen* en is er een legitimatie voor verandering nodig?
- Welk *type evaluatie* is gewenst? Zie verder.
- Is sprake van een *begeleidingscommissie* voor het evaluatieonderzoek? Zo ja, wat is daarvan de samenstelling, taak en werkwijze?

- Is bekend wie na afloop allemaal *gebruiker* is van de evaluatiebevindingen? Zo ja, is bekend wat hun verwachtingen zijn over het onderzoek? Worden ze betrokken in een evaluatie?
- Is in de opzet sprake van *goede maatvoering*: is de evaluatie niet te uitvoerig, te duur, komen de resultaten te laat?

Stap 5: Keuze en inhoud van de evaluatie-opzet

- Evalueren betekent beoordelen in een vooruitblik of terugblik: *ex ante evaluatie* of *ex post*?
- Ligt het in de rede om het bij een *terugblik* te laten of staat de evaluatie ook in het kader van *prospectie*?
- Is de evaluatie *extern* of *intern*? Hoe uitgebreid?
- Wordt gekozen voor een evaluatie waarbij nagegaan wordt of doelen bereikt zijn (doelbereiking), welke de uitvoering was en of beleid daadwerkelijk de gewenste effecten opleverde?
- Is een *goal free - aanpak* gewenst door juist te kijken naar verwachte of onverwachte neveneffecten van beleid?
- Is beleid al zo lang in werking dat niet alleen prestatie (output) maar ook effecten (outcomes) zichtbaar zijn? Als geen effecten zichtbaar zijn is dan een *early warning evaluation* nodig en mogelijk: door te kijken of de uitvoeringsmaatregelen nog in lijn zijn met de beoogde effecten?
- Geldt de 'wet van de evaluatieve gebreken': geen/wel nulmeting, te vroege/late meting, before-aftermodel met experimentele en controlegroep, enz.? (zie verderop in deze tekst)

Stap 6: Reconstructie van de beleidsinhoud

- Wat was de *belangrijkste beweegreden* voor het beleid?
- Hoe luidt de *beleidstheorie achter het beleid*: Wat zijn de doelstellingen? Was sprake van meetbare doelstellingen? Kwamen officieuze doelstellingen voor? Welke doelstellingenboom? Welke doelstelling was de belangrijkste?
- Veranderden de doelstellingen sinds de vaststelling van het beleid of niet?
- *Middelenkeuze*? Nog verandering?
- *Verwachtingen* t.a.v. beleidsuitvoering, prestaties en doelstellingsrealisatie, geschiktheid en omvang van de personele en financiële middelen?
- Wat waren samengevat de *belangrijkste veronderstellingen* achter het beleid?
- Zijn daarmee voldoende *maatstaven* voor beoordeling verkregen?

Stap 7: Prestatie- en effectmeting

- *Operationalisatie* van prestaties en van doelstellingen?
- Welke prestaties en bedoelde effecten zijn *waargenomen*, door wie en wat is de kwaliteit daarvan?
- *Additionele* informatieverwerving? Bij wie: welke waarnemingseenheden? Wat is de beste waarnemingsmethode (survey, observatie, expertmeeting)? Is een combinatie gewenst?
- Is sprake van *doelbereiking*: is beleid een succesnummer?

Stap 8: Uitvoeringsanalyse

- Hoe verliep de in- en uitvoering van beleid qua transparantie voor doelgroepen, personele uitvoeringscapaciteit, beleidsmiddelenkeuze, keuze van uitvoerders?
- Wat is de *satisfactie* bij betrokken actoren over de beleidsinhoud en uitvoering? Welke knelpunten volgens wie? (Framing?)

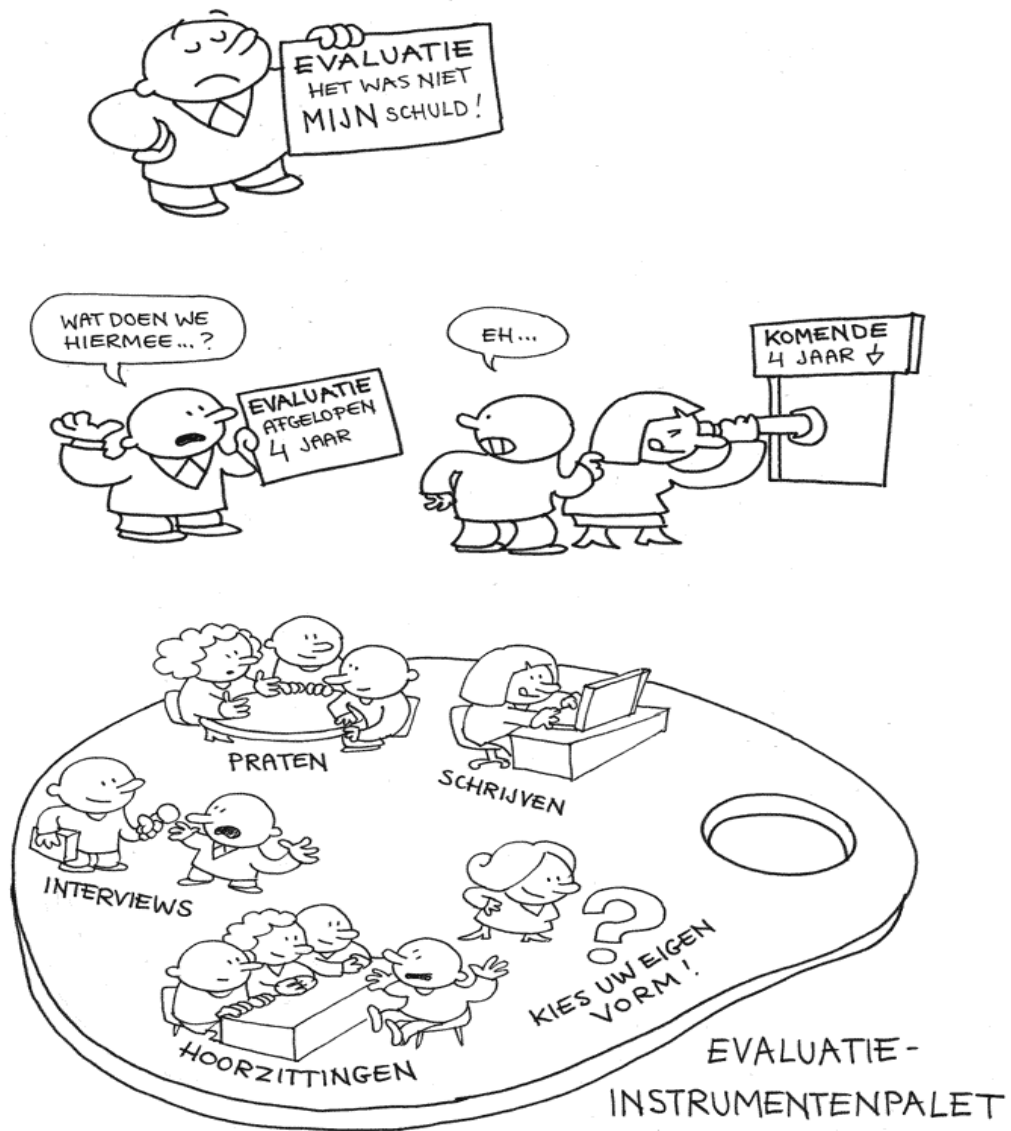
- *Succes- en faalfactoren?*
- *Wat was de ‘discourscoalitie’?*
- *Veroorzaakte het beleid ook de gewenste effecten of was sprake van externe factoren en omstandigheden?*

Stap 9: Rapportage

- *Ga nogmaals na wat de gebruikers verwachten en vergelijk die met bevindingen.*
- *Wat zijn de hoofdconclusies in termen van empirie?*
- *Wat is de betekenis van de bevindingen in het licht van enkele scenario’s, zoals: is er reden het gehele beleid af te schaffen? (‘termination’); is er reden het gehele beleid onverkort voort te zetten? (‘continuïteit’); op naar een proces van beleidsopvolging?; naar een nieuw paradigma? Argumenteer waarom welke aanbeveling te geven is.*
- *Stel het rapport op.*
- *Maak een samenvatting en stel een persbericht op.*

Er bestaan ook andere stappenplannen: zie bij Bressers (1988).

DEEL II: TYPEN BELEIDSEVALUATIE



6 DE KLASSIEKE BELEIDSEVALUATIE: EFFECTIVITEITSANALYSE

Beleidsvaluatie betreft de evaluatie van beleidsprocessen zoals bij voorbeeld de totstandkoming van beleid, de besluitvorming, de organisatie van het beleid, de in- en uitvoering van beleid, de effecten van beleid.

Het meeste onderzoek wordt gedaan naar prestaties en effecten van beleid. Een bekende aanpak bij het beoordelen van beleid is de effectiviteit van beleid (Bressers, 1993). Hierbij wordt antwoord gegeven op de vraag:

- wat zijn de doelstellingen achter het beleid (= gewenste effecten)?;
- worden die doelstellingen bereikt?;
- en zo ja, komt dat door het beleid zelf of iets anders, iets dat geen onderdeel was van het beleid?

Dergelijk onderzoek is klassiek. Het wordt *effectiviteitsonderzoek* genoemd.

Doelbereikingsonderzoek is hiervan een belangrijk onderdeel naast de analyse van beleid en/of van andere omstandigheden die bijdragen aan de doelbereiking. Internationaal wordt in geval van doelbereikingsanalyse gesproken van *'the goal achievement approach'* of *'goal attainment'* (Weiss, 1972, 1976). Deze aanpak is natuurlijk gerechtvaardigd omdat een beleid beoordeeld mag worden op basis van wat officieel ook is uitgesproken en democratisch is vastgelegd en besloten. 'Beleidsvaluaties hebben vaker het karakter van een doelbereikingsevaluatie dan van een effectiviteitsevaluatie' (Bressers en Hoogerwerf, 1991; Bongers en Den Hertog, 2007: 36).

Een andere maatstaf is legitimiteit en een andere doelmatigheid van beleid. Het is dus ook mogelijk onderzoek te doen naar het bereiken van doelen met geringe middelen (doelmatigheid) of bij voorbeeld de acceptatie van beleid. Henri Potman (1989) is op dit onderwerp 'beleidsacceptatie' gepromoveerd. Een aantal jaren geleden gold de Wet geluidhinder, die toen omstreden was. Potman heeft de acceptatie van die wet gezien. Hij verrichtte dus een beleidsvaluatie.

Vertrek vanuit een beleidstheorie

Maar een maatstaf is bij evalueren eigenlijk niet genoeg. Aan beleid zelf zijn vaak al maatstaven te ontleen. Welke dan?

Belangrijke maatstaven voor beoordeling van beleid zijn te vinden in de *achterliggende gedachten* achter een beleid, dus de veronderstellingen. Bestuurskundigen spreken in dat verband van de beleidstheorie van een bepaald beleid (Hoogerwerf, 1984; Leeuw en Van de Graaf, 1988; Pröpper & Reneman, 1993; Van Heffen, 1996; Leeuw, 2003; Dunn, 2012). Achter in feite elk beleid gaat dus een 'beleidstheorie' schuil.

Een *beleidstheorie* is het geheel van veronderstellingen over oorzaken en gevolgen van een verschijnsel en ook over doelen of doelstellingen van beleid om iets met of tegen een verschijnsel te doen, de toegepaste beleidsmiddelen en de relatie ertussen (Van Heffen, 1996). Dat betekent dat achter een wet als de Wet Milieubeheer of om het even welke op modificatie gerichte wet of maatregel van bestuur, een beleidstheorie schuilgaat. Als er bij voorbeeld een nitraatmaatregel komt voor landbouwers, zit daar een redenering achter, een beleidstheorie.

De beleidstheorie is in feite *de motor* achter beleid die ervoor moet zorgen dat een beleidsinterventie *werkt* (Pawson en Klein Haarhuis, 2005: 45). De beleidstheorie wordt ook

wel de *programmatheorie* achter beleid genoemd (Pawson en Klein Haarhuis, 2005). Bressers (1993, 2008) spreekt van een *beleidsveldmodel*.

Formele evaluatie

William Dunn (2012: 324) spreekt in dit verband van formele evaluatie. Hij omschrijft formele evaluatie als volgt:

'an approach that uses descriptive methods to produce reliable and valid information about policy outcomes but evaluates such outcomes on the basis of policy-program objectives that have been formally announced by policy makers and program administrators'.

De belangrijkste assumptie achter deze formele evaluatie is volgens Dunn dat wat formeel als doelstellingen geannonceerd wordt ook daadwerkelijk iets zegt over de waarde of betekenis van een beleidsprogramma. Dat zal vaak wel het geval zijn, maar soms kan beleid ook een middel zijn om een kritiek wind uit de zeilen te nemen. Machiavelli zou het wel weten. In zijn geest kun je bij voorbeeld bedenken: je geeft als rijksoverheid organisaties voor ontwikkelingshulp budget voor uitvoering van ontwikkelingstaken om ze zo zoveel aan je te binden dat je ze domesticeert en monddood maakt. Door uitvoerders te *verafhankelijken* neem je tijdig kritiek op de omvang van het ontwikkelingsbudget wind uit de zeilen.

Wie gaat evalueren (ex ante of ex post), kan dus beginnen met de veronderstellingen (inclusief de doelstellingen) eerst op een rij te zetten en die te operationaliseren in onderzoekbare termen. Zijn die veronderstellingen niet duidelijk en niet precies dan kan beleid vermoedelijk ook geen gewenste doelstellingen bereiken. Mogelijk is het beleid dan een wazig compromis geweest of een vage vorm van cadeauverstreking of een vorm van (co)financiering waarbij aan de andere financierende partij overgelaten wordt om doelstellingen te specificeren.

Veelal wordt als eis gesteld aan beleid dat doelstellingen in beleidsnota's zijn geformuleerd. Dat moet bij voorkeur SMART gebeuren, dus moeten de doelstellingen Specifiek, Meetbaar, Actueel, Realistisch en Tijdsgebonden (dus SMART) zijn (Bongers en Den Hertog, 2007: 35). Dat wordt ook wel rationalistisch genoemd.

In werkelijkheid is overheidsbeleid een kwestie van cognitie en interactie of – anders gesteld – van uitdenken en uitvechten (Hoppe, 2012). Dat betekent dat beleid vaak ook een compromis is en inconsistenties bevat. Dat moet een evaluator openbaren maar daar moet hij ook mee leren leven. Het maakt dat zijn (of haar) vertrekpunt niet ideaal is.

Elke evaluatie begint dus met de analyse van de beleidsinhoud (wat is het beleid) en het ontwikkelen van een beeld over wat een overheid eigenlijk met het beleid nastreeft.

7 DE OPZET VAN EEN KLASSIEKE BELEIDSEVALUATIE: ENKELE DILEMMA'S

Hier komt ex post evaluatie aan bod. In het voorgaande is aangegeven dat een onderzoeksofzet idealiter zou moeten plaatsvinden op basis van voor- en nameting met controlegroepen. Kan dat wel?

Kan het zoals het moet of moet het maar zoals het kan?

Er moeten keuzen op het vlak van de onderzoeksofzet en methoden en technieken van onderzoek gemaakt worden. Een aantal belangrijke keuzen heeft betrekking op de volgende punten.

Figuur 1: *Aspecten van de keuze van een ofzet voor ex post beleidsevaluatie*

- 1 Ideaal is een voor- en nameting.
- 2 Ideaal is de keuze van een experimentele groep en een controlegroep.
- 3 Ideaal is een variatie in beleidsimpuls.
- 4 Ideaal is het meten van effecten, niet slechts van prestaties.
- 5 Ideaal is een tijdige opdracht tot evaluatie.
- 6 Ideaal is als de officiële beleidsdoelstellingen ook de echte doelstellingen zijn.
- 7 Ideaal is als de beleidsdoelstellingen stabiel blijven.
- 8 Ideaal is als de effecten hard zijn te maken.
- 9 Ideaal is als evaluatoren zich bewust zijn van de eventualiteit van een netwerkcontext.

Deze aspecten worden hierna toegelicht.

1 Ideaal is een voor- en nameting.

Er dient er in een ex post evaluatie vóór de beleidsimpuls of beleidsinterventie al een meting te hebben plaatsgevonden, dan kan vervolgens op een later moment of op meerdere momenten een nieuwe vergelijkbare meting plaatsvinden. Deze meting vooraf wordt een nulmeting genoemd. Een verschil tussen de voor- en nameting(en) moet dan, als de omstandigheden gelijk blijven, op de cruciale variabelen duiden op een beleidseffect. Het beste is als ook nog voldaan wordt aan een volgende eis.

2 Ideaal is de keuze van een experimentele groep en een controlegroep.

Strikt genomen kun je niet zeggen dat een verschil tussen voor- en nameting evident duidt op een beleidseffect want er kunnen andere invloeden zijn buiten het beleid die het effect bewerkstelligen. De onderzoeker moet dit nagaan.

Stel de onderzoeker bevindt zich in een land waarvan de overheid een beleid gericht heeft op de tempering van gebruik van zonnebanken omdat het aantal mensen met huidkanker toeneemt. Stel we beperken ons tot het menselijk gezicht. Als mensen de laatste vijf jaar een meer gebruind gelaat hebben na de zomermaanden, meer bezoeken aan huisartsen met klachten over huisirritaties worden gebracht en doorverwijzingen naar ziekenhuizen op dit vlak plaatsvinden, stel we weten dat, dan kan dat liggen aan:

- a) het vaker gebruiken van zonnebanken, en/of
 - b) een toename in de hoeveelheid zonne-uren, en/of
 - c) de kracht van de straling waaraan mensen zijn blootgesteld, en/of
 - d) aan het feit dat subtropische en tropische vakantiegebieden meer in trek zijn geraakt.
- Ik laat hier het gebruik van middelen met een bruiningseffect terzijde.

Het is gewenst om een experimentele groep te hebben en een controlegroep met gelijke kenmerken.

De experimentele groep onderscheidt zich van de controlegroep doordat die een beleidsimpuls krijgt en de controlegroep niet. Eigenlijk zou je zowel een voor- als een nameting moeten verrichten bij de experimentele groep en de controlegroep.

3 *Ideaal is een variatie in beleidsimpuls.*

Misschien dat de evaluatoren ook oog hebben voor het op zo doelmatig mogelijke wijze bereiken van de doelstellingen. Men zal dan proberen om de overheidsinterentie te differentiëren naar kracht en instrument. De lichtste route is favoriet als je zo even veel bereikt als met een zwaar of duurder instrument. Om dat vast te stellen, moet er sprake zijn van twee of drie experimentele groepen.

In het lokale verkeersveiligheidsbeleid is zo een aanpak denkbaar. Je kunt de verkeersveiligheid bevorderen door de verkeerssnelheid te temperen via elkaar snel opeenvolgende bobbel in de weg of via elkaar niet snel opvolgende bobbel. Of via een combinatie van bobbel en bepaalde kilometerzonerings en eventueel wegversmallingmaatregelen.

4 *Ideaal is het meten van effecten, niet slechts van prestaties.*

In de beleidspraktijk wordt gekozen voor een beleid, waarbij meestal bepaalde instrumenten ingezet worden. De beleidsuitvoering impliceert dan toepassing van die instrumenten. In een verkeersveiligheidsbeleid impliceert dat bij voorbeeld een instructie geven aan bepaalde personen om over te gaan tot aanschaf van borden voor 30-kilometerzones en/of van materiaal voor de aanleg van een wegversmalling. De feitelijke aanleg is dan *de beleidsprestatie* (output). Het bereiken van minder verkeersincidenten en feitelijk verkeersslachtoffers is het gewenste *effect* (outcome). In de praktijk wordt vervolgens vaak gezegd dat een beleidsevaluatie de gewenste effecten in beeld moet brengen. Zijn die effecten wel of niet opgetreden? Wat laat evaluatie echter vaak zien?

Dat het nog te vroeg is geweest om het beleidsevaluatieonderzoek te verrichten. Effecten zijn door dat vroege moment nog helemaal niet zichtbaar. Men meet slechts dat de drempels of wegversmallingen zijn aangelegd. In de praktijk ontstaat vaak verwarring over het meten van prestaties of effecten. Nogal eens worden prestaties voor effecten aangezien.

Stel dat een overheid een subsidie geeft aan een schouwburg en daarbij eist dat de schouwburgdirectie zich richt op culturele verheffing van de stadsinwoners. Hoe dat vast te stellen? Vaak zal een directie dan aangeven dat er bij voorbeeld vijf typen programma's zijn: het lichte muziekgenre, klassieke muziek, toneel, cabaret en ballet. De directie wil naar een gemiddelde stoelbezetting van 60 procent, over alle voorstellingen in een jaar gezien. Wat blijkt bij een beleidsevaluatie? Er is geen nadere keuze gemaakt van het gewenste effect, dus wat gerekend wordt tot culturele verheffing en wat niet, maar slechts de stoelbezetting is aangegeven. Dan beantwoordt de beleidsevaluatie niet aan haar doel. De evaluatie zegt dan niets over de verheffing terwijl dat toch een doel was. Kon dit ook niet? Toch wel. De directie had in overleg met de subsidieverstrekker vooraf kunnen bepalen dat verheffing te operationaliseren is als:

- De aanwezigheid van jaarlijks meer dan in totaal minimaal 3000 bezoekers van klassieke muziek en toneel;
- Een gewenste stoelbezetting in de grote zaal van de schouwburg van gemiddeld per voorstelling op het gebied van klassieke muziek en toneel van 60 procent;
- Opname van minimaal 40 procent ander toneel dan blijspelen (komedies).

De werkelijkheid is dat een schouwburgdirectie misschien helemaal niet denkt aan een evaluatiecriterium of er geen belang bij heeft of de handen vrij wil houden. En de subsidieverstrekker heeft wellicht ook geen aandacht geschonken aan de vaststelling van een maatstaf om naderhand het succes of het falen van een subsidie te meten.

We zien dat evaluatoren vaak een toevlucht moeten nemen tot een *next best*-maatstaf. Dat kan een prestatie maatstaf zijn bij gebrek aan zicht op een effect.

5 *Ideaal is een tijdige opdracht tot evaluatie.*

De evaluator wil graag een tijdige opdracht, en een evaluatiedesign kiezen met voor- en nameting(en), met experimentele en controlegroep en met een variatie naar beleidsimpuls en met zicht op effecten. Dan zal de evaluator ook nog graag nagaan of er sprake is van invloeden van buiten, om de oorzakelijkheidsanalyse (komt het effect doro beleid of een extern effect?) te verrichten.

6 *Ideaal is als de officiële doelstellingen ook de echte doelstellingen zijn.*

Het is gangbare praktijk om de officiële overheidsdoeleinden als evaluatiecriteria te nemen. Het beleid is immers democratisch vastgesteld. Er ligt een analyse aan ten grondslag, het beleid is qua doelstellingen en middelen doordacht. Adviesorganen als de Raad van State hebben er hun licht over laten schijnen. Dus ligt het voor de hand om deze doelstellingen ook serieus te nemen. Waarom dan andere criteria nemen?

Er behoeven geen andere criteria te worden genomen. Een orgaan als de Algemene Rekenkamer doet dat dan ook niet. Maar er zijn twee waarschuwingen van belang.

Ten eerste, gaat er soms een doelstelling achter een wet of andere maatregel schuil die, als je deze noemt, al geen kans meer maakt om gerealiseerd te worden. Zo was de WUB, de Wet Universitaire Bestuurshervorming, in de jaren zeventig van de vorige eeuw een wet die ook ertoe strekte om kritiek op het universitaire bestel en met name kritiek van studenten op hun gebrekkige medezeggenschap de wind uit de zeilen te nemen. Maar dit punt is *nooit officieel erkend*.

Wie evalueert, doet er goed aan om dit soort 'accenten' toch op te sporen. Dat kan door middel van analyse van de wetsgeschiedenis of interviews met bewindslieden en parlementariërs. Op *uitsluitend* gebruik van officiële doeleinden bij beleidsevaluaties bestaat dan ook kritiek (Bressers en Hoogerwerf, 1995: 23).

De tweede waarschuwing betreft dat bepaalde doelstellingen belangrijker zijn dan andere. Toen de wetsevaluatiecommissie die de Wet algemene bepalingen milieuhygiëne (Wabm), waarin ik zitting had, in de eerste helft van de jaren tachtig werd geïnstalleerd zei de toenmalige minister Leendert Ginjaar (minister van Volksgezondheid en Milieuhygiëne in het kabinet-Van Agt I) dat de commissie wel moest weten dat een bepaalde doelstelling *belangrijker* was, in zijn ogen, dan een andere.

7 *Ideaal is als de doelstellingen stabiel blijven.*

Een specifiek punt is ook de vraag of officiële doelstellingen stabiel blijven, dus in de loop van de tijd niet veranderen doordat bepaalde accenten verlegd worden en er doelstellingen bij komen of verdwijnen ten opzichte van de oorspronkelijke doeleinden. De wereld staat tenslotte niet stil en ook in politieke arena's gebeurt van alles.

Idealiter blijven doeleinden stabiel, dat is het makkelijkst voor evaluatoren. Evaluatoren doen er evenwel goed aan rekening te houden met wat in evaluatiekringen de '*dynamiek van doelstellend gedrag*' heet (Herweijer, 1981; Bressers en Hoogerwerf, 1995: 23; Bongers en Den

Hertog, 2007: 37). Doelstellingen kunnen gaan rollen. Dat wil zeggen, de ene doelstelling gaat na verloop van tijd belangrijker worden dan een andere, bij voorbeeld door veranderde maatschappelijke omstandigheden en/of de komst van een politiek bestuur van andere samenstelling. Misschien blijft een doelstelling wel bestaan maar er worden meer of minder middelen aan een doelstelling gekoppeld. Een evaluator kan zich dan de vraag stellen of dat een desastreus gevolg heeft gehad of helemaal niet.

8 *Ideaal is als de effecten hard zijn te maken.*

De evaluatoren onderzoeken en onderzoeken maar zijn de effecten ook echt zichtbaar te maken?

Ray Pawson (2006) is de man van *evidence-based policy*. Hij wil bereiken dat beleid zo onderzocht wordt dat je bij wijze van spreken ook weet dat als je een bepaald medicijn geeft de kans heel groot dat een gemiddelde patiënt, waarvoor het medicijn ook een oplossing zou moeten zijn, ook eerder beter wordt dan misschien langs natuurlijke weg al het geval zou zijn. Hij streeft met zijn *evidence-based policy* bewijs na alvorens een beleidsinterventie wordt gepleegd. Maar makkelijk is dat niet (Nelen, 2008). Waarom?

Ten eerste, bestaande kennis over een (voorgenomen) beleidsprogramma is meestal veelsoortig en gefragmenteerd.

Ten tweede, in onderzoek ontbreekt vaak een (quasi-) experimentele opzet.

Ten derde, als we al iets over effecten weten, welk onderdeel van het beleidsprogramma is daar dan verantwoordelijk voor?

De effecten van een programma zijn dus 'zelden' hard te maken, aldus Pawson & Klein Haarhuis (2005: 46).

De oplossing? Ik zie twee mogelijkheden voor evaluatoren.

1) *Informatiebundeling*. Probeer zoveel mogelijk informatie te verzamelen over processen waarlangs beleid zijn uitwerking heeft, die gebaseerd is op uiteenlopende bronnen van informatie. Kijk wat notulen, brieven, mails, sms-berichten, dossiers en andere documenten opleveren, alsmede observaties van gedrag, interviews met sleutelfiguren en objecten van beleid, en schriftelijke enquêtes. Filter overeenkomst en verschil uit. Analyseer eventuele tegenspraak in het materiaal (Pawson en Klein Haarhuis, 2005: 46).

2) *Early warning evaluation* (ex post). Is een beleid nog amper ingevoerd en kunnen effecten nog niet blijken dan kan toch geëvalueerd worden door te kijken of de maatregelen in de beleidsuitvoeringsfase nog sporen met de bedoelingen van de beslissers: 'een logische stappentoets'; ook wel consistentietoets genoemd. Deze evaluatie heet *early warning evaluation*. Met deze aanpak wordt gepoogd een keten van stappen te construeren vanaf de besluitvorming, de voorlichting, de invoering van beleid tot de prestaties. Zijn de punten in de keten consistent, zo dat er een lijn zichtbaar wordt die gaat richting gewenste effecten?

Het doel hiervan is natuurlijk om vroeg geattendeerd te worden op *wegdrijven* van de bedoelingen of doelstellingen, zodat bijstelling nog mogelijk is.

9 *Ideaal is als evaluatoren zich bewust zijn van de netwerkcontext.*

Er is nog een bijzonder probleem bij evaluatieonderzoek: wiens handelen wordt aan evaluatie onderworpen? Dat van een overheid of het handelen van meerdere netwerkpartners die van elkaar afhankelijk zijn omdat geen enkele overheid over alle taken, bevoegdheden en middelen beschikt?

Hier is achter te komen als je kijkt naar wie taken en bevoegdheden heeft op een bepaald beleidsterrein en wie over welke middelen beschikt. Het komt wel voor dat een evaluator niet in de gaten heeft dat één overheidsbestuur maar *beperkte* taken en bevoegdheden heeft en zich ook niet autonoom met iets kan bemoeien omdat andere overheidsactoren en actoren daarbuiten ook over taken, bevoegdheden en/of middelen beschikken. Dan zijn de actoren van elkaar afhankelijk. Ieder voor zich kan niet veel uitrichten, de ‘partners’ hebben elkaar nodig. Geen enkele van de actoren afzonderlijk heeft alle doorzettingsmacht. Geen enkele partij kan afdwingen wat die wil. Om tot actie te kunnen komen, moeten de actoren aan netwerkconstitutie doen (de goede partijen rond tafel krijgen) en moeten ze komen tot netwerkmanagement: een gezamenlijk perspectief ontwikkelen, meer concrete doelverbinding aangaan en komen tot middelenballing (Klijn en Koppenjan, 2001; Van den Heuvel, 2001; Klijn, 2005; Bekkers en Lips, 2008).

Neem de verdroging of vernatting van De Peel (Bekkers en Lips, 2008). Op dit vlak heeft geen enkele overheidsorganisatie of private partij alle doorzettingsmacht. Geen overheid heeft hier alle taken, bevoegdheden en middelen in handen. Organisaties zijn dus van elkaar afhankelijk. Om de verdroging of vernatting aan te pakken, hebben we onder andere te maken met enkele departementen, enkele provinciebesturen (Noord-Brabant en Limburg), een waterschap, gemeenten, grondeigenaren, pachters. Deze partijen dienen bij beleidsvorming te komen tot *gedeelde beeldvorming* over wat het probleem is en *wat de oplossingsrichting*, die uitmondt in *doelbundeling* en zelfs *doelverrijking* en *middelenbundeling*. Fixaties op bepaalde onhaalbare oplossingen moeten weg en ook blokkades moeten verdwijnen. Om het eens te worden zal men het eigen belang en de eigen ‘mental map’ moeten overstijgen.

Wat betekent dit voor evaluatie? Evaluatie in deze context zou eigenlijk als nuttig gezien moeten worden en gedeeld moeten worden door *alle* partijen. De evaluatie moet dan niet gericht zijn op zichzelf, op slechts een actor maar op de netwerksamenstelling, netwerkdoelstellingen en netwerkproductiviteit. De evaluator heeft dan dus gedeeltelijk andere maatstaven nodig.

Het komt veel voor dat de aanwezigheid van beleidsnetwerken vol afhankelijkheden niet gezien wordt door de opdrachtgevers van evaluatieonderzoek en evaluatoren. De evaluatoren komen dan met een criterium als slagvaardig bestuur en versterking van slagvaardig bestuur (‘spierballen’). Dat is niet verstandig. Slagvaardigheid is een criterium dat lijkt verbonden met een overheid die alle doorzettingsmacht bezit. In plaats daarvan zouden evaluatoren de netwerkproductiviteit moeten bekijken. Dus meer aandacht is gewenst voor ‘stroop’ en voor ‘schikken en plooiën’ dan voor de ‘spierballen’.

Dit misverstand doet zich soms voor bij de rol van provincies. Dan wordt de rol van het provinciebestuur gezien als een bestuur dat zijn doorzettingsmacht onvoldoende neemt terwijl miskend wordt dat de provincie sterk afhankelijk is van anderen (Korsten, 2011).

Dergelijke *netwerkgerichte evaluaties* komen voor maar ze zijn schaars. Hoe kan dat?

- 1) Een eerste factor is dat een overheidsbestuur soms een evaluatie over zichzelf en voor zichzelf wil omdat dit bestuur ook zelf de evaluatie financiert.
- 2) Een tweede verklaring ligt in het moeizame overleg om het eens te worden over een netwerkgerichte evaluatie. Partijen moeten het eens zien te worden over een evaluatieopdracht. Het gevaar van een ‘garbage can’ dreigt: iedereen wil iets anders maar niet of weinig betalen.

3) Een derde verklarende factor ligt in de sterke kracht van de suggestie dat deze ene overheid dan misschien wel afhankelijk lijkt maar wie wil dat toegeven. De reflex om toch vanuit zichzelf een evaluatie op te stellen is sterk.

Op bepaalde terreinen zijn netwerkgerichte evaluaties zelf nagenoeg onmogelijk. Toen kritisch gekeken moest worden naar de Herculesramp nabij Eindhoven (een vliegtuigongeluk met vele doden) verscheen een tiental rapporten. Elke betrokken organisatie liet een onderzoek instellen vanuit het eigen perspectief en ten behoeve van zicht op de eigen organisatie. Nuttig om zich te verantwoorden en te willen leren maar *'blame avoidance'* ligt ook op de loer (zie ook het werk van Hood). Iedere organisatie wil weten of ze wel of niet gefaald heeft en mogelijk ook de eigen stoep schoon vegen.

Tineke Abma (1996) is sterk voorstander van netwerkgerichte evaluaties waarin het perspectief van verschillende actoren aan bod komen. Ze noemt het *responsief evalueren*. Strengere evaluatoren noemen responsieve evaluatie echter geen evaluatie omdat het hen teveel doet denken aan het iedereen een bijdrage laten leveren aan beleidsvorming.

Conclusie: wet van de evaluatieve gebreken

Genoemde punten maken duidelijk dat er meestal *geen ideaal design voor onderzoek* mogelijk is. 'Dan is het een beetje behelpen'. Soms zal iedereen onder de bevolking of doelgroep voor beleid in aanmerking willen komen en is er dus geen groep te vinden die uitgesloten is van beleid en tot controlegroep 'gebombardeerd kan worden'. Dan is de keuze voor een experimentele groep en een controlegroep voor de evaluator helemaal niet te maken. Ook het werken met variatie in beleidsimpuls lukt vaak niet. En tenslotte, soms wordt een opdracht tot evaluatie pas verstrekt als het beleid al enige tijd loopt. Dan is het laat. Een nulmeting is dan onmogelijk.

Alleen al deze omstandigheden bewerkstelligen wat genoemd wordt *de wet van de evaluatieve gebreken*. De wet impliceert: er is altijd wel een manco aan evaluatieonderzoek (Leeuw, 2008: 26). In de praktijk kunnen er meer manco's zijn dan de genoemde.

8 HET TE VROEGE EVALUATIEMOMENT: EARLY WARNING EVALUATION ALS OPLOSSING

Stel, een beleidsprogramma in de vorm van een wet is door de volksvertegenwoordiging geaccepteerd (ook door de Eerste Kamer) en kan in het Staatsblad. De invoeringsdatum is nabij. Het departement laat een persbericht het licht zien. De organisatie van de beleidsuitvoering wordt opgepakt en er wordt een invoeringsprogramma opgetuigd als eerste onderdeel van de uitvoering. Ambtenaren (uitvoerders) gaan op cursus. Het eerste boekje over het nieuwe beleid van een juridisch onderlegd schrijver die vertrouwd is met de beleidssector ziet binnenkort het licht. En uitgeverij gaat aan het werk met een losbladige uitgave, cd roms, en wellicht ook met een discussieplatform voor een 'community' van betrokkenen bij en kenners van de wet.

Misschien heeft de oppositie het beleid gehoord en is er een motie aangenomen die de minister verplicht om binnen 'korte tijd' te komen tot een evaluatie van het beleidsprogramma. De minister heeft nog even getracht om de Tweede Kamer tegemoet te komen door, na een hint in die richting in het advies van de Raad van State, voor te stellen om in de wet een *jaarverslagartikel* op te nemen. Eens per twee jaar zou er een jaarverslag komen met bevindingen van monitoring van de wetsuitvoering. De Tweede Kamer nam daar echter geen genoegen mee. Men wilde iets zwaarders. Het werd *niet een horizonbepaling* (de wet houdt na vijf jaar op te bestaan en vooraf dan een wetsevaluatie) maar het werd een ad hoc toegezegde *evaluatie*. Wat dan vervolgens? Wachten tot de beleidseffecten zichtbaar zijn kan een jaar of drie-vier gaan duren. Zo lang wilde het parlement niet afwachten. De Kamerleden weten: als er geen druk op de ketel gezet wordt, komt er niks van.

Daarop besluit het betreffende departementaal sectorhoofd, gehoord de directeur-generaal, een maand of vijf na introductie van het programma een 'aanbesteding op uitnodiging' te starten om een evaluator te vinden. Begrotingruimte daarvoor was eerder al geschapen.

Wat moet de evaluator in een offerte al aanbieden? Voor een *ex ante evaluatie* is het moment strikt genomen te laat. Een *before-aftermeting* met experimentele groep en controlegroep is ook niet mogelijk; een nulmeting was niet voorzien. Maar de evaluator kan overwegen meerdere meetmomenten te kiezen, want de wet is immers al in uitvoering. Direct maar een meting? Maar wat dan?

Een mogelijkheid is de *early warning evaluation* (EWA). Dit is een concept waarmee ik zelf ervaring heb opgedaan nadat in de vorige eeuw de Innovatienota het licht zag (Gooren, Korsten en Prakke, 1983; Gooren en Korsten, 1983).

Wat een early warning evaluation is

Een early warning evaluatie heeft de volgende basiskennmerken (ook Gooren, 1985).

1. De early warning evaluation (EWA) is een *ex post evaluatie*.
2. Deze EWA vindt plaats op een moment dat het beleid *nog niet lang in uitvoering* is. EWA is dus aan de orde in een *vroege* fase van de beleidsimplementatie.
3. Er zijn in- en uitvoeringsmaatregelen genomen maar beleidseffecten zijn nog *niet of nauwelijks* zichtbaar. De evaluatie kan dus geen doelbereikingsanalyse zijn of effectiviteitsanalyse.

4. De evaluatie is gericht op de analyse van de opeenvolgende stappen van 'beleid tot effect'. Het is een soort *'ketenanalyse met beoordeling op consistentie van maatregelen en acties'* waarbij nagegaan wordt of de voorlichting, de uitwerking van doelen, middelenkeuze en andere elementen consistent zijn tot opzichte van de gedachte route van 'beleid tot effect'. Zodat daarmee *een voorspellende antwoord* gegeven kan worden op de vraag of de doelstellingen überhaupt later bereikt zullen *kunnen* worden. Sporen de al getroffen maatregelen met de veronderstellingen achter het beleid?
5. De evaluatie is bedoeld voor *beleidsaanpassing, -correctie of -uitwerking*. De beleidsmakers worden door een EWA vroeg gewaarschuwd als beleid dreigt weg te drijven van de oorspronkelijke intenties.

Wat de EWA niet biedt, is een fundamentele evaluatie van de vraag of het beleid überhaupt wel zinvol was en of de doelstellingen haalbaar zijn. Het is ook *geen doelstellingsvrije evaluatie* en *geen doelmatigheidsevaluatie*. Eerder is het een uitvoeringsevaluatie in een vroege fase gericht op de vraag of er een lijn zit in de uitwerking van beleid die de kans op het bereiken van de doelstellingen groot maakt.

De EWA is *bij uitstek* gebruiksggericht.

Het waarom van early warning evaluation

Het waarom van deze vorm van evaluatie is makkelijk te achterhalen en te formuleren.

1. Er is sprake van *onzekerheid* over de kwaliteit van de beleidsinhoud en/of uitvoering. De oppositie en anderen in het parlement hebben bedenkingen over het beleid maar zij hebben het beleid niet willen tegenhouden. De evaluatie is voor hen een 'aflaat', een 'reiniging van de ziel' waarna alles weer door kan gaan. Een aflaat is een lichte straf voor een lichte zonde.
2. De evaluatie is een soort *tweede kans* om nog te repareren in een stadium als de belangen niet echt gesteld zijn. Er is dan nog ontvankelijkheid voor verandering.

In het voorbeeld is niet aangegeven waar de bedenkingen ten aanzien van de wet betrekking op hebben. Wellicht is een vraag of de uitvoerende eenheden in het land wel tijdig 'in de startblokken staan'. En mogelijk is sprake van vrees voor gebrekkige handhaving of gebrekkige steun in het veld.

Voorwaarden

Tot de voorwaarden voor het succesvol verrichten van een EWA behoort het volgende.

- De evaluator moet allereerst een ervaren evaluator zijn met kennis van de organisatie van de uitvoering van beleid en dus van implementatievraagstukken.
- Ten tweede moet de evaluator wets- en veldkennis hebben.
- Ten derde moet de evaluator vermogen hebben om schakels in de uitvoering van deze specifieke wet te zien en te koppelen en overzicht te krijgen, iets waarvoor René Torenvlied (1996) in zijn dissertatie *'Beleid in uitvoering'* aandacht vroeg.
- Vermoedelijk is het onderzoek makkelijker uit te voeren als sprake is van een *conditioneel beleidsprogramma* dan van een *doelprogramma*. Een conditioneel

programma is een 'als... dan'- programma. Bij voorbeeld is te denken aan iemand die een identiteitsbewijs wil. Als iemand Nederlander is, ingeschreven is in gemeente X en voor de kaart betaalt (de condities), kan een identiteitskaart verstrekt worden. Stel even dat het zo werkt, dan is dit een kwestie van checken en verstrekken. Een evaluator kan bij een dergelijk uitgewerkt systeem een hele keten uittekenen. Bij doelprogramma's met veel beleidsvrijheid voor ambtenaren wordt dit veel lastiger.

Werkzaamheden

Hoe de evaluator te werk kan gaan?

1. *Een referentiebeeld opbouwen.* Het verrichten van deze EWA is mogelijk voor evaluatoren die het beleid en de beleidsgeschiedenis goed kennen en zeer op de hoogte zijn van de invoeringsmaatregelen. De evaluator moet kunnen uittekenen hoe die uitwerking en uitvoering logischerwijze zullen moeten verlopen. Zonder dat beeld kan het niet.
2. *De schakeling in de ketting opsporen, dus de logische stappen op een rij zetten.* Uit dat referentiekader moet afgeleid kunnen worden wat de stappen zijn die achtereenvolgens gezet zouden moeten worden. Het opsporen van het gewenste uitvoeringspatroon vereist enige durf van een evaluator. Maar deze kan een expertmeeting organiseren en interviews hebben met de beleidsontwerpers om dit *ideaaltypisch beeld of model* op te bouwen? Dat lijkt voor de hand te liggen. Op die manier kan achterhaald worden wat de schakels zijn, wat de cruciale momenten zijn in de uitvoering waarop het beleid naar de consistente uitvoering kan doorlopen dan wel naar 'de afgrond' kan gaan. Ook komen kritische onderdelen of fasen in beeld.
3. *De gegevensverzameling.* De evaluator gaat gegevens verzamelen. Hoe hij of zij hier moet opereren hangt grotendeels af van de wetsinhoud, kennis van het uitvoeringsprogramma en de verschillende actoren die een rol spelen. Het multi-actor perspectief kan aan de orde zijn. Kunnen bepaalde actoren elkaar tegenwerken. Welke belangen spelen? Hoe zijn de machtsverhoudingen? Dreigen perverse effecten?
4. *De vergelijking tussen gegevens en het referentiekader.* Hierbij gaat het om de verzamelde gegeven langs het ideaaltypisch patroon te leggen en te zien of sprake is van congruentie en consistentie.

Conclusie

De early warning evaluation is een interessante oplossing voor als een evaluatiemoment eigenlijk te vroeg is gekozen.

9 DE VEELHEID AAN MAATSTAVEN: DE NORMENBANK ALS OPLOSSING

Geen evaluatie zonder maatstaven. Dit onderwerp verdient dus nader aandacht. Ik ga in op de vraag welke maatstaven zoal in evaluatiestudies en daarbuiten voorkomen. Het resultaat van die exercitie zal u verrassen.

Maatstaven: verschillend in de beleidscyclus?

Welke maatstaven zijn zoal te kiezen zijn in een beleidsevaluatie hangt grotendeels af van het onderwerp van evaluatie. Bij een evaluatie van een totstandkoming van beleid zijn andere criteria aan de orde dan bij het nagaan van de beleidsimplementatie of beleidseffectiviteit. Ik werk dat kort uit.

Bij de *totstandkoming van beleid* is aan de orde wat eigenlijk de aanleiding tot overheidsactie is. Wat is precies het probleem of de uitdaging? Gaat het om een probleem op het vlak van energie, volkshuisvesting, arbeidsmarkt, winkelvoorzieningen, parkeren, migratie, milieu, integratie, onderwijs, straat- of uitgaanscriminaliteit, cohesie, cultuur, recreatie of sport, pedofilie, cybercrime, emancipatie? Hoe wordt het probleem gepercipieerd en gedefinieerd? Is dit iets voor overheidshandelen of moet de overheid hier buiten blijven? En als de overheid zich ermee inlaat, wat wordt dan het beleidsvoorstel? Het is mogelijk dat hier volkomen verschillend naar gekeken.

Neem de in 2012 nabij Den Helder aangespoelde bultrug. Topexperts meldden zich om te bezien wat nog te doen viel. Deze walvis was volgens sommigen ten dode opgeschreven; niet veel aan doen; laat de natuur zijn werk doen. Anderen meenden dat de walvis humaan behandeld moest worden. Ze gaven het dier ook een naam: Johanna. Daardoor kwam het dier als het ware dichterbij. Er kwam nog net geen condoleanceregister. De vraag was voor de sceptici van zoveel 'gedoe': wat heeft de reddingsoperatie gekost en wie betaald dat? En, de dierverdedigers (Kamerlid Dion Graus e.a.) vroegen zich af: moet er een protocol komen voor het omgaan met aangespoelde grote vissen? De Tweede Kamer besloot uiteindelijk niet tot specifieke beleidsmaatregelen. Daarmee was in de ogen van veel burgers een absurditeit, namelijk een protocol, voorkomen.

Bij het nagaan van wat *van de beleidsuitvoering terecht komt* is beleidseffectiviteit aan de orde en daarmee ook de doelbereiking.

Maar in het algemeen is te denken aan de volgende criteria bij beleidsevaluatie:

- responsiviteit: signaleert het overheidsbestuur de serieuze maatschappelijke voorkeuren en eisen en verwerkt het die?;
- doelgerichtheid van het beleid: welke doelstellingen en subdoelstellingen moeten worden gerealiseerd volgens het officieel democratisch vastgesteld document?;
- legitimiteit: zijn burgers betrokken bij de beleidsvorming?;
- proportionaliteit: wordt de lichte en goedkopere beleidsinstrumenten gekozen of zware en dure zware?;
- consistentie: is tegenspraak in het beleid zelf vermeden?;
- effectiviteit: worden gewenste doelstellingen bereikt?

Maatstavenvloed

Om welke criteria gaat het nu in beleidsevaluatie, als we de ontwikkeling in de tijd plaatsen en daarbij ook criteria van visitatiecommissies en uit jurisprudentie betrekken?

Figuur 2: Beoordelingscriteria in decentraal overheidshandelen, 1958-2008

<i>Criterion in overheidshandelen</i>	1958	1968	1978	1988	1998	2008
Rechtstatelijk bestuur	++	++	++	++	++	++
Rechtmatig bestuur	++	++	++	++	++	++
Integriteit van bestuur(ders)/voorbeeldrol	+	+	+	+	+	++
Autonomie van lokaal bestuur	+	++	+	+	+	++
Beginselen van behoorlijk bestuur	+	++	++	++	++	++
Betrouwbaar bestuur	++	++	+	+	+	o
Sober bestuur	++	++	o	o	o	o
Haalbaarheid van bestuurlijke aanpak	+	++	+	+	+	o
Toegankelijk bestuur	+	+	+	o	o	o
Openbaarheid van bestuur (besluiten)	o	o	+	o	o	+
Transparantie van handelen (benchmarks)	o	o	o	o	o	++
Initiatiefrijk bestuur	o	o	+	+	++	++
Planmatig bestuur	o	+	++	+	+	o
Probleemgericht bestuur	o	o	+	+	+	++
Responsief bestuur t.o.v. samenleving	o	o	+	+	+	++
Handhaafbaarheid/kwaliteit handhaving	o	o	o	+	+	++
Kwaliteit van dienstverlening	o	o	+	+	++	+++
Cohesiegericht bestuur	o	o	o	+	+	++
Doelgerichtheid van bestuur	o	o	+	+	+	++
Consistent beleid	o	o	+	+	+	+
Prioriteiten bepalen	+	+	+	+	+	++
Effectiviteit van beleid	o	o	o	+	+	++
Legitimiteit/steun/draagvlak van beleid	o	o	+	+	++	++
Uitvoerbaarheid van beleid	o	o	+	++	++	++
Verzelfstandiging/privatisering	o	o	o	+	+	o
Doelmatig bestuur	o	o	o	o	+	++
Bedrijfsmatig bestuur	o	o	o	++	++	+
Behoud privacy	+	+	o	o	o	o
Slagvaardig bestuur (i.c. doorpakken)	o	o	o	o	o	+
Interactief bestuur/inspraak	o	o	+	+	++	+
Beginselen van behoorlijke regelgeving	o	o	o	+	+	+
Eenvoudige regels/deregulering	o	o	o	+	+	+
Gelijkheid nastreven	o	+	++	+	+	o
Beginselen van behoorlijke behandeling	o	o	+	++	++	++
Kwaliteit rechtsbescherming	o	+	+	+	+	+
Klantgerichtheid van bestuur	o	o	+	++	++	++
Vraaggestuurd bestuur/maatwerk	o	o	o	o	o	++
Inspanningsverplichtend	+	+	o	o	o	o
Prestatiegerichtheid/resultaatgerichtheid	o	o	o	o	+	+++
Overzichtelijkheid van bestuur en beleid	o	o	o	o	o	+
Programmasturing/begroting	o	o	o	o	o	+
Proportionaliteit in handelen (ook politie)	o	o	o	o	+	+
Voorzorg/preventie	o	o	o	+	+	+++
Controle op dagelijks bestuur/evaluatie	o	o	o	+	+	++
Effectrapportages (milieu-; etc.)	o	o	o	+	+	+
Monitoring van voortgang	o	o	o	o	+	+++
Meer (horizontale) verantwoording	o	o	o	o	o	++
Administratieve lasten verlagen	o	o	o	+	+	+
Bestuurskrachtig bestuur	o	+	+	+	+	++
Strategische beleidskwaliteit	o	o	o	o	+	++
Gemeentelijke regiekwaliteit	o	o	o	o	+	++
Governance, netwerkproductiviteit	o	o	o	+	+	++
Kwaliteit van (keten)samenwerking	o	o	o	o	o	++
Lerende overheid/ zelfreinigend vermogen	o	o	o	o	+	+++
Integraal bestuur	o	o	o	o	o	++
Aantal criteria met minimaal één plus	12	16	25	35	42	46
Aantal criteria met minimaal twee plus	4	7	5	7	11	33

Legenda: o=afwezig; + = enigszins ++= sterk aanwezig; +++= zeer sterk

In de jaren zestig van de vorige eeuw was nog slechts sprake van een beperkt aantal criteria om beleid mee te beoordelen, zoals rechtstatelijkheid, rechtmatigheid, betrouwbaarheid van bestuur, soberheid, haalbaarheid. Denken in termen van effectiviteit, legitimiteit en doelmatigheid was nog nauwelijks aan de orde. Logisch want in de jaren zestig was de overheid nog betrekkelijk klein in termen van begrotingsomvang en aantallen ambtenaren en was spreken over 'beleid' nog nauwelijks aan de orde, laat staan 'en vogue'. Maar in de decennia daarna is er een ware *vloedgolf* aan nieuwe maatstaven of criteria gekomen, zoals bij voorbeeld maatstaven op het vlak van planmatig bestuur (de commissie-Vonhoff sprak van de carnavalsoptocht van planprocedures), beginselen van behoorlijke regelgeving (zie de dissertatie van Inge van der Vlies), beginselen van behoorlijke behandeling (ombudsmancriteria), slagvaardigheid, proportionaliteit, openbaarheid en tegenwoordig transparantie, netwerkproductiviteit, verantwoording, zelfreinigend vermogen (denk aan de beginselen voor goed bestuur, zoals opgesteld onder minister Ter Horst).

Werden in 1958 pakweg een stuk of tien criteria waargenomen, in 2010 waren dat er al meer dan 40: een verviervoudiging (Korsten, 2005b; 2010).

Die *maatstavenvloed* is het gevolg van disciplinaire studies van wetenschappers, parlementaire debatten en jurisprudentie en van visitatiecommissies die gemeenten, politiekorpsen, woningcorporaties, ziekenhuizen en omroepen doorlichtten (Korsten, 2005b, 2010). De maatstaven zijn de gewoonste zaak van de wereld geworden maar het overzicht ging bij evaluatoren, beleidsvoerders en velen daarbuiten ontbreken.

Figuur 3: Code voor goed bestuur

<p>1. <i>Openheid en integriteit</i> Het bestuur is open en integer en maakt duidelijk wat het daaronder verstaat. Het bestuur geeft in zijn gedrag het goede voorbeeld, zowel binnen de organisatie als daarbuiten.</p> <p>2. <i>Participatie</i> Het bestuur weet wat er leeft in de maatschappij en laat zien wat het daarmee doet.</p> <p>3. <i>Behoorlijke contacten met burgers</i> Het bestuur zorgt ervoor dat hijzelf en de organisatie zich behoorlijk gedragen in contacten met burgers.</p> <p>4. <i>Doelgerichtheid en doelmatigheid</i> Het bestuur maakt de doelen van de organisatie bekend en neemt de beslissingen en maatregelen die nodig zijn om de gestelde doelen te behalen.</p> <p>5. <i>Legitimiteit</i> Het bestuur neemt de beslissingen en maatregelen die het mag nemen en die in overeenstemming zijn met geldende wet- en regelgeving. De beslissingen zijn te rechtvaardigen.</p> <p>6. <i>Lerend en zelfreinigend vermogen</i> Het bestuur verbetert zijn prestaties en die van de organisatie, en richt de organisatie hier op in.</p> <p>7. <i>Verantwoording</i> Het bestuur is bereid zich regelmatig en ruimhartig jegens de omgeving te verantwoorden.</p> <p>Min. van BZK (2009)</p>

Twee gevolgen: normenbank en code

De maatstavenvloed heeft twee gevolgen gehad.

Ten eerste, de Algemene Rekenkamer, die ook te maken kreeg met beleidsevaluatie, ging rond 1990 en later 'een normenbank' aanleggen om te voorkomen dat men na verloop van tijd bijvoorbeeld in een derde of vierde onderzoek naar overheidssubsidies, van jaren later, andere criteria zou aanleggen dan eerder geschied was. En om de kennis te benutten van eerder onderzoek naar hetzelfde thema.

Ten tweede, vanwege de maatstavenvloed is getracht tot een korter abstract lijstje van essentiële maatstaven te komen: de 'code voor goed bestuur', die enkele jaren geleden is opgesteld onder verantwoordelijkheid van de bewindslieden Guusje ter Horst (PvdA) en Ank Bijleveld (CDA) van Binnenlandse Zaken en Koninkrijksrelaties. Figuur 3 geeft dit overzicht.

Tot de essentiële criteria voor beoordeling van overheidshandelen behoort de noodzaak om beleid te evalueren op effectiviteit. Immers, beleid dat niet beantwoordt aan wat een overheidsbestuur wil bereiken en ervan verwacht, is ongewenst. Onderdeel van de code zijn ook criteria als doelgerichtheid, interactiviteit met burgers en steun voor beleid, alsmede zelfreïnigend vermogen.

De 'code voor goed bestuur' zit helemaal op de golflengte van 'een lerende overheid' die poogt steeds beter te handelen en aan *verbetermanagement* doet (verg. ook Van der Knaap, 1997). Daan van Beek spreekt van gerichtheid op *een intelligente overheid* en voormalig gemeentesecretaris van Amsterdam Erik Gerritsen (2011) spreekt van de wenselijkheid van *een slimme overheid*.

Maatstaven in een netwerkcontext

Beoordelen is niet alleen een zaak van maatstaven kiezen maar ook van *geschikte maatstaven* kiezen al naar gelang sprake is van een netwerkcontext of niet. NUBL is een voorbeeld van een netwerkcontext. Wat betekent NUBL?

Casus

De Nadere Uitwerking Brabant/Limburg (NUBL) was een samenwerkingsproject van de provincies Noord-Brabant en Limburg en de ministeries van LNV, V&W en VROM. Het samenwerkingsproject NUBL, waarin de ROM-projecten De Peel en Midden-Brabant zijn opgenomen, is uitgewerkt in een startconvenant dat deze partijen in 1992 hebben getekend. Hierin geven partijen aan dat - om de bestaande milieukwaliteiten in stand te houden dan wel te verbeteren en de gewenste ruimtelijke ontwikkelingen mogelijk te maken - een Plan van Aanpak moet worden opgesteld. Tevens wordt afgesproken dat dit plan gericht is op uitvoering van projecten in het kader van de *geïntegreerde aanpak* voor het *ruimtelijk beleid en milieubeleid*. De uitvoeringsprojecten zullen in nauwe samenwerking met regionale belangenorganisaties, bestuurlijke organisaties, maatschappelijke groeperingen en het bedrijfsleven worden ontwikkeld. Het Plan van Aanpak NUBL is dus een uiting van gezamenlijk concreet aan de slag willen en gaan.

De reden voor de samenwerking ligt in de stagnatie van de regionale economische ontwikkeling en de cumulatie van problemen in een omvang die groter is dan veel elders in Nederland. NUBL was een kans om verbetering te brengen op het raakvlak van landbouw, economie, recreatie en toerisme, milieu, water, ruimtelijke ordening en natuur. Er was ruime financiering voor de projecten mogelijk. De plannen zouden in de periode tot en met 1997 van start moeten gaan.

NUBL had onder meer betrekking op De Peel. De Peel ligt op de grens van Brabant en Limburg, tussen de kernen Asten, Deurne, Horst en Panningen. Binnen dit gebied liggen plaatsen als Griendtsveen en Helenaveen. Dagblad De Limburger hierover begin 1997: 'De Streekcommissie Peelvenen organiseert twee voorlichtingsbijeenkomsten. De eerste op (...) 1997 in (...) Nederweert en de tweede op (...) in (...) te Helenaveen. (...). De voorzitter en enkele leden van de Streekcommissie zullen vertellen over het werk van de Streekcommissie en hoe zij de streek bij hun werk willen betrekken.'

'Medio 1996 heeft de Streekcommissie Peelvenen het heft in handen genomen om samenhang in de uitvoering van beleid te brengen. Het Peelvenenproject is met name bedoeld om de samenwerking tussen de overheden, doelgroepen en bewoners in de Peelregio te versterken. De leden van de verschillende belangengroepen hebben afgesproken samen met de streek de problemen aan te pakken. Het gaat hierbij om problemen op het gebied van natuur, landbouw, recreatie, water, milieu, landschap en economie. Inmiddels zijn door initiatieven uit de streek ongeveer 20 projecten in uitvoering genomen. De Streekcommissie wil u graag informeren over haar werkzaamheden, de stand van zaken van de projecten en hoe zij u in het vervolgtraject wil betrekken. Want de Streekcommissie wil graag met de streek in gesprek gaan over de toekomst van de Peelvenen.'

Het ROM-gebiedenbeleid bleek in 1998 alleen nog maar een bestuurlijk plansucces. Er was sprake van plannen maar niet van veel meer, zoals blijkt uit de ministeriële evaluatie en het boek 'De proef op de ROM'. NUBL bleek geen goed beleidsmanagement gehad te hebben. Het had beter gekund, aldus de hoogleraar P. Glasbergen in een 'audit' van de zelfevaluatie door het programmabureau NUBL. Netwerkmanagement is blijkbaar niet makkelijk. Maar hoe een provinciebestuur in den brede te beoordelen?

Voor het begrijpen van de kwaliteit van een provinciaal bestuur is het niet alleen nuttig maar ook noodzakelijk het wetenschappelijk werk over beleidsnetwerken te kennen. Dat dreigde bij twee recente visitaties van de provinciale bestuurskracht, die van de provincies Limburg en van Zeeland, even vergeten te worden ondanks de aanwezigheid van drie voormalige commissarissen der koningin in de commissies. Maar er trad na discussie in de visitatiecommissies en het beraad met de adviseurs tijdig een 'herstel in denken' op. Waar ging het om?

Er dreigde bij de beoordeling van elk provinciebestuur, haar organisatie en bestuurlijke opgaven en bestuurskracht een en hetzelfde beoordelingskader te ontstaan voor *alle* provinciale taken. Dat zou bij nader inzien een blunder geweest zijn. Er bestaan namelijk grofweg twee soorten provinciale taken of taakclusters: taken waarbij het provinciebestuur (vrijwel) alle doorzettingsmacht in eigen hand heeft en taken waarbij sprake is van afhankelijkheid van andere organisaties. Ik licht dit toe.

Doorzettingsmacht. Ten eerste zijn dat de bestuurlijke taken (opdrachten en andere opgaven) waarbij een provinciebestuur zelf over vele aan die taken gerelateerde bevoegdheden beschikt en over middelen beschikt. In dat geval, dus als 'gedeputeerde staten aan alle knoppen zitten', is het provinciebestuur nauwelijks afhankelijk van andere actoren in een beleidsnetwerk en kan het zelf in principe veel macht doorzetten in de vorm van besluiten en beleidsrealisaties. Denk aan het onderhoud van provinciewegen. Wie veel taken, bevoegdheden en middelen in eigen handen heeft, kan beoordeeld worden in termen van doeleinden, voortgang maken en dus beoordeeld worden in termen van slagvaardigheid. Van bestuurlijke drukte heeft een dergelijk bestuur in dat geval geen last. Netwerkvorming is hier *niet* aan de orde want er *is* geen

(beleids)netwerk. Regiekwaliteit is dit geval een onzinnig criterium want het provinciebestuur is hier zowel sturingsactor als belangrijkste of enige uitvoerder van provinciaal beleid.

Gespreide macht. Daarnaast komt het veelvuldig voor dat een provinciebestuur beschikt over een zekere wettelijke taak maar juist nauwelijks over bevoegdheden en middelen, zeker niet over alle bevoegdheden. Dat is het tweede taakcluster. Gebiedsgericht beleid behoort hiertoe. Dan is een provinciebestuur als het iets wil bereiken juist buitengewoon afhankelijk van andere actoren en moet samenwerking en onderhandeling verder brengen. In die tweede context van taakvervulling in een netwerk van mede-overheden beschikt een provinciebestuur dus *niet alleen* over doorzettingsmacht. Een provinciebestuur krijgt hier alléén niets voor elkaar. Als provinciebestuur heb je de *andere actoren nodig* want ze beschikken - positief gezien - over middelen ('resources') die nodig zijn voor het realiseren van een gemeenschappelijk doel en negatief gezien over rem- of vetomacht ('blokkademacht'). Bij dit twee taakcluster moet je beoordelen:

- op adequaatheid van de netwerkconstitutie ('alle partners aan tafel?'),
- op netwerk- of ketenproductiviteit (sporen doelstellingen van van elkaar afhankelijke partners?; is sprake van schakeling tussen organisaties of elkaar hinderen?; kunnen financiële middelen gebundeld worden?)
- en op regie (o.a. initiatiefname; afstemming).

In netwerkcontexten is slagvaardigheid van een actor alleen, een provinciebestuur, op zich iets dat als 'een tang op een varken slaat'. In een netwerkcontext is slagvaardigheid van een provinciebestuur immers een niets zeggend iets omdat het gaat om *de slagvaardigheid van het totaal* aan 'netwerkpartners', die van elkaar afhankelijk zijn en elkaar dus nodig hebben. Het geheel moet slagvaardig worden en het moet niet zo zijn dat sprake is van 'free riders' (meeliften zonder inspanningen en een aandeel te leveren). Een netwerk komt in actie als partners zich willen verdiepen in de perceptie van de ander van een probleem-oplossingsrichting, en als er een klimaat ontstaat om de bestaande belangenconfiguraties te overstijgen. Een 'dialogo van doven' moet worden voorkomen. Het gaat hier om het leerstuk van procesmanagement (De Bruijn, Ten Heuvelhof en In 't Veld, 2002).

Bij het tweetal taakclusters gaat het dus om *gedeeltelijk verschillende criteria ter beoordeling*. Uiteindelijk is wel gemeenschappelijk of het beleid democratisch gefundeerd is, rechtvaardig is, rechtmatig, doelmatig en legitiem maar bepaalde criteria zijn gerelateerd aan een taakcluster, zoals met deze tweedeling duidelijk is gemaakt.

Ik zal de betekenis van gedifferentieerde criteriatoepassing verder toelichten aan de hand van eigen ervaringen met het visiteren van twee provinciebesturen op 'bestuurskracht', die van Limburg en die van Zeeland.

Bestuurskrachtmeting in Limburg en Zeeland

We maken nu een sprongetje in de tijd en naar een andere context, en bezien of netwerkdenken voor de visitatie van provinciebesturen van belang is.

Enkele jaren terug werd ik gevraagd om lid te worden van een visitatiecommissie die de bestuurskracht van de provinciebestuur Limburg tegen het licht moest houden. Ik zou dat werk aanvatten met een minister van staat en voormalige minister, CdK (van Noord-Holland) en burgemeester, te weten Jos van Kemenade en met als derde visitatiecommissielid een oud-commissaris der koningin (van Overijssel) en voormalig directeur-generaal openbaar bestuur van het ministerie van BZK Jan Hendriks. Dit ervaren trio liet zich bijstaan door een gerenommeerd adviesbureau, Berenschot. De commissie zou de bestuurskracht bekijken door uit te gaan van een selectie van bestuurlijke opgaven van het provinciebestuur en daarvan te

bezien wat er in de uitvoering van terecht kwam. Daaruit zouden dan wel enkele wijze lessen te trekken zijn die het provinciebestuur kon benutten voor de toekomst. Het visitatietrio kwam met de fractieleiders en gedeputeerde staten overeen dat een aantal geselecteerde opgaven geschikt was. Deze opgaven, die ambities bevatten, gaven de essentie van het provinciebestuur weer, omvatten meer dan de helft van het provinciebudget en werden ook anderszins als een 'redelijke' afspiegeling gezien van de provinciale taken (en daarmee verbonden bevoegdheden). De aanpak van de bestuurskrachtmeting kon zo op steun van alle politieke kanten rekenen en was daarmee onomstreden.

Beoordelingscriteria

De visitatiecommissie had zich in haar planopzet nadrukkelijk ingelaten met de beoordelingscriteria van het provinciebestuurlijk handelen (Berndsen, Fraanje, Korsten en Kort, 2007). Spiegelen is nuttig maar je moet wel brillen hebben om te kijken en normen ter beoordeling van wat je ziet. Waarop het provinciebestuur te beoordelen? De bestuurskunde doet ons hier een handreiking. Vanuit het perspectief van doelgerichtheid, systeemlegitimititeit en specifieke beleidssteun, doeltreffendheid, doelmatigheid, rechtmatigheid, slagkracht, vermijding van bestuurlijke drukte? Deze keuze van criteria moest recht doen aan het provinciebestuur, dus aan het middenbestuur. Dat betekende dat de keuze van dergelijke criteria niet voldoende was. Het provinciebestuur moet als *middenbestuur* immers ook op diverse terreinen coördineren en afstemmen. Denk maar eens aan de rol op het gebied van waterbeleid. Waterschappen voeren uit maar provinciebesturen sturen de waterschappen op hoofdlijnen, rekening houdend met het nationale beleid.

De visitatiecommissie formuleerde met de adviseurs een aantal criteria ter beoordeling van de bestuurskracht. So far, so good! Verliep het proces van beoordeling in de uitvoering van de visitatie ook soepel?

Een van de meest opvallende punten uit het visitatieproces was, achteraf, dat er bij de formulering van provinciale taken door de commissie aanvankelijk geen onderscheid was gemaakt in taken waarvoor de provincie zelf volledig verantwoordelijk is en waarbij dit bestuur ook over alle bevoegdheden en middelen beschikt, versus een situatie waarbij taken, bevoegdheden en middelen die gespreid waren over actoren, zoals veelal bij gebiedsgericht beleid het geval was en is. Er was door de visitatiecommissie en de ondersteuners dus *geen* rekening gehouden met het feit dat een provinciebestuur ten aanzien van *bepaalde* opgaven over veel doorzettingsmacht beschikte en over tal van andere in het geheel *niet* omdat het afhankelijk was. Dat betekende dat het trio visitatoren zich moest bezinnen op de beoordelingscriteria bij beleid waarbij het provinciebestuur beschikte over veel doorzettingsmacht versus beleid dat in een netwerkcontext tot stand komt en/of wordt uitgevoerd en het provinciebestuur dus over weinig doorzettingsmacht beschikt.

Waarom niet gezien?

Hoe kon het gebeuren dat dit onderscheid aanvankelijk niet gezien werd terwijl toch twee van de drie visitatoren nota bene zelf provinciebestuurder waren geweest?

Dat kwam omdat de ondersteunende adviseurs sterk op de private sector waren gericht (geweest) en omdat er geen actuele, toegankelijke en dus handzame catalogus bestaat van provincietaken die in netwerkvorm worden opgepakt. De ondersteunende adviseurs veronderstelden dat criteria als initiatief en slagvaardigheid ongedifferentieerd naar taken en bevoegdheden volledig van toepassing zou kunnen zijn op alles wat een provinciebestuur doet. Dat was en is echter een misverstand. Een provinciebestuur heeft op diverse beleidsterreinen helemaal niet de mogelijkheid om een proces naar haar hand te zetten. Ze is als middenbestuur slechts mede-overheid. Een provinciebestuur heeft bijna geen enkel direct

contact met burgers. Haar partners, subsidieontvangers of klanten zijn tussengeschakelde organisaties.

Afhankelijkheid

Overheidsorganisaties kunnen in het algemeen een wettelijk voorgeschreven taak hebben en ook over de (beslis)bevoegdheden en middelen beschikken om opgaven op een taakgebied te realiseren. Als dat het geval is, heeft een overheidsorganisatie doorzettingsmacht, tenminste als er voldoende steun is voor besluiten en geen partij besluiten dwarsboomt. Op tal van gebieden heeft een overheidsorganisatie wel een wettelijke taak maar slechts beperkte bevoegdheden en middelen. In die situatie is sprake van afhankelijkheden van andere overheidsorganisaties en actoren daarbuiten. Dan zijn interbestuurlijke relaties aan de orde.

Een bekend voorbeeld daarvan is de wenselijke vernatting van De Peel en het tegengaan van verdroging. Het bevorderen van vernatting is alleen mogelijk als van elkaar afhankelijke partijen samenwerken. In geval van De Peel ging het om enkele ministeries, enkele provinciebesturen, waterschappen, gemeenten en grondeigenaren. Samenwerking is nodig omdat geen partij zijn wil (of macht) kan opleggen aan een andere. Geen enkele partij kan de vernatting alleen realiseren. Er is geen bestuurlijke hiërarchie gewenst maar wat men in Duitsland noemt: 'Politikverflechtung'. Als andere 'partijen' niet gezamenlijk meedoen, komt er niks van de grond.

In een situatie van afhankelijkheid gelden niet zonder meer dezelfde beoordelingscriteria als in een situatie waarin voor een provinciebestuur geen afhankelijkheid bestaat. Indien organisaties van elkaar afhankelijk zijn, is het gewenst om te komen tot

- netwerkconstitutie (goede partijen aan tafel) en
- netwerkproductiviteit.

Die productiviteit kan intentioneel blijken uit goede bedoelingen maar belangrijker is dat sprake is van (een zekere) doelcongruentie en ook een middelenbundeling. *Doelcongruentie* is nodig om te vermijden dat de strijd voortduurt over waarheen een gebied zich ontwikkelt. Middelenvervlechting is nodig omdat geen enkele partij genoeg middelen heeft om te realiseren wat nodig is. *Cofinanciering* dient zich als een mogelijkheid aan.

Regie

In een beleidsnetwerk zitten niet zo maar de goede partijen aan tafel en ze zijn niet 'als vanzelf productief'. Het is nodig dat er regie ontstaat door een procesarchitect en dat blokkades vermeden worden en doorbraken ontstaan. Een provinciebestuur kan een rol in deze vervullen als ze een relatief machtige positie vervuld, hetzij qua bevoegdheden, hetzij financiële middelen, hetzij beide. Provinciebestuurders die zelf krachtdadig zijn in de ene context waarin het provinciebestuur doorzettingsmacht heeft, moeten in een netwerkcontext juist behoedzaam opereren.

Dat is wat de visitatoren in Limburg geleerd hebben. Zij en de adviseurs moesten niet zo maar overall slagvaardigheid, doelgerichtheid en beleidseffectiviteit van handelen eisen (criteria die makkelijk zijn te hanteren als een overheid de volledige doorzettingsmacht heeft) maar daar waar een netwerkcontext aan de orde is ook de netwerkconstitutie, de netwerkproductiviteit, de eventuele fixaties of blokkades en de regie bezien.

Deze belangrijke les van onderscheid maken naar taakclusters is ook geleerd in het proces van visitatie van de provincie Zeeland. Deze in 2010 uitgevoerde visitatiecommissie bestond uit oud-gedeputeerde van Groningen Marc Calon, oud-CdK van Noord-Brabant Hanja Maij-Weggen, voormalig CZ-voorzitter Mike Leers en hoogleraar Arno Korsten. Ook tijdens deze visitatie moest bij nader inzien het onderscheid in taaksoorten gemaakt worden.

Figuur 4: Matrix van provinciale taken

		Geheel van bevoegdheden en middelen van het provinciebestuur	
		Veel	Weinig of geen
Opgaven van het provinciebestuur	Wettelijke opgaven	a. Weinig afhankelijkheid van andere overheidsactores: grote doorzettingsmacht voor provinciebestuur	c. Veel afhankelijkheid: weinig doorzettingsmacht van provinciebestuur; netwerkproductiviteit gewenst
	Autonome opgaven	b. Weinig afhankelijkheid van andere actores. Middelen als lokaas.	d. Veel afhankelijkheid als gevolg van beperkte bevoegdheden en/of middelen: weinig doorzettingsmacht. Netwerkconstitutie en netwerkproductiviteit gewenst

Figuur 5 verschaft daarbij nog aanvullend inzicht doordat een onderscheid gemaakt wordt in wettelijke en autonome taken. Bij autonome taken die een provinciebestuur oppakt, kan ook weer sprake zijn van een context van veel afhankelijkheid van andere organisaties of weinig.

Figuur 5: Voorbeelden van enige taken van het provinciebestuur van Zeeland vlg. B&A Groep

		Geheel van bevoegdheden en middelen van het provinciebestuur	
		Veel	Weinig of geen
Opgaven van het provinciebestuur	Wettelijke opgaven	a. Weinig afhankelijkheid van andere overheidsactores: grote doorzettingsmacht voor provinciebestuur. Voorbeelden: Ecologische hoofdstructuur; steunfunctie Wmo voor gemeenten; vergunningverlening	c. Veel afhankelijkheid: weinig doorzettingsmacht van provinciebestuur; netwerkproductiviteit gewenst. Voorbeelden: nieuwe Wet ruimtelijke ordening.
	Autonome opgaven	b. Weinig afhankelijkheid van andere actores. Middelen als lokaas. Voorbeelden: Instandhouding van mbo, hbo en wo in Zeeland (OAZ)	d. Veel afhankelijkheid als gevolg van beperkte bevoegdheden en/of middelen: weinig doorzettingsmacht. Netwerkconstitutie en netwerkproductiviteit gewenst. Voorbeelden: provinciaal toeristisch beleid; havenontwikkeling; waterveiligheid; verkeersveiligheid.

Reflectie: nieuwe criteria

Wie als visitatiecommissie de bestuurskracht van een provinciebestuur beoordeelt, uitgaande van taakgerelateerde opgaven en ambities en de uitvoering daarvan, zal altijd beoordelingscriteria hanteren. De keuze daarvan is nog niet zo eenvoudig omdat historisch gezien sprake is geweest van een grote toename van criteria ter beoordeling van overheidshandelen (Korsten, 2010). Het gaat al lang niet meer alleen om bekende referentiecriteriën (wenselijkheden, eisen) als rechtmatigheid, doelgerichtheid, doeltreffendheid, doelmatigheid, behoorlijk bestuur, legitimiteit van overheidshandelen, integriteit. De overheid is zich namelijk in *toenemende* mate gaan inlaten met andere besturen en de samenleving en delen daarvan, en ze doet dat op *tal van manieren*. Dat leidde tot *nieuwe criteria* om beleid te realiseren zoals bij voorbeeld het vermijden van bestuurlijke drukte (filevorming houdt immers op), ‘regiekwaliteit’ (bij elkaar brengen), proportioneel handelen (lichte middelen gebruiken waar dat kan), transparantie. De historische groei in criteria verhindert niet om een lijst van criteria op te stellen voor het beoordelen van een middenbestuur (Berndsen, Fraanje, Korsten en Kort, 2007).

Kan men deze criteria nu onverkort toepassen op alle taken van een provinciebestuur: op taken die wel of niet in een netwerkcontext worden 'opgepakt' en uitgevoerd? Dat was de kernvraag van dit betoog. Het antwoord luidt negatief. Er bestaan namelijk taken waarover een provinciebestuur (nagenoeg) alleen kan besluiten en taken waarbij dat niet het geval is. Besluit een provinciebestuur alleen dan is beoordeling eenvoudig. Immers, alle taken, bevoegdheden en middelen komen dan uit het provinciebestuur zelf. Dan is 'bestuurlijke drukte' géén issue en dus geen beoordelingspunt want 'je doet zelf wat je besluit'. Dan opereert een provinciebestuur sterk hiërarchisch. Dan is netwerkproductiviteit *geen* criterium want er bestaat geen netwerk op dat gebied. In geval een provinciebestuur daarentegen niet alleen kan besluiten – en dat komt vaak voor bij een middenbestuur, zeker bij ruimtelijke ordening en gebiedsgerichte issues - is sprake van een *verstrengeling* tussen van elkaar afhankelijke overheden. Denk maar eens aan de ontwikkeling van een bedrijventerrein, dat moet passen in de provinciale en gemeentelijke ruimtelijke ordening en grondpolitiek. Denk ook aan het bestrijden van verdroging in een beekachtig, moerassig of drassig gebied. Dan is het nodig te kijken of de van elkaar afhankelijke overheden – met elk eigen taak, bevoegdheden en middelen - elkaar ontmoeten en overleggen om het eens te worden over doeleinden en over het in een pot doen van middelen (doelcongruentie en cofinanciering). En dan is inzet van blokkademacht ongewenst. Regie is hier juist wel een criterium als een provinciebestuur voor de taakaanpak afhankelijk is van andere actores (zoals waterschappen, gemeenten, samenwerkingsverbanden, grondeigenaren, etc.) die ook over eigen verantwoordelijkheden en daarmee eigen taken en bevoegdheden beschikken en middelen kunnen inzetten. Het provinciebestuur kan de regierol vervullen maar die rol kan ook bij andere overheden liggen, wat mede afhankelijk is van wettelijke kaders en/of (financiële) macht en/of afspraken.

Oude reflex

Opmerkelijk nu is dat in de praktijk de neiging tot een oude reflex bestaat bij visitatiecommissies en ondersteunende adviseurs, namelijk om criteria die horen bij een zelf besluitende en uitvoerende provincie ook los te laten op contexten waarin een provincie juist maar weinig te zeggen heeft, namelijk in een netwerkcontext (van van elkaar afhankelijke overheden). Met andere woorden, van een provinciebestuur wordt in die oude reflex verwacht dat ze hiërarchisch optreedt - denk aan woorden als: stevig, stoer, beslissend, slagvaardig - in een netwerkcontext. Welnu, dat kan niet. Dat leidt tot weinig of niets. Als een provinciebestuur in een netwerk opereert van van elkaar afhankelijke overheden dan moeten provinciebestuurders partners bij elkaar brengen, bestaande tegenstellingen overstijgen in een nieuw concept met een focus en verbinden, verzoenen, verknopen. Daarvoor kan een provinciebestuur bij voorbeeld ook met de geldbuidel rammelen. De wettelijk vastgelegde 'afhankelijkheid' - lees het vervlochten bestuur - dwingt daartoe. Daarmee betreden we het gebied van 'het faciliteren van interactieprocessen', van 'co-creatie' en van 'vitaliseren van gestagneerde organiseerprocessen', allemaal onderdelen uit het leerstuk van netwerkmanagement (Glasbergen, 1989; Van Eeten en Termeer, 1996; Van Eeten, 1999; De Bruijn, Ten Heuvelhof en In 't Veld, 2002; Termeer, 2006).

De ervaringen met visitatieprocessen leert dat goede provinciebestuurders in de gaten hebben wat het ene taakgebied en het andere vereist. Zelfs krachtige bestuurders met visie – zoals de Limburgse gedeputeerden Vestjens (VVD) en Driessen (CDA) - bleken in een netwerkcontext behoedzaam, faciliterend en interacterend te opereren terwijl ze juist de leiding nemen waar het provinciebestuur over een wettelijke taak en juist veel eigen bevoegdheden en middelen beschikt.

DEEL III: DE KEUZE VAN EEN EVALUATIE-OPZET EN DE EVALUATIELADDER

10 ER IS MEER DAN DE KLASSIEKE EVALUATIE: TYPEN BELEIDSEVALUATIE OP EEN RIJ

Eerst weer even terug naar de klassieke beleidsevaluatie. Vandaar bouw ik het verhaal op om zichtbaar te maken dat er meer typen beleidsevaluatie bestaan en er keuzen te maken zijn in concrete gevallen van onderzoek.

Het *effectiviteitonderzoek* wordt ex post uitgevoerd. Kern daarvan is, aldus Carol Weiss (1972, 1976) de *doelbereikingsanalyse* ('goal achievement approach') (Bekkers, 2007: 299). Daarin wordt een antwoord gegeven op de vraag: bereik ik met beleid wat ik wil bereiken? Maar de effectiviteitanalyse kent een tweede onderdeel: bereik ik de doelstellingen juist door het beleid zelf of door andere factoren of omstandigheden? Dat noemen we wel de *oorzakelijkheidsanalyse*.

Bij dergelijk onderzoek kan men veronderstellingen naar boven halen en bezien of ze juist blijken.

Bij voorbeeld: er wordt een verkeersveiligheidsbeleid opgetuigd. Doel is: minder slachtoffers door verkeersincidenten in wijken. Middelen: verkeersdrempels en andere verkeerssnelheidremmers. Resultaat: het aantal incidenten en ongelukken met letselschade neemt in twee jaar in stad X af in wijken met drempels versus wijken waar geen maatregelen zijn getroffen. Komt dat door de aanleg van drempels en andere remmers? Stel het blijkt ook dat de politie auto's meer is gaan controleren op rijnsnelheid en stel dat die controles meer in de wijken met minder slachtoffers plaatsvonden, zonder dat van bewuste politiek van de politie sprake was. Dan is de beleidsmix nuttig, inclusief de controles. Misschien moet na de evaluatie die dit aan het licht brengt het beleid worden aangepast met blijvend meer controles of de plaatsing van snelheidattenderingspalen, vanuit *de veronderstelling* dat die eenzelfde werking op verkeersdeelnemers hebben als politiecontroles. Immers snelheidscontroles zullen niet structureel in de betreffende wijken plaatsvinden.

Direct is nu ook duidelijk dat de gebruiksgerichte evaluatie niet voor de 'Bühne' plaatsvindt maar ook van nut is voor de bezinning op beleid. Dat kon in geval van de beoordeling van beleid alleen door een intelligente onderzoeksopzet met een interventiegebied en wijken die niet (frequent) aan de prikkel met verkeerscontroles blootgesteld waren.

Met het noemen van de klassieke evaluatie die uitgaat van doelstellingen of breder beleidsveronderstellingen is, is het klassieke type evaluatie genoemd. Een variant is de '*goal free evaluation*'. Wat houdt die in?

Doelstellingsvrije beleidsevaluatie ex post

Evaluatie van beleid kan een effectiviteitsanalyse impliceren, die vertrekt vanuit achterliggende doelstellingen en veronderstellingen maar ook een doelstellingsvrije evaluatie (goal free evaluation) zijn. Wie op zoek gaat naar neveneffecten van beleid, dus naar effecten die wel optreden maar nooit bedoeld waren, moet niet uitgaan van de officiële doelstellingen maar moet juist op zoek gaan naar bezwaren die vóór de besluitvorming over het beleid of direct daarna zijn geuit. En de evaluator kan 'in het veld' vragen naar veronderstelde of al gebleken neveneffecten. Critici en anderen kunnen aanwijzingen geven waar die neveneffecten te vinden zijn.

Indelingen

Er bestaan in werkelijkheid echter meer typen evaluatie. Het is als met Eskimo's. Eskimo's kunnen veel soorten sneeuw onderscheiden. Wie geen Eskimo is zal misschien alleen een onderscheid kunnen maken tussen natte en droge sneeuw, poeder- en korrelsneeuw. Zo is het ook met evaluatie. Kenners kunnen veel soorten onderscheiden en niet-kenners niet.

Ik geef enkele onderscheidingen en noem de volgende typen:

1. *Verplichte evaluatie* of *vrijwillige evaluatie*: evaluatie als gevolg van een beleid met een eindtermijn ('sunset legislation'), een evaluatiebepaling in een wet of verordening, of ad hoc te starten.
2. Een onderscheid naar evaluatie vóór de besluitvorming over een beleidsprogramma: *ex ante evaluatie* versus *ex post evaluatie*. De klassieke aanpak van effectiviteitsanalyse is een *ex post*-aanpak;
3. Een evaluatiemoment snel na besluitvorming, dat wil zeggen als er nog wat aan beleid te veranderen is: *early warning evaluatie*;
4. Evaluatie in de vorm van alleen terugkijken of ook vooruit kijken: *retrospectieve* en *prospectieve evaluatie*. De meeste beleidsevaluaties op rijksniveau zijn terugblikkende evaluaties. Die hebben het nadeel dat ze over het algemeen weinig handreikingen geven voor de toekomst. Daarom wordt wel gezegd: kom met een prospectieve evaluatie. Dan is de bedoeling om terug te blikken op de manier waarop beleid is toegepast en zich ontwikkeld heeft maar ook op zoek te gaan naar de doorontwikkeling en de mogelijkheid van toepassing van 'best practices' die elders voorkomen;
5. Evaluatie van een *beleidsprogramma* of evaluatie van *een organisatie* met een geheel van beleidsprogramma's'. Deze gaan vaak onder de vlag van een *visitatie* van gemeenten, ziekenhuizen, politiekorpsen, woningcorporaties, een omroep;

Ik zal nog eens aangeven wat de *ex ante* en *ex post*-aanpak is om langs die weg enkele andere typen te verduidelijken. Het stil staan bij de klassieke aanpak is tevens bedoeld om deze aanpak er bij de lezer in te hameren als een ankerpunt, als een punt van waaruit steeds te vertrekken is.

Toelichting: ex ante en ex post evaluaties

Ex ante evaluaties impliceren dat *voorafgaand* aan besluitvorming over een beleid nagegaan wordt wat gewenste of ongewenste gevolgen zijn van een of meer beleidsalternatieven. Bij *ex post* evaluaties gaat het om de beoordeling van beleid *nadat* het beleid is vastgesteld en volop in uitvoering is. De meeste studies die je tegenkomt zijn *ex post* studies. Het oogmerk 'verantwoording' brengt vaak al *ex post* evaluaties met zich. Maar ook wie beleid wil ombuigen laat vaak eerst een evaluatie verrichten. Immers, er moeten argumenten op tafel komen waarom tot ombuigen wordt overgegaan en welke kant die opmoet.

Ex ante evaluatie is evenwel ook belangrijk. Ik ben er altijd een warm voorstander van geweest wat onder meer bleek uit beschouwingen over *uitvoeringsgericht ontwerpen van beleid* (Elmore spreekt van 'forward & backward mapping').

Een voorbeeld: Stel er zijn alternatieven voor een beleid om de achterstand van autochtone jongeren in het Friese onderwijs te corrigeren. De volgende alternatieve beleidsroutes zijn in de fase van beleidsvoorbereiding genoemd:

- a) een programma mogelijk maken dat gericht is op meer *individuele begeleiding* van jongeren ('remedial teaching') in het Friese basisonderwijs: extra gelden voor het aantrekken van begeleidende leerkrachten;
- b) een programma gerichte op *klassenverkleining* in alle leerjaren van het Friese basisonderwijs, waardoor er ook meer aandacht van onderwijskrachten naar leerlingen kan uitgaan;
- c) een programma om op vrijwillige basis tegen betaling van een zekere bijdrage door ouders gedurende drie dagen per week driekwartier *extra scholing* te geven in alle leerjaren van het Friese basisonderwijs;
- d) een combinatie van a en b.

Stel dat dit de alternatieven zijn waar na allerlei afwegingen de voorkeur naar kan uitgaan maar er moet nog wel een keuze gemaakt worden. Dan kan een ex ante evaluatie gericht zijn op *het inschatten* van de effecten per alternatief en een vergelijking daarvan. Alternatief a en b kan duurder zijn dan c. Stel. Misschien dat een tweetal panelbijeenkomsten van onderwijskundigen en onderwijssociologen, een Nederlands panel en een internationaal panel, uitwijst dat de maatregelen a en b het beste samen kunnen gaan maar slechts een beperkt effect sorteren omdat de genetische factor veel meer de onderwijsprestaties in alle leerjaren van het basisonderwijs verklaart. Stel. En, stel, alternatief c stuit op weerstand bij Friese burgers en leerkrachten die niet willen worden opgezadeld met inning. Dan zou een uitkomst kunnen zijn dat een bewindspersoon geen enkele van deze maatregelen doorzet.

Ex ante evaluatie kan zo gezien ertoe bijdragen *om helemaal niet te beginnen aan beleid dat onvoldoende effect sorteert. Of om nog eens goed na te denken over de kosten en baten van beleid.*

Ex ante evaluatie maakt dus kritisch over bedoelde effecten (zie Hellendoorn, 2001; Hanemaayer, 2012).

Ex ante evaluatie is vooral nuttig bij die – doorgaans controversiële - beleidsissues waarvan veel onbekend is (wat zijn de oorzaken?) en waarover betrokkenen het maar moeilijk eens kunnen worden over de probleemanalyse en de oplossing?

Het was Ad Geelhoed, de topambtenaar die in de jaren tachtig van de vorige eeuw op nationaal niveau de dereguleringsoperatie leidde, die na het bekijken van tal van wetsontwerpen en wetten ook een krachtig pleitbezorger werd van *structureel ingebedde ex ante evaluaties*.

Horen we die geluiden meer recent nog? Ook Frans Leeuw (2008), directeur van het Wetenschappelijk Onderzoeks- en Documentatie Centrum (WODC) op het gebied van justitie en veiligheid, laat zich kritisch uit over de vele 'sprongen in het donker', die gemaakt worden, bij voorbeeld op het terrein van het beleidsmatig omgaan met justitiabelen. Sommig beleid ontstijgt de status van charlatanerie en fictie niet, stelt hij.

Dick Hanemaayer (2012), die 25 jaar beleidsevaluator was, geeft als hartenkreet
'dat overheidsbeleid vaak niet goed wordt voorbereid',
 en
'dat 'bij overheidsbeleid minimaal gebruik wordt gemaakt van ex ante evaluatie'.

Er moet dus *veel meer gedaan worden aan ex ante evaluatie*, menen de ervaringsdeskundigen Leeuw en Hanemaayer. Dat geldt voor alle overheidsniveaus. Niet blind meer, maar vooral bij weerbarstig en dus complexe problemen en daaraan gerelateerd beleid. Veelal zal dat beleid zijn in een netwerkcontext. Denk maar aan preventie tegen voetbalvandalisme.

Zijn ex ante evaluaties meer dan koffiedik kijken?

Er zijn methoden denkbaar die impliceren dat met een redelijke grote waarschijnlijkheid kunnen aangeven wat er wel of niet van veronderstellingen terecht komt. Simulatie van de uitvoering van een wetsontwerp is zo een middel (verder Hellendoorn, 2001; Hanemaayer, 2012).

William Dunn geeft in zijn boek *'Public Policy Analysis'* (5^e editie) een overzicht van deze methoden voor wat hij noemt *'forecasting expected policy outcomes'*. Daartoe behoren onder meer haalbaarheidstudies ('feasibility assessments'), trendanalyses, toekomstverkenningen, de Delphi methode en cross-impact analysis (Dunn, 2012: 165-180).

Inputevaluatie als vorm van ex ante evaluatie

Deze beschouwing gaat over typen evaluaties. Ex ante evaluaties zijn er in soorten. Effectrapportages zijn als ex ante evaluatie te zien. Deze zijn aspectmatig en doorgaans niet integraal gericht. Ik geef nog een bijzonder type: inputevaluatie. Inputevaluatie is gericht op de inspanningen die moeten worden verricht om bepaalde doelstellingen te realiseren. Dan zal het gaan om de kwantitatieve en kwalitatieve hulpbronnen, zoals mensen en middelen, die nodig zijn om beleid 'op te tuigen' en uit te voeren (Bekkers, 2007: 299).

Procesevaluatie als vorm van ex ante of ex post evaluatie

Een iets ander type evaluatie is de procesevaluatie (Bekkers, 2007: 299). Deze kan voorafgaande aan de belangrijkste besluitvorming over een beleid worden uitgevoerd of naderhand.

Met procesevaluaties wordt beoogd om te inventariseren wat nodig is aan processen om bepaalde doelstellingen te bereiken. De focus is dan gericht op de organisatie van het beleid en de beleidsuitvoering in het bijzonder, ook in interbestuurlijk perspectief. Bij voorbeeld, welke *invoeringsaanpak* van het beleid op het vlak van de Wet maatschappelijke ontwikkeling (wmo) is nodig? Welke *condities* zijn gewenst om te bereiken dat gemeenten na decentralisatie van taken en bevoegdheden deze wmo-taken op zich nemen? (Peters, 2006). Moeten *voorbeeldprojecten* ('pilots') worden gestart, zoals het geval was bij de Wet dualisering gemeentebestuur, zoals gemeenten van elkaar kunnen leren?

Efficiency-evaluatie als ex post of ex ante evaluatie

Een efficiency-evaluatie betreft een evaluatie van de verwachte of feitelijke kosten die de uitvoering van beleid met zich brengt. De vraag kan zijn hoe de kosten zich verhouden tot de verwachte of feitelijke baten (Bekkers, 2007: 300).

Vragen op dit vlak zijn bij voorbeeld:

- Maakt vergrijzing de AOW onbetaalbaar?
- Naleving en handhaving: loont het de moeite? (Prinsen en Vossen, 2003);
- Opsluiten of sleutelen?: kosten en baten van detentie en resocialisatie? (Slotboom en Wiebrens, 2003);
- Alcoholcontroles in het verkeer: wat leveren ze op? (Wesemann, 2003);
- Brandveiligheid in de horeca: wat zijn de kosten en baten van additionele maatregelen? (Suurmond en Van Velthoven, 2003).

Deze evaluatie is daarmee een ex post evaluatie. Het kan echter ook gaan om een ex ante evaluatie. Vaak zal aan de decentralisatie van een bepaalde taak naar gemeenten een taxatie van de kosten vooraf gaan.

Responsieve evaluatie

De beleidsevaluaties, zoals de klassieke aanpak van de effectiviteitsanalyse en de goal free evaluation worden wel *rationele evaluaties* genoemd. Hiertegenover staan *responsieve evaluaties*. Tineke Abma (2000, 2001) is vooral een verdediger hiervan. Ze wil vooral de kijk van verschillende actoren op beleid verduidelijken en niet alleen uitgaan van de overheidsblik. Haar analyses richten zich vooral op de gezondheidszorg.

Met haar pleidooi heeft ze nog niet veel voeten aan de grond gekregen omdat haar aanpak teveel een beleidsontwerp-aanpak is, waarbij ze alle partijen en zeker ook zwakkere, wier belang doorgaans slecht vertegenwoordigd is, aan bod wil laten komen.

Retrospectieve en prospectieve evaluatie ex post

Ik kom bijeen volgend onderscheid. Veel evaluaties zijn ex post en retrospectief en niet expliciet prospectief gericht. Beleidsvoerders zijn daarover wel eens ontevreden omdat deze evaluatie weliswaar aangeven wat bereikt is, maar niets zeggen over wat te doen om alsnog doelstellingen te bereiken of het doel doelmatiger te bereiken of iets anders. Ze zijn anders gesteld te weinig toekomstgericht.

Een voorbeeld. Toen de Vlaamse overheid in 2011-2012 haar protestedenbeleid wilde laten beoordelen door een visitatiecommissie –Reynaert/Korsten wilde ze juist niet zozeer een ‘beoordeling in cijfers’ maar een prospectieve evaluatie. De Vlaamse overheid wilde handreikingen krijgen voor een bepaling van toekomstige afspraken met de grote steden Brussel, Antwerpen, Gent, Mechelen, Leuven, Brugge, etc.. Ze wilde met andere woorden niet alleen een terugblik zien van wat met de afspraken de afgelopen jaren terecht gekomen was in termen van prestaties en effecten maar ook een vooruitblik.

Advies: Aan deze casus is een advies te ontlenuen. Een dergelijke gerichtheid van de evaluatie, retrospectief of meer prospectief, kon alleen aan het licht komen in de fase van de opstelling van een offerte en door een intensieve interactie van de visitatiecommissie met de opdrachtgever. De fase van opdrachtverlening is dus van zeer groot gewicht om verwachtingen af te stemmen over wat mogelijk is en feitelijk gedaan kan worden.

Van welke aard is het meeste beleidsevaluatieonderzoek?

Een WODC-onderzoek geeft daar meer informatie over. Sinds 1995 zijn 75 wetsevaluaties verricht. Klein Haarhuis en Niemeijer (2008) stelden vast dat 60 procent van de onderzochte wetsevaluaties *doelbereikingsstudies* waren. De vraag in welke mate wettelijke doelstellingen geheel of gedeeltelijk bereikt worden staat daarin centraal.

In tweederde van deze studies wordt ook aandacht besteed aan processen van uitvoering en/of houdings- en/of gedragsverandering bij de doelgroep. We zien dus dat doelbereikingsevaluatie vaak samengaat met (partiële) *procesevaluatie*. Tien procent van de evaluaties is een zuivere procesevaluatie. Onder de wetsevaluaties bevinden zich geen effectiviteitsevaluaties in strikte zin. Dat betekent dat of effecten bereikt worden door beleid of omstandigheden of gebeurtenissen buiten de wet niet is nagegaan.

11 DE EVALUATIELADDER: KIEZEN VOOR EEN LICHTE OF ZWARE EVALUATIE?

Geen enkel bestuur van een organisatie laat alle beleidsprocessen evalueren; geen enkel provinciebestuur, geen enkel gemeentebestuur. Dat betekent dat er vermoedelijk beleid is dat zich ook minder voor evaluatie leent en dat het kiezen van speerpunten in de afweging van wat een bestuur aan evaluatie onderwerpt nuttig is. En inderdaad, het ene beleid is zinvoller te evalueren dan het andere.

Routinematige beleidsprocessen behoeven vermoedelijk geen evaluatie. Het opstellen van een straatnamenplan is daartoe te rekenen.

Ook beleid waarvan de betekenis al snel zichtbaar is, hoef je niet nog eens te gaan doorlichten. Dat is 'overdone'. Als een gemeentebestuur van een plattelandsgemeente een dorpentocht heeft gehouden om voorgenomen beleid per plaats uit te leggen en de indrukken zijn overal dat dit een succes was (hoge opkomst, goede sfeer, helderheid in presentaties, levendige discussie, etc.) dan kan een schriftelijke rapportage over een evaluatie eigenlijk kort zijn en misschien is die wel overbodig.

Maar we kennen ook defensiebeleid: moet een JSF of Saab JAS 39 Gripen (twee vliegtuigtypen die genoemd zijn als opvolger van de F16) aangekocht worden of niet? Deze vraag is te plaatsen in enkele scenario's over de toekomst van de Nederlandse krijgsmacht. Daarvan kan een ex post evaluatie deel uit maken van troepenuitzendingen naar crisishaarden. We kennen klimaatbeleid en wat al niet. Er valt veel te evalueren.

Het laat zich op basis van het voorgaande aanzien dat er *een evaluatieladder* te maken is. Die ladder loopt vanaf het ene einde – beleid niet of heel licht evalueren – naar het andere einde – beleid zeker stevig aan evaluatie onderwerpen.

Geen evaluatie nodig

Welk beleid heeft geen evaluatie?

1. Beleid dat eerst recent aan evaluatie is onderworpen en waarover dus al veel bekend is. Dan is een evaluatie een beetje dubbelop.
2. Beleid dat volledig in medebewind wordt uitgevoerd en waarbij de invloed om te komen tot beleidsverandering in enigerlei vorm nagenoeg nihil is. Veelal is het dan aan de Rijksoverheid om een evaluatie te starten. Wie aan de knoppen ziet, moet laten evalueren, niet wie alleen naar de knoppen kan kijken.
3. In geval van een beleid dat geen speerpunt is van een overheidsbestuur maar dat secundair of tertiair is.
4. Beleid dat toch afloopt omdat het eenmalig en dus tijdelijk is.
5. Dat beleid waarmee niet veel budget gemoeid is. De evaluatie zou al snel duurder zijn dan het beleid zelf.
6. Dat beleid waarover geen of weinig verantwoording nodig is. Een straatnamenplan heeft geen evaluatie tenzij het bestuur overal straatnamen heeft gekozen van dictators en fascistoïde figuren. Weinigen willen wonen in de Hitlerlaan of de Avenue Mussolini.

7. Dat beleid waarvoor al een ander verantwoordingsinstrument is voorzien, zoals een jaarverslag.
8. Beleid dat evident onderdeel is van een bredere evaluatie, zoals een visitatie van de gehele organisatie.
9. Beleid dat evident een succes is en waarvan niemand wil weten waarom.

Zware evaluatie

Wat heeft naar verwachting juist een zware evaluatie?

1. Beleid dat behoort tot de autonome beleidsruimte voor een overheidsbestuur en te beschouwen is als een reactie op een *'wicked problem'*, dus een weerbarstig of controversieel probleem. Deze moeilijke problemen impliceren dat de kennis over beleid beperkt is en consensus over het beleidsprogramma moeilijk bereikt wordt. De aanpak wordt vaak ook later gevolgd door een aanpassing omdat bepaalde doel-middelcombinaties niet werken. Denk aan de bestrijding van voetbalvandalisme of aan provinciaal beleid tot uitplaatsing van agrarische bedrijven uit gebieden met te veel ammoniakuitstoot (problematiek van een landbouwontwikkelingsgebied, een LOG).
2. Beleid dat past in de programmabegroting en waarvoor de drie W's van toepassing zijn. Dan is evaluatie een voortzetting van een ingezette lijn.
3. Beleid dat behoort tot de speerpunten van een bestuur.
4. Beleid waarmee veel budget gemoeid is en/of een aanzienlijke ombuiging (in casu bezuiniging) overwogen is of kan worden.
5. Dat beleid waarbij evaluatie verplicht is ingevolge een wetsartikel of verordening of evaluatie ad hoc afgesproken is (bijv als gevolg van ene motie in een raads- of statenvergadering).
6. Dat beleid dat betrekking heeft op sterke gedragsveranderingen van burgers. Omdat niet aannemelijk is dat burgers eenvoudig tot gedragsveranderingen komen. Anders gesteld, er zullen veronderstellingen achter het beleid schuil gaan, die discutabel zijn.
7. Dat belangrijke beleid waarvoor geen andere verantwoordingsinstrumenten bestaan.
8. Dat belangrijke beleid dat al enige tijd niet aan evaluatie is onderworpen.
9. Complex beleid dat een pilot is voor vele andere overheden; waarvoor elders dus ruime belangstelling bestaat.

In de praktijk kom ik nog geen ladder tegen. De provincie Noord-Brabant heeft wel een evaluatieprogramma gemaakt. Dat programma is door gedeputeerde staten besproken met provinciale staten.

Deze discussie over een ladder roept onmiddellijk vragen op over of bij voorbeeld monitoring wel evaluatie is en benchmarking ook. Echter eerst iets over een lichte vorm van evaluatie.

Pseudo-evaluatie

Pseudo-evaluatie impliceert het op basis van beschrijving weergeven van geldige en betrouwbare informatie over beleidseffecten, zonder het ondernemen van een poging om effecten te waarden in relatie tot personen, groepen of de samenleving.

De achtergrond van dergelijke evaluatie is dat maatregelen onomstreden en evident worden geacht waardoor er helemaal geen maatstaven op het materiaal gelegd hoeven te worden (Dunn, 2012: 323). Stel dat een parlamentsmeerderheid jaar na jaar heeft aangegeven dat beken, sloten en vaarten elk jaar schoner moeten worden dan zal een studie die uitwijst dat in deze vormen van waterafvoer al vier jaar op rij het water schoner is geworden niet direct de vraag oproepen naar een duurzaamheidsmaatstaf of andere maatstaf.

DEEL IV: EX POST EVALUATIE NADER BEZIEN



12 MONITORING: LAAG OP DE EVALUATIELADER

Wat het is

Beleidsvaluatie kan dan wel breed worden opgevat als het beoordelen van beleidsprocessen, zoals het ontwerpen van beleid, de organisatie van het beleid, de besluitvorming of de uitvoering en effecten van beleid, William Dunn (2012) stelt dat beleidsvaluatie zich *hoofdzakelijk* richt op het vaststellen van waardeoordelen over *effecten* van beleid. Monitoring moet daarbij onderscheiden worden. Monitoring richt zich op de productie van feitelijke informatie over beleidsprocessen, met name van prestaties en (slechts) soms van effecten.

Vaak is een monitor interbestuurlijk ingebed. Een voorbeeld betreft de 19 febr. 2013 door minister Plasterk voorgesteld monitoring van de decentralisaties op het vlak van jeug, werk en zorg.

Citaat: 'Het Rijk blijft na de decentralisaties verantwoordelijk voor de werking van het stelsel. Hiertoe zal het Rijk met gemeenten afspraken maken over een intensieve monitoring van de bereikte resultaten en de gemeentelijke uitgaven over het geheel van het brede sociaal domein (en dus niet per onderdeel) door te meten en te benchmarken. Doel van deze monitor is drieledig:

Ten eerste biedt de monitor het kabinet inzicht in het functioneren van het stelsel zodat het kan worden aangepast indien het stelsel niet tot de juiste resultaten leidt. De uitkomsten zijn input voor bestuurlijk overleg tussen Rijk en gemeenten.

Ten tweede biedt de monitor inzicht aan gemeenten in hun eigen prestaties ten opzichte van andere gemeenten om van elkaar te leren en het horizontaal verantwoordingsproces te ondersteunen.

Ten derde biedt de monitor de informatie voor de Tweede Kamer over de behaalde resultaten in het sociaal domein. Gegevens voor de monitor worden in samenhang verzameld, verwerkt en gepresenteerd en zal onder regie van BZK worden ontwikkeld'.

Monitoring registreert vaak wat gebeurde, soms hoe het gebeurde en zelden waarom het gebeurde. Dat is dus '*magere*' evaluatie in een klein jasje. Beleidsvaluatie in een ruimer jasje laat zien of beleid daarbij verschil maakt(e).

Bongers en Den Hertog (2007: 32) plaatsen monitoring in de eerste generatie beleidsvaluaties.

Ook de methodoloog Peter Swanborn (2002) laat zien dat monitoring laag op de evaluatieladder staat. Hij stelt monitoring bij veel organisaties gebruikelijk is *zonder* dat er een specifieke interventie aan verbonden is of aan gerelateerd kan worden. Stel dat een monitor aangeeft hoeveel inbraken in een jaar plaatsvonden in een bepaalde gemeente. Monitoring wordt dan 'scannen' en staat zo volkomen los van overheidshandelen.

Monitoring wordt nogal eens gebruikt als het nagaan van de voortgang in beleid. Zo is er ooit een monitor geweest van het grotestedenbeleid (GSB). Ik signaleer ook een Netwerk Ecologische Monitoring (NEM) en een Monitoring Waterbeleid (WIB) (De Kool, 2007). Aan monitoring toegekende functies zijn volgens Dennis de Kool signaleren, sturen, verantwoorden, leren en communiceren.

In de praktijk gaan er dilemma's achter de keuze van monitoring schuil. De volgende komen aan bod.

Soberheid of volledigheid: Moet een monitor gericht zijn op details of op hoofdlijnen van beleid? Documenten die over hoofdlijnen gaan kunnen tot ongenueanceerde uitspraken leiden? Wie de details gaat opzoeken, kan wellicht de hoofdlijn niet meer vinden. Oplossing: stel een basisdocument op en een verdiepingsdocument.

Beschrijven versus beoordelen: Beleidsvoerders hebben doorgaans niet alleen behoefte aan beleidssignalen ('cijfers') maar ook aan verklaringen, interpretaties en oordelen (wel aangeduid als 'verhalen achter de cijfers'). Daarom wordt ook wel een onderscheid gemaakt tussen kale monitoring en evaluatie. Evalueren betreft zeker het beoordelen van ontwikkeling en oplossing: dus is het advies om monitoring van verdere stappen op de evaluatieladder gescheiden te houden (De Kool, 2007). Monitors zijn *in feite niet geschikt om te evalueren* maar kunnen wel een opstap vormen voor evaluatie. Daarbij is het in het algemeen nuttig, ook bij evaluatiestudies, om feiten en oordelen te scheiden.

Flexibiliteit versus stabiliteit: Bij monitoring is continuïteit in het signaleren van groot belang. Hoe ontwikkelt iets zich in de tijd? Daar staat tegenover dat beleid dynamisch is, niet stil staat, en informatiebehoeften in de loop van de tijd veranderen. Steeds weer naar nieuwe cijfers vragen in monitoring is echter niet goed. Het kan de vergelijkbaarheid in de tijd aantasten. Oplossing: neem een ruggengraat van vaste informatie op in een monitor en laat toe dat elk jaar een bepaalde hoeveelheid nieuwe informatie meegenomen mag worden, dus toegevoegd kan worden.

Concrete doelen versus vage ambities: In de praktijk is sprake van een neiging om te komen tot SMART geformuleerde doelstellingen (specifiek, meetbaar, acceptabel, realistisch en tijdgebonden) maar daar staat tegenover dat de praktijk steeds weer uitwijst dat er behoefte is aan een zekere vaagheid in de ambities (deuren open houden) om naderhand wendbaar te zijn en met doelstellingen te kunnen schuiven. Beleid is bovendien een compromis tussen politieke fracties die vaak verschillend aankijken tegen zaken. Dat leidt eerder tot meer dan tot midner vaagheid. Oplossing: Geven en nemen. Vertrouwen hebben in een monitor.

Conclusies

- 1) Monitoring registreert vaak wat gebeurde, hoe dat wat gebeurde of gebeurt zich in de tijd ontwikkelde, soms hoe het gebeurde en hoogst zelden waarom het gebeurde. Dat is dus heel '*magere*' evaluatie in een klein jasje. Beleidsevaluatie in een ruimer jasje laat zien of beleid daarbij verschil maakt(e). Monitors zijn in feite niet geschikt om echt met maatstaven overheidsactie te evalueren maar kunnen wel een opstap vormen voor evaluatie.
- 2) Monitoring is een instrument om te inventariseren, te signaleren, over bevindingen te communiceren en zich te verantwoorden. Maar monitoring is *geen* echte vorm van evaluatie.
- 3) Meestal zitten in monitors geen oordeel of maar beperkt. Monitoring staat dus heel laag op de evaluatielader. Waar staat benchmarking?

13 BENCHMARKING: HOGER OP DE EVALUATIELADDER

Benchmarking, wat het is

Het gaat bij benchmarking om cijfermatige vergelijking, om vergelijking van organisaties of eenheden en om verbetering. Beter gesteld, benchmarking betekent dat organisaties of delen van organisaties hun prestaties (laten) vergelijken met die van andere organisaties, om zo aanknopingspunten voor verbetering van het eigen functioneren te krijgen (Nelissen en De Goede, 1999; Bowerman, Ball & Francis, 2001; De Groot e.a., 2004; Hakvoort en Klaassen, 2004; Van Helden, Tillema en Van Esch, 2005; De Bruijn, 2001, 2007). Zo lieten regionale opleidingscentra zich jaarlijks benchmarken.

Vergelijking vindt plaats op één manier zodat een vergelijking makkelijk mogelijk is. Dan gaat het om vergelijking van *kwantitatieve* gegevens in de vorm van prestaties (output), zelden van effecten (outcomes). Er is dus meestal *géén* sprake van een vergelijking van effecten (outcomes). In die zin is *geen* sprake van ex post beleidsevaluatie, zoals we die kennen in de vorm van een doelbereikingsanalyse en effectiviteitsaanpak (Weiss, 1972). Er wordt in benchmarks ook minder diep gegraven.

Plaats op de evaluatieladder

Benchmarking startte ooit in het Nederlandse bedrijfsleven, vóór de tweede wereldoorlog. Het was toen en daarna iets dat veel weg had van bedrijfsvergelijkende statistiek om productieprocessen te verbeteren. Nadien heeft het concept benchmarking een evolutie doorgemaakt naar prestatieverbetering in meer algemene zin: systematisch opgezet, meer continu toegepast, gericht op verbetering (Hakvoort en Klaassen, 2004: 121-132).

Benchmarking is in mijn ogen echter, evenals monitoring, een heel lichte vorm van evaluatie en die vorm staat dus ook laag op mijn evaluatieladder.

Ik ga hier nader op benchmarking in omdat ik zelf als toezichthouder bij een Riagg en Roc ervaring heb gehad met de inhoud en gebruik van benchmarks en er onder mijn leiding onderzoek is gedaan naar benchmarking bij sociale diensten. Ik verrichtte ook zelf benchmarkonderzoek. Zie: *Gemeenten in rapportcijfers* (Abma en Korsten, 2009a, b).

We rekenen benchmarking tot het domein van deze beschouwing over beleidsevaluatie omdat sprake is van een beoordeling aan de hand van maatstaven. Beoordeling, daar draait het bij evaluatie immers bijna altijd om.

Toch is benchmarking anders dan veel beleidsevaluaties. Waarom? Vier redenen:

1) Van een benchmarking van bij voorbeeld *landbouwbeleid*, *emancipatiebeleid* of *personeelsbeleid* is geen sprake. Bijzonder aan benchmarking is namelijk dat *géén* sprake is van beoordeling van *afzonderlijke* beleidsprogramma's maar van instituties als geheel, dus van organisaties (met al hun beleid) of organisatiedelen of productiecomplexen. Denk aan een benchmarking van waterschappen, ziekenhuizen, onderwijsinstellingen of studierichtingen, gemeenten.

Een voorbeeld van een benchmark is 'De Staat van de Gemeente'. Dit is een met de Vereniging van Gemeentesecretarissen (VGS) ontwikkeld instrument dat is gebruikt om gemeentelijke organisaties met elkaar te vergelijken op basis van onder meer enquêtes onder burgers. Daarbij werden per gemeente gegeven over burgers in vijf rollen bij elkaar gebracht: de burger als kiezer, wijkbewoner, belastingbetaler, onderdaan, participant (Abma en Korsten, 2009a, b). De manier van rapporteren door

het rapporterende instituut was dan dat een rapport gemaakt werd met kwantitatieve gegevens van vijftien vergelijkbare gemeenten.

2) Meestal wordt geen eigenstandig beoordelingskader gebruikt maar worden indicatoren, in de vorm van een aspect van een verschijnsel, toegepast die in een cijfer zijn te vatten.

3) De analyse is kwantitatief, niet kwalitatief.

4) Er verschijnt een totaalrapport, vaak geen afzonderlijk rapport per bestudeerde en deelnemende organisatie.

Benchmarking is altijd achteraf (ex post), nooit ex ante.

Beoordelingskader

Een referentiepunt voor het eigen functioneren wordt gevonden in een aantal andere organisaties (vaak meer dan 10) maar stel nu dat de andere organisaties ronduit slecht functioneren, wat dan? Dan zegt een vergelijking niet veel. Daarom is bij benchmarking een *referentiekader* nodig. Dat kader wordt doorgaans gevormd door *de best* presterende organisatie. Dan kan een echt bestaande organisatie zijn maar dat kan ook een denkbeeldige organisatie zijn die geconstrueerd wordt door alle beste scores op vergelijkingspunten bij elkaar te zetten. In het laatste geval wordt de beste organisaties, opgebouwd uit 'best practices', een *fictieve* creatie van de beste organisatie door de onderzoeker. Wellicht bestaat een dergelijke beste organisatie in werkelijkheid niet.

Het is ook mogelijk om de gemiddelde organisatie uit een vergelijking op een aantal criteria als ijkpunt te laten functioneren, waarna bij een vergelijking het behoren tot *de beste helft* van de organisaties als 'goed' geldt.

Functies van benchmarking

Welke functies vervult benchmarking?

1) *Leren*. Bij veel overheidsorganisaties ontbreekt de tucht van de markt om goed te presteren. Benchmarking kan dan een instrument zijn om toch te prikkelen tot prestatieverbetering. Het gaat bij benchmarking om vergelijking gericht op prestatieverbetering met als referentie de goede of beste praktijken die uit het onderzoek zelf komen. Zo gezien, is benchmarking dus bedoeld om een organisatie tot 'leren' (uitmondend in verbeteren) te brengen. Deze functie geldt algemeen als de belangrijkste functie (Van Helden, Tillema en Van Esch, 2005)

2) *Verantwoording*. Het is ook mogelijk dat het doel is 'transparante verantwoording'. Dat was bij voorbeeld het geval bij bedrijfsvergelijking van het zuiveringsbeheer van waterschappen (Van Helden, Tillema en Van Esch, 2005).

3) *Bekostiging*. Benchmarking wordt ook wel gebruikt voor de normering van externe bekostiging. Bij voorbeeld: de rijksoverheid vergoedt voor die en die hoeveelheid en kwaliteit van prestaties.

4) *Bezinning op ambities*. Uitkomsten van benchmarking worden hier en daar wel gebruikt om de bestuurlijke ambities te stabiliseren of op te schroeven, iets wat men in de literatuur weinig leest. Benchmarking kan een opstap naar discussie vormen en om vragen te stellen als 'waarom scoort onze organisatie op dat punt nou beter of slechter?'

Een directeur van een regionaal instituut voor geestelijke gezondheidszorg waar ik toezichthouder was, zei ooit tegen mij:
‘Deze Riagg wil tot de vijf beste in het land behoren. Eens zien of dat ook uit de benchmark komt’.

5) *Bezweringsformule*. Een benchmark kan ook als bezweringsformule dienst doen in de conversatie tussen een toezichthoudend orgaan en een directie of college van bestuur. Een benchmark kan voor rust zorgen.

‘Als lid van een Raad van Toezicht van een regionaal opleidingscentrum voor mbo-onderwijs zag ik sinds 2000 elk jaar benchmarks langs komen waaraan alle Roc’s in het land hadden deelgenomen. De waarneming dat “onze” Roc in de beste helft zat, werd altijd gemeld. Die mededeling had iets geruststellend in zich, waar iedereen zich ook bewust van was. Maar soms was een benchmark in combinatie met verklaringen over wat achter de cijfers schuil ging een aanzet om een bepaald onderwerp of beleidsvraagstuk nader te bekijken.’

Voorwaarden

Benchmarking kan gedijen als aan bepaalde condities wordt voldaan. Ik noem er enkele.

1 Een vergelijking moet zinvol zijn.

Een benchmark is vaak een vergelijking van alle academische ziekenhuizen in Nederland of een vergelijking van vijftien sociale diensten van ongeveer een gelijk aantal personeelsleden of cliënten. Daaruit blijkt dat een vergelijking zinvol moet zijn. Het is vaak zinloos om een groot academisch ziekenhuis te vergelijken met een heel klein ziekenhuis, of om de sociale dienst van een grote gemeente als Amsterdam met die van pakweg Barneveld te vergelijken.

2 De indicatoren moeten relevant zijn.

De indicatoren voor de organisatievergelijking moeten zo relevant zijn dat ze ook echt iets zeggen over *de kern* van de organisatie en of te meten deel daarvan. Hier gaat het wel eens fout omdat *alleen het kwantificeerbare en makkelijk meetbare* meegenomen wordt in een vergelijking, of dat nu de essentie is of niet.

3 Een breed palet van indicatoren is gewenst.

Een *breed palet* aan prestatie maatstaven blijkt bevorderlijk voor het krijgen van draagvlak bij bestuurders, managers en andere gebruikers om benchmarking te starten of aan een initiatief van anderen mee te doen (Van Helden, Tillema en Van Esch, 2005: 15).

Beperkingen en grenzen van vergelijking

Benchmarking kent beperkingen. Ik noem er enkele.

1) Vooral kwantitatieve vergelijking, geen kwalitatieve.

Benchmarking is sinds 1990 in toenemende mate toegepast in het openbaar bestuur maar vooral in sectoren waar het mogelijk is te tellen en te turven bij voorbeeld in de vorm van bedden, beddenbezetting, de ratio (verhouding) tussen verzorg(st)ers en verzorgden enz.. Zo zijn benchmarkingoperaties toegepast bij thuiszorgorganisaties, regionale instituten voor geestelijke gezondheidszorg, waterschappen, waterleidingbedrijven, sociale diensten. Daarmee wordt al duidelijk dat benchmarking zich leent voor het opstarten van discussies over doelmatigheid. Een benchmark van geheime diensten of inlichtingendiensten als de AIVD zal men vergeefs zoeken. Overigens is de AIVD wel eens tegen het licht gehouden (Commissie-Havermans, 2004).

2) Een tweede beperking van benchmarking houdt in dat ze vooral beschrijvend, niet verklarend zijn.

Wie een benchmarking start en deelneemt aan een vergelijking moet ook beseffen dat de vergelijking vaak vragen oproept over de duiding en over wat er achter de cijfers schuilgaat. De gebruiker moet dus de grenzen van een benchmark kennen.

3) *Beperkte gebruikswaarde.*

De bereidheid, naar aanleiding van benchmarking van waterschappen op het punt van bedrijfsvoering zuiveringsbeheer (bvz), om echt met verbeteringsacties in de vorm van plannen te komen is bij waterschappen 'wisselend' gebleken (Van Helden, Tillema, Kuppens en Dekking, 2005:38). Een tegenvallend resultaat. Een op elke vier waterschappen (27 procent) ondernam geen enkele actie op basis van de benchmarking en 36 procent laat slechts een bescheiden actiebereidheid zien. Dat is samen 63 procent. 'De link tussen bedrijfsvergelijking en actiebereidheid is vrij zwak' concluderen Van Helden c.s. dan ook.

De auteurs zeggen dat benchmarking wel 'werkt' omdat waterschappen over het algemeen een positief oordeel hebben over de bedrijfsvergelijking maar benchmarking werkt 'nog niet goed genoeg'.

De voortzetting van de bedrijfsvergelijking heeft volgens de auteurs alleen zin als deze aanzet tot prestatieverbetering. 'De link moet worden versterkt', zeggen ze. Zeker omdat ook blijkt dat het niet vanzelfsprekend de zwakke waterschappen zijn of de waterschappen die echt prestaties moeten verbeteren, die ook aan het verbeteren slaan. Een urgentie leidt nog niet zo maar tot bestuurlijk en ambtelijk urgentiebesef.

Remmend op gebruik van uitkomsten uit rapportages van benchmarking is dat een organisatie altijd wel ergens minder goed en ergens anders juist wel goed in is. Dat leidt tot een 'ach, ja...'-houding.

Een tweede factor die remmend werkt op gebruik van benchmarkuitkomsten luidt: vaak is in rapportages over benchmarking wel sprake van beschrijvende vergelijking maar *niet* van verklaring waarom iets is zoals het is of bepaalde tendenties blijken. Terwijl die verklaring van prestatieverschillen nogal eens juist nodig is om de gebruikswaarde te vergroten, zoals Van Helden, Tillema en Van Esch (2005: 15) stellen. Bij de bedrijfsvergelijking zuiveringsbeheer (bvz) werd slechts 'in beperkte mate aan deze voorwaarde voldaan', aldus de genoemde onderzoekers.

Wat heeft verder dan wel positieve invloed op de gebruikswaarde? Wat werkt accelererend? Uit het bvz-onderzoek onder waterschappen blijkt: *grote externe druk*, soms ook concurrentie, om cijfers te gebruiken, leidt tot meer gebruik van uitkomsten van benchmarking.

Daaraan is een krachtige aanbeveling te ontleen: zorg ervoor dat bezinning op evaluatie-uitkomsten verplicht is, dat er druk op de ketel is van de waterschapsbesturen zelf. De Unie van Waterschappen zou dus kunnen proberen te werken aan het bereiken van een betere benchmarkcultuur. De auteurs zeggen dat het introduceren van visitaties om *per waterschap* met verbeterpunten te komen te overwegen is. Dit zou voor de sector als geheel ook kunnen gebeuren. Visiteren geeft meer druk. Het zijn immers deskundigen en andere kenners van de waterschapsorganisaties die visiteren.

Verder werkt positief op gebruik van uitkomsten van benchmarking in:

- dat de rijksoverheid in de rijksvisie op de waterketen uit 2003 voor benchmarking in de waterketen pleit; *een stimulans helpt*;

- de rapportage moet duidelijk zijn en de hoeveelheid 'huiswerk' om de uitkomsten naar de eigen organisatie toe te trekken of vertalen moet beperkt zijn;
- er moet *aandacht georganiseerd worden*, in de vorm van agendering in werkgroepen en/of organen, zodat er ook gesproken kan worden over uitkomsten;
- het delen van informatie uit benchmarks in een netwerk en *erover overleggen*, is positief omdat zo vakgenoten met elkaar overleggen;
- de aanwezigheid van *bestuurlijke ambitie* (de lat hoog leggen);
- het vrijmaken van middelen voor vernieuwing gericht op prestatieverbeteringen;
- *positieve interactie* tussen bestuur en ambtenaren die stimulerend werkt op kijken naar uitkomsten en zich erop bezinnen;
- ambtenaren enige *speelruimte* laten om verbeteringen te doordenken en/of aan te brengen;
- het '*anderen doen het ook*'-effect of '*wij moeten mee*'-effect.

3) *De kracht van de indicatoren*. Raken de indicatoren, als sleutelindicatoren, de kern van de kwaliteit van de belangrijke taken van organisaties of schieten ze er langs heen? Indicatoren die de kern vatten zijn lang niet altijd opgenomen in benchmarking.

4) *De oordelen zijn 'dun'*. De rapportages over benchmarks zijn doorgaans sober. Dat betekent dat benchmarking verre van uitmunt in de kwaliteit van de oordelen. Ik refereer hier aan benchmarking van mbo's en rioleringen.

Figuur 6: Vergelijking van drie op kwaliteit gerichte methoden van gegevensverzameling

	Monitoring	Benchmarking	Zelfevaluatie . visitatiecommissies
Functie: Interbestuurlijke signalering	Vaak	Meestal niet	Soms
Functie: Verantwoording	Ja	Ja	Ja
Functie: Gericht op leren/verbeteren	Niet zeker	Ja, expliciet	Ja
Openbaarheid van verstrekte gegevens	Meestal niet actief	Ja	Ja, tenzij....
Gegevensverzameling over beleid of	Beleidsprogramma	Beleid/ sector	Organisatie
Uitvoering	Extern, met verstrekking gegevens	Extern, met verstrekking gegevens	Extern, met zelfevaluatie
Referentiekader	Geen beoordelingskader Van waaruit gegevens worden verzameld	Geen referentiebeeld. Vergelijking met totaal of gemiddelde	Vrijwel altijd een beoordelingskader met normering (= 'governance' beeld)
Kwantitatief en kwalitatief gericht	Kwantitatief, beperkt kwalitatief	Nadrukkelijk kwantitatief	Kwantitatief en kwalitatief
Oordelen over gegevens		Rangorde	Beoordeling op basis van normen. Rapport per beoordeelde eenheid

Waardering van benchmarking

Wat houdt de waardering voor benchmarking in?

Er zijn ervaringen opgedaan met een benchmarkingproject van de Nederlandse waterschappen, de zogenaamde bedrijfsvergelijking zuiveringsbeheer (bvz). Waterschappen zijn overheidsorganisaties met een specifieke taak. Waterschappen houden zich met name bezig met waterkwantiteitsbeheer (dijkbeheer, waterpeilbeheer) en met een deel van het waterkwaliteitsbeheer (met name de zuivering van afvalwater). Wordt het instrument gewaardeerd door de besturen van waterschappen en zijn concreet ook verbeteracties waargenomen?

Waterschappen waarderen in het algemeen het instrument benchmarking (bzz) in hun sector. Het geeft een goed inzicht in de prestaties en geeft 'redelijke aanknopingspunten voor prestatieverbetering' (Van Helden, Tillema, Kuppens en Dekking, 2005: 38). Waterschappen oordelen over het algemeen 'gematigd positief' over hun prestaties op bepaalde gebieden en minder op andere.

Conclusie

De algemene conclusie van deze beschouwing luidt dat benchmarking betrekkelijk laag op de evaluatieladder staat. De analyses zijn sterk kwantitatief, een rapportage met duiding van de cijfers ontbreekt vaak en een rapportage voor tien of meer deelnemende organisaties is voorzien waardoor een individuele organisatie nog een eigen vertaalslag moet maken (wat dan niet of niet grondig gebeurt).

In beginsel is benchmarking een bruikbaar instrument is om overheidsorganisaties aan te zetten tot het verbeteren van prestaties maar tevens bleek dat aanvullende instrumenten nodig zijn om tot prestatieverbetering te komen, zoals druk op de ketel zetten en visitaties starten. Visitatiecommissies spreken indringend met bestuurders, ambtenaren en doelgroepen en benutten zelfevaluaties. Bij benchmarking is doorgaans zoets als een benchmarkcommissie afwezig.

Veel onderzoek wijst in de richting dat benchmarking lang niet altijd brengt wat ervan verwacht wordt. Uitkomsten van benchmarking wordt vaak wel aan gebruikszijde bekeken maar makkelijk ter zijde gelegd. De doorwerking is hier en daar betrekkelijk gering. Symbolisch gebruik komt voor. Dan wordt gesteld: 'we waren toch al van plan de kant op de gaan die het onderzoek opwijst!', een niet zelden gratuite mededeling.

Advies: Wie aan benchmarking gaat doen, moet daarom nadrukkelijk bij de inrichtingsbeslissingen aandacht besteden aan inrichtingsbeslissingen en gebruikswaarde.

14 ZELFEVALUATIE EN VISITATIECOMMISSIES: HOOG OP DE EVALUATIELADDER

Evaluëren met behulp van visitatiecommissies en zelfevaluatie, hoe gaat dat en hoe kunnen we die aanpak waarderen?

Ik kom nu bij zelfevaluatie ten behoeve van een visitatieproces en visitatieprocessen die geen zelfevaluatie kennen. Ik heb daarmee zelf ruime ervaring opgedaan door onderwijsvisitaties in het wetenschappelijk onderwijs te 'ondergaan', aan onderzoeksvisitaties deel te nemen en ik heb zelf een visitatie in het hbo geleid en een midtermevaluatie van een opleiding aan de Erasmus Universiteit. Bovendien was ik lid van twee visitatiecommissies ter beoordeling van twee provinciebesturen en hun organisatie (Zeeland, Limburg) en lid van een commissie die zich bezig hield met de bestuurlijke toekomst van Groningen.

Plaats op de evaluatieladder

Visiteren met toepassing van zelfevaluaties kan als vorm van gebruiksgesichte evaluatie kan hoog op de evaluatieladder worden geplaatst. Niet omdat deze visitatie wetenschappelijk gericht zou zijn, want dat is niet zonder meer het geval. Maar wel omdat er sprake is van

- zorgvuldig zoeken naar relevante, geldige, precieze en betrouwbare gegevens,
- het visitatieproces iteratief is en de gevisiteerden betrokken zijn (het gaat niet over hen en zonder hen maar juist met hen), en
- er sprake van onafhankelijke beoordeling door 'vakgenoten' op basis van een beoordelingskader dat door de gevisiteerden in de sector bekeken en aanvaard is.

Meestal is deze visitatie *ex post* en *retrospectief* (Korsten, 2004 a,b en c; Schillemans, 2009).

Populariteit

Visitaties kennen sinds een aantal jaren een grote populariteit. Die danken ze voor een groot deel aan de bijdrage die ze kunnen leveren aan leren in de publieke sector. Visitatierapporten bevatten immers een brede feedback op het handelen of niet-handelen van overheden als gemeenten of provincies, corporaties of politiekorpsen. Bezinning op de organisatie kan tot verbetering leiden. Ze zijn in potentie dus duidelijk meer dan loze manier van verantwoording over geleverde prestaties (Korsten, 2004a,b en c; Schillemans, 2009: 162; Grooteman, 2004).

Conditie en inrichtingsbeslissingen

Sinds ongeveer 1985 komen in Nederland visitaties voor. Aanvankelijk werden ze vooral gestart in het academische onderwijs en hoger beroepsonderwijs, omdat de colleges van bestuur daartoe het initiatief namen. Er viel in het academisch onderwijs niet te ontkomen aan meer verantwoording en leervermogen dan de onderwijsinspectie kon oproepen. De staf van de Vereniging van Samenwerkende Universiteiten werkte daarom ten behoeve van de CvB's een plan uit. In dat aanvaarde plan stond dat alle opleidingen eens per vijf jaar zouden worden doorgelicht op kwaliteit in de vorm van een combinatie van zelfevaluatie op basis van een protocol en een onafhankelijke visitatiecommissie van vakgenoten en onderwijskundigen zonder banden met afzonderlijke opleidingen. Deze praktijk ging van start en is over het algemeen een succes gebleken. Recent is deze visitatievorm overgegaan in een *accreditatieproces*.

De visitatiecommissies voor academisch onderwijs genoten in de eerste twee gehouden onderwijsvisitatierondes legitimiteit, hun oordelen werden als gezaghebbend ervaren en de bij opleidingen betrokken bestuurders, docenten, studenten en ondersteuners herkenden zich in het geschetste beeld van de afzonderlijke opleiding en het geheel van opleidingen. De bevindingen werden ook erkend. Van bezwaar maken was nauwelijks sprake. Logisch, want elke opleiding had de kans gehad om informatie in te brengen over tal van onderwijsaspecten,

zoals rendementen en doorstroming. En men had de kans gehad op een conceptrapport commentaar te geven.

Rond visitaties zijn dus tal van inrichtingsbeslissingen genomen (Schillemans, 2009). Op basis van deze geijkte praktijk zijn *beginselen van behoorlijke visitatie* te formuleren (Korsten, 2003a, 2004a, 2004b, 2004c).

Beginselen van behoorlijk visiteren

Op grond van evaluaties zijn de volgende beginselen van behoorlijke visitatie geformuleerd:

- 1) Er is een initiatiefnemer tot visitatie.
- 2) Er wordt een protocol vastgesteld met de opzet, procedure, het kwaliteitskader, de inhoud van de zelfevaluatie of zelfstudie etc.
- 3) De visitatie vindt plaats onder leiding van een visitatiecommissie.
- 4) De commissie is onafhankelijk.
- 5) De commissie hanteert een kader ter beoordeling van het onderwerp van beoordeling en procesplan.
- 6) De visitatie ontvangt een zelfevaluatie dat voldoet aan eisen uit een protocol.
- 7) De visitatiecommissie brengt een bezoek ter plaatse om de zelfevaluatie te checken en aanvullende gegevens te verwerven.
- 8) Er is een iteratief proces: betrokkenen (gevisiteerden) worden gehoord en krijgen een conceptrapport voor commentaar voorgelegd.
- 9) De commissie brengt een rapport uit per gevisiteerde opleiding.
- 10) Het visitatierapport is openbaar.

Als aan deze beginselen is voldaan is de kans zeer groot dat de gevisiteerden tevreden zullen zijn over het proces en de uitkomsten. Ze zullen zich herkennen in de 'foto' en de 'foto' erkennen.

Geijkte praktijk

Visitatieprocessen hebben ook plaats gevonden in de wereld van grotestedenbeleid, gemeenten, politiekorpsen, woningcorporaties, ziekenhuizen, media, interdepartementale beleidsprojecten.

De praktijk van visiteren in het grotestedenbeleid vertoonde veel gelijkenis met die in het hoger onderwijs. De commissie, onder leiding van Elco Brinkman, was samengesteld uit ex-bestuurders op rijks-, provincie- en gemeenteniveau. De commissie-Brinkman ontving op basis van een protocol een zelfevaluatie per gemeente, ging ter plaatse kijken en in gesprek en stelde een conceptrapport op waarop reactie van het gemeentebestuur mogelijk was. Dat leverde een openbaar rapport op. Zie daar de gelijkenis.

Deze ervaringen *bevestigen* het belang van de 'beginselen van behoorlijk visiteren'. Dat betekent dat de kans groot is als een visitatieproces *volledig* gebaseerd is op deze beginselen het visitatieproces ook positief gewaardeerd wordt in de sector en bij de gevisiteerden.

Deze bevestiging bleek opnieuw uit analyse van de gang van zaken bij de Limburgse bestuurskrachtmeting in meer dan 30 gemeenten. Ook hier acceptatie van het normenkader en de procesgang, herkenning van de 'foto' uit het visitatierapport en erkenning van de uitkomsten.

Bij politiekorpsen en ziekenhuizen was sprake van een zekere afwijking van deze beginselen. Ook politiekorpsen stelden wel een zelfevaluatie op maar deze werd doorgaans *niet* openbaar. De ziekenhuisvisitaties zijn zo gevoelig dat het rapport zeker *niet* openbaar is.

Het laten uitvoeren van visitaties door adviesbureaus *zonder* visitatiecommissie is in het algemeen ongewenst omdat besturen nogal eens de neiging vertonen om adviesbureaus te sturen. Dit bleek mij als lid van een begeleidingscommissie bij voorbeeld uit de opstelling van het provinciebestuur van Zuid-Holland bij een bestuurskrachtonderzoek uitgevoerd door het bureau Van Naem en Partners.

Retrospectief en ook prospectief?

Doorgaans zijn visitaties retrospectief. Er wordt dus gekeken naar de 'performance', de gebleken prestaties. In eerste onderwijsvisitaties van universitaire opleidingen zijn geen vooruitblikken te lezen of aanbevelingen. Maar bij latere veranderde dat.

Onderwijsvisitatiecommissies houden wel degelijk een spiegel voor. Dus komen zaken naar voren die moeten verbeteren. De onderwijsinspectie zit daarop ook toe en daarmee verkrijgt je dus de *externe druk* die van positieve invloed is op gebruik of benutting van uitkomsten, iets dat rond benchmarking bij waterschappen al van belang bleek.

De visitatie van Limburgse gemeenten was in de eerste ronde ook retrospectief en qua toekomstoriëntatie 'kaal'; zonder aanbevelingen. In de tweede ronde is deze commissie-Rutten opgeschoven. Ze bleef wel oordelen maar ging ook meer meedenken met gemeentebesturen. Dat werd gewaardeerd.

Soms is een visitatie niet retrospectief maar meer nadrukkelijk prospectief. Een voorbeeld hiervan is de visitatie van het grotestedenbeleid van de Vlaamse steden die in 2011 in opdracht van de Vlaamse overheid is uitgevoerd. Deze visitatiecommissie-Reynaert/Korsten was retrospectief maar ook prospectief omdat de commissie geacht werd met de Vlaamse overheid en de stadsbesturen mee te denken over verdere mogelijkheden en door 'best practices' aan te reiken. In de voorbereidingsfase op de visitatie bleek dat volstaan met louter retrospectie ongewenst was.

Conclusie

Visiteren is een nuttige vorm van evalueren. Vooral de bijzondere organisatiewijze valt op en het feit dat de organisatie in het hele proces van het begin tot het einde meegenomen wordt. De zelfevaluatie dwingt organisaties al snel tot het verwerven van overzicht en het zich vormen van een beeld over waarin de organisatie en het bestuur sterk en zwak is. Inmiddels zijn over visiteren inzichten verworven, zoals de beginselen van behoorlijk visiteren.

Succes is bij visiteren niet altijd gegarandeerd. De samenstelling van de commissie, de teamhomogeniteit en de aanwezige competenties bleken van belang. Een commissie dient namelijk gevoel te hebben voor het zorgvuldig verloop van een procedure, voor onderzoek, voor goede interactie met gevisiteerden en voor 'wijze' beoordeling.

DEEL V: EX ANTE EVALUATIE NADER BEZIEN

15 EX ANTE EVALUATIE

Ex ante evaluatie en het ontwerpen van beleid

Ex ante evaluaties (EAE) strekken ertoe om in te schatten of het wel lukt om te bereiken wat je als overheid wilt bereiken als je tot een bepaald beleid zou besluiten, of helemaal niet en waar je dan strandt. Ex ante evaluatie gaat in feite over de zinvolheid van een beleidsvoornemen, een beleidsalternatief, aldus Hakvoort en Klaassen (2004: 53; ook: Hellendoorn, 2001; Hanemaayer, 2012).

De overheid streeft na om met ex ante evaluatie (EAE) *voorafgaand* aan de (centrale) besluitvorming over beleid de beleidsinhoud, het proces en het bereiken van resultaat van beleid te verbeteren. Ex ante evaluatie kan zo een onderdeel worden van de fase van beleidsvoorbereiding en het ontwerpen van beleid. Ze zijn bij uitstek prospectief gericht.

Getracht wordt een 'eerste klas enkele reis Titanic' te voorkomen, dus een deceptie of beleidsfiasco.

Ex ante evaluatie: enkele typen

Aan welke evaluaties valt te denken? Aan de volgende:

- kosten-batenanalyse en kosten-effectiviteitsanalyse (Suurmond en Van Velthoven, 2003; Slotboom en Wiebrens, 2003; Wesemann, 2003);
- multicriteria-analyses (Voogd, 1988; De Vries, 1992, 1994);
- consequentie-analyses (Becker, 1987);
- ex ante schatten van de uitvoerbaarheid van beleid;
- effectrapportages (Koppenjan, 1995; Korsten, 1997).

Van deze typen zijn de eerste drie genoemde in principe *integrale* evaluaties, die in principe gericht is op het nagaan of alle effecten bereikt kunnen worden, en is de effectrapportage een *partiële* evaluatie omdat deze slechts gericht is op bepaalde belangen of bepaalde effecten. Denk aan de milieu-effectrapportage waarin de verwachte gevolgen voor het milieu van een bepaald beleid wordt nagegaan. Dat is een *partiële evaluatie*.

Op deze types ga ik kort in.

Integrale monetaire methoden: kosten-batenanalyse als voorbeeld

De *kosten-batenanalyse* (kba) wordt wel een monetaire methode genoemd. Kenmerkend hiervoor is om alle kosten en baten in geld uit te drukken (Hellendoorn, 2001). Het is dan ook begrijpelijk dat het ministerie van Financiën de toepassing hiervan stimuleerde. De minister van Financiën heeft in verband de begrotingsverantwoordelijkheid een belang bij meer baten dan kosten en in het algemeen bij lage(re) kosten.

Er wordt ook wel gesproken over een *maatschappelijke kosten-batenanalyse* (mkba). Hierbij gaat het om de vraag of de maatschappij in haar totaliteit een nettovoordeel heeft van het uitvoeren van een project, zoals bij voorbeeld de inpoldering van de Markerwaard of de aanleg van de Tweede Maasvlakte.

Bij toepassing van dit soort methoden wordt soms een referentiesituatie gebruikt. Dat is meestal de situatie die naar voren komt als het project niet zou doorgaan (ook wel nulsituatie genoemd).

Alle kosten-batenanalyses hebben als probleem dat sommige aspecten of elementen van beleid niet in geld uit te drukken zijn (Hakvoort en Klaassen, 2004: 60).

Daarnaast bestaan er *kosten-effectiviteitsanalyses*, kea's. Deze zijn verwant aan kosten-batenanalyses, ook wel kba's genoemd. De overeenkomst bestaat hierin dat gepoogd wordt om een zoveel mogelijk gekwantificeerd overzicht te krijgen van voor- en nadelen van beleidsalternatieven. Zo kan een vergelijking *tussen* de alternatieven ontstaan en *rangorde* van alternatieven (Hakvoort en Klaassen, 2004: 61). Een voorbeeld van een kosten-effectiviteitsanalyse betreft de analyse van alternatieve aanlandingsplaatsen voor vloeibaar aardgas (LNG).

Integrale niet-monetaire methode: multicriteria-analyse als voorbeeld

Multicriteria-analyse (mca) is een evaluatiemethode om tussen diverse alternatieven een rationele keuze te maken op basis van meer dan één onderscheidingscriterium.

- Dat is het eerste bijzondere: er is nadrukkelijk sprake van *meerdere beoordelingscriteria*. Bijvoorbeeld: met mca kunnen scores op economische, ecologische en sociale criteria bij elkaar worden opgeteld, om alternatieve trajecten voor bij voorbeeld een nieuwe weg te rangschikken.
- Het kan bij mca gaan om uitdrukking in *diverse eenheden*: meters, oppervlaktes, tijd, euro's. Er is dus niet uitsluitend sprake van een uitdrukking in geld (euro's). Het is bovendien mogelijk dat er een verschil in zwaarten bestaat tussen de beoordelingscriteria; het vraagstuk van *de gewichten*.
- Zowel kwantitatieve *als kwalitatieve* gegevens worden in de analyse meegenomen.

De doelen van een multicriteria-analyse zijn het ordenen van gegevens, het transparant maken van beslissingsprocessen en het ondersteunen van beslissers. De methode lijkt op maatschappelijke kosten-batenanalyse (mkba), maar bij mca hoeven niet alle criteria te worden uitgedrukt in monetaire eenheden.

Een multicriteria-analyse is bij voorbeeld toe te passen bij besluitvorming over een nieuwe huisvestingslocatie van een schoolgebouw (Voogd, 1988; De Vries, 1992). Maar er zijn ook adere toepassingen mogelijk. Denk aan de keuze voor een maaltijdvoorziening in Tytsjerkstradiel, over het sociaal leefklimaat in Roermond, de inrichting van het Lauwersmeer, de vestigingsplaatskeuze voor een kerncentrale (De Vries, 1994).

Consequentie-analyse

Bij consequentie-analyse gaat het erom om 'voorafgaand aan een beleidsactiviteit zo veel mogelijk na te gaan welke effecten en neveneffecten van deze activiteit verwacht mogen worden' (Becker, 1987: 117). Een dergelijke analyse kan volgens Henk Becker worden toegepast als een overheid bij voorbeeld deregulering overweegt. Wat gebeurt er als je die en die regels gaat schrappen?

Volgens Becker (1987: 117) moet je zo een analyse starten met een probleemanalyse: verken welke actoren betrokken zijn, welke hun onderlinge relatie en interacties zijn en welke knelpunten ze ervaren? Daarna maak je een basisanalyse: wat is de (voor)geschiedenis van een probleem en oplossing, en wat zijn de in dit verband relevante oorzaken en gevolgen. Vervolgens zie je wat beïnvloedbaar is door een overheid en wat niet? Tegen dit decor kan een onderzoeker vervolgens nagaan welke beleidsalternatieven zich aandienen. De consequentie-analyse houdt in dat nagegaan wordt van elk alternatief wat de bijdrage kan zijn

aan de oplossing of reductie van een probleem. De kern vormt het aftasten van de te verwachten effecten en neveneffecten voor de belangrijkste alternatieven. Daarna is aan de alternatieven een rangorde toe te kennen, een volgorde van wenselijkheid.

Volgens Becker zijn effectrapportages, zoals een milieu-effectrapportage een vorm van consequentie-analyse.

Overzicht van de omvang van ex ante evaluatie

Dean Steenbeek (1997) heeft in de jaren negentig 28 EAE's op tien departementen aangetroffen. Hiervan zijn er medio 1997 (vermoedelijk) veertien in ontwikkeling en veertien in gebruik. De afzonderlijk opgevoerde MDW-toetsen worden als een toets geteld. MDW staat voor marktwerking en deregulering, een operatie die liep tijdens het kabinet-Kok (1994-1998).

Bij personen die EAE's laten uitvoeren is sprake van *drie uitgangspunten*, namelijk dat beleidsvoerders:

1. in zekere mate in staat zijn om doelgerichte beïnvloeding van sociale systemen voor te bereiden en ook uit te voeren;
2. 'van te voren kennis kunnen verwerven over te verwachten effecten en neveneffecten van deze beïnvloedingsactiviteiten, welke kennis binnen zekere grenzen betrouwbaar is'. Blijkbaar bestaat er vertrouwen in de mogelijkheden tot voorspelling;
3. 'bereid zijn om van de resultaten van (...) analyses kennis te nemen om de uitkomsten van deze analyses te betrekken bij de besluitvorming' (zie Becker, 1987: 125).

Beleid wordt blijkbaar niet lastig doorkruist door machtspolitiek. Informatie is belangrijk.

Van de veertien waargenomen operationele toetsen zijn er drie als zeer bekend in de rijksdienst te kwalificeren: de wetgevingstoets, de MDW-toets en de ketenbenadering. 'Redelijk bekend' zijn de Tafel van Elf, de emancipatie-effectrapportage en 'uitvoeringsgericht werken' (Steenbeek, 1997: 36). De Tafel van Elf is beschreven in het tijdschrift *Beleidsanalyse*.

Bepaalde toetsen worden vaak gebruikt, andere veel minder. Hoe is de stand van zaken per medio 1997? Meer dan 100 maal per jaar worden gebruikt: de wetgevingstoets; uitvoeringsanalyse en toetsing; RO-toets.

De frequentie 20-100 maal scoort: MDW-toets; Tafel van Elf; tarieven-toets; handboek subsidies; checklist coördinatie bestuursbeleid; ketenbenadering.

Een frequentie van minder dan 20 maal per medio 1997 kent: handleiding subsidie-instrument; de emancipatie-effectrapportage; gezondheidseffectrapportage; uitvoeringsgericht werken; economische effectrapportage infrastructuur. Het gebruik is hoog als een toets ook formeel is vastgelegd.

Medewerkers die met toetsen werkten, vonden een toets over het algemeen *nuttig* (1997: 38). Men is vrij eensgezind over het nut.

Waarom kwam de toets er? Onderzoek toont dat beleidsambtenaren wel gewend zijn om beleid te maken maar minder benul hebben van de implementatie. Een toets moet de uitvoerbaarheid, handhaafbaarheid, doeleffecten of neveneffecten in beeld brengen. Het uiteindelijke doel van EAE's is om de kwaliteit van beleid te verbeteren.

Kraan-Jetten, Brand en De Valk (1998) publiceerden in het tijdschrift *Beleidsanalyse* een checklist van kwaliteitsnormen bij ex ante evaluaties.

Figuur 7: Ex ante evaluatie-toetsen op rijksniveau per medio 1997

Rubriek	Toets	Departement / context	In ontw. / in gebruik
A invloed op doeleffecten	1 handleiding subsidie-instrument	LN&V	operationeel
	2 checklist tarieven	V&W	operationeel
	3 checklist certificatie en normalisatie	V&W;relatie met marktwerking-project	
	4 handboek subsidies	V&W	operationeel
B de proceskwaliteit: a uitvoerbaarheid beleid b handhaafbaarheid van beleid	5 checklist voor uitvoeren en beoordelen van ex ante evaluaties	Financiën	
	6 handboek advisering en implementatie wetgeving	Financiën	
	7 wetgevingstoets	Justitie	operationeel
	8 handleiding effectmeting uitvoerbaarheid en handhaafbaarheid (MDW-toets)	Justitie	<i>operationeel</i> MDW-toets
	9 de tafel van elf: toets op mate van naleving door gedragswetenschappelijke analyse op 11 dimensies	Justitie	operationeel
	10 uitvoeringstoets rechterlijke organisatie	Justitie	operationeel
	11 ex ante uitvoeringstoets (EAUT)	Justitie	
	12 checklist handhaafbaarheid	Justitie	
	13 EU-checklist	Justitie	
	14 uitvoeringsanalyse en -toetsing	OC&W	operationeel
	15 'gegroeide toetsingspraktijk beleidsontwikkeling	OC&W	
	16 tienstappenplan beleid en wetgevingskwaliteit	SZW	
	17 ketenbenadering (checklist voor resultaatgericht beleid)	V&W	operationeel
C inhoud van het beleidsinstrument en neveneffecten	18 checklist bestuursbeleid	BiZa	operationeel
	19 veiligheidseffectrapportage	BiZa	
	20 bedrijfseffectentoets (MDW-toets)	EZ	<i>operationeel</i> MDW-toets
	21 emancipatie-effectrapportage	SZW	operationeel
	22 economische effectrapportage infrastructuur	V&W	operationeel
	23 verkeers veiligheids effectrapportage (VVER)	V&W	
	24 milieutoets (MDW-toets)	VROM	<i>operationeel</i> MDW-toets
	25 uitvoeringsgericht werken in vijf stappen (U&H-toets)	VROM	operationeel
	26 gezondheidseffectrapportage (GER)	VWS	operationeel
27 ramingsmodel zorgsector	VWS		
28 gehandicapten informatie doordenksysteem	VWS		

Bron: Steenbeek, 1997

16 EEN SUBTYPE VAN EX ANTE EVALUATIE: EFFECTRAPPORTAGES

Integrale ex ante evaluatie

Ex ante evaluaties impliceren het voorafgaand aan besluitvorming over een beleidsprogramma analyseren van de vraag of de beleidsdoelstellingen en middelen tot de gewenste uitvoering kunnen leiden, doelstellingen (dus beoogde effecten) wel bereikt kunnen worden en ongewenste neveneffecten uitblijven. Dat betekent dat de gehele beleidstheorie, dus het geheel van veronderstellingen, achter een beleidsvoornemen doorgelicht kan worden.

Dergelijke evaluaties worden verricht in geval van grote twijfel over de vraag of het beleid wel beantwoord aan het doel en/of de baten wel opwegen tegen de kosten en/of de kosten niet te groot worden en/of wel voldoende rekening is gehouden met reële knelpunten in de uitvoering.

Is er reden voor dergelijke twijfel? De literatuur geeft aan dat veel beleid anders uitpakt dan voorzien wordt en dat er dus reden kan zijn om vooraf al na te denken of het beleid wel gewenst is of vooraf nog aanpassing vergt. Er zijn dus heel vaak serieuze argumenten te geven voor de vraag of 'dit' beleid wel tot uitvoering moet komen.

Effectrapportages vormen een subtype: het zijn partiële beleidsevaluatie ex ante.

Willen we het verschijnsel effectrapportage doorgronden, dan is het nodig enige begrippen te gebruiken. Bij beleid kunnen we spreken van 'input', 'output' en 'outcomes'. Bij input gaat het om de middelen, bij output om de prestaties, en bij outcomes om de effecten van een beleid. Als we een tunnel onder een weg aanleggen om de verkeersveiligheid te verbeteren, is de tunnel de (beleids)prestatie en de verbeterde verkeersveiligheid het beoogde doel. Is er daadwerkelijk sprake van een verbeterde verkeersveiligheid door de tunnelaanleg dan treedt de gewenste outcome op, het effect waar het allemaal om begonnen is.

Een effectrapportage is in dit voorbeeld goed voorstelbaar. Een verkeersveiligheids-effectrapportage kan, vóór een besluitvormend orgaan besluit tot een tunnel of twee tunnels of tot een andere oplossing, of wellicht tot niets, in kaart brengen wat we mogen verwachten van het alternatief, of het neveneffecten heeft, bijvoorbeeld sociale onveiligheid, en of een alternatief eigenlijk wel gewenst is.

Effectrapportage, wat het is

Een effectrapportage is een manier van ex ante evaluatie (verder: Korsten, 1997). In een effectrapportage wordt vooraf, vóór de besluitvorming plaatsvindt, gekeken naar effecten van een voorgenomen beleidsmaatregel of naar effecten van enkele alternatieve maatregelen. Dat gebeurt niet zonder reden natuurlijk. De ervaring leert dat het bereiken van wat een overheid wenst, lang niet altijd plaatsvindt. Omdat gewenste effecten niet altijd volledig of in voldoende mate bereikt worden, wordt niet volstaan met evaluatie achteraf, maar proberen overheden al *vooraf* te kijken of effecten te bereiken zijn. Ex ante evaluatie lijkt dus voor de hand liggend als effecten twijfelachtig zijn.

Effectrapportages zijn weliswaar goed voorstelbaar maar niet eenvoudig uit te voeren, zoals nog zal blijken. Ze komen ook daadwerkelijk in allerlei gedaanten bij de Nederlandse overheid voor. Tot een van de bekendste effectrapportages behoort de milieu-effectrapportage, de mer. Bij de mer is het niet gebleven. Er zijn sinds pakweg 1980 steeds vaker effectrapportages voorgesteld. Ze intrigeren blinkbaar. Volgens sommigen roept de ene rapportage zelfs de andere op, waardoor men van een *wedloop* kan spreken.

Vragen rond effectrapportages

Aan effectrapportages zijn allerlei andere aspecten te onderscheiden (Korsten, 1997). Dat bleek al uit het voorgaande. We noemen er nog enkele:

- effectrapportages, wat zijn het eigenlijk en welke typen bestaan?;
- waarom komen ze voor?;
- waar komen ze voor, op welk overheidsniveau?;
- komen ze alle tot wasdom?;
- hoe verrichten we effectstudies?;
- waar loopt de wapenwedloop van steeds meer rapportages op uit?;
- welke gevolgen hebben de rapportages op zich?

Op enkele aspecten gaan we kort nader in.

Er bestaan meer soorten effectrapportages dan de mer. Zo komen er hier en daar wel *jeugdeffectrapportages* en *emancipatie-effectrapportages* voor.

Effectrapportages zijn om allerlei redenen een intrigerend verschijnsel. Waarom? Belangengroepen zien er een manier in om de reflectie op hun belang te verankeren en hun belang dus steeds weer op de agenda te krijgen. Milieugroepen bepleiten mede om die reden een milieu-effectrapportage.

Hoewel het uitvoeren van dergelijke effectrapportages lastig is, worden er sinds 1980 steeds nieuwe effectrapportages bepleit. Een van de kabinetten Lubbers pleitte voor een *bedrijventoets* en later viel zelfs het woord '*Europa-toets*'.

Het al wat oudere '*Draaiboek voor de wetgeving*' laat zien dat bij wetgeving in het bijzonder gelet wordt op effecten voor het particuliere bedrijfsleven, de sociaal-economische ontwikkeling en de marktwerking. Dat is grotendeels het gevolg van een bezinning op marktwerking, deregulering en wetgevingskwaliteit. Volgens het draaiboek moet ook gelet worden op de gevolgen voor het milieu, en op de uitvoerbaarheid en handhaafbaarheid.

Oog voor effecten niet vanzelfsprekend?

Als we speciale effectrapportages verrichten, is het blijkbaar niet vanzelfsprekend om bij reguliere beleidsontwikkeling naar effecten te kijken. Dat is bij nader inzien ook zo.

Jaren terug werd bij overheidsbeleid zeker niet zo op effecten gelet. We waren in Nederland optimistischer over het halen van effecten. In een tijd van wederopbouw was woningbouw belangrijk, en uiteraard ook het halen van de bouwvolumes. De vorderingen in de bouw waren eenvoudig te berekenen. Het budget voor beleid was in die tijd in het algemeen meestal wel beschikbaar en als een beleid niet direct het beoogde effect had, werden er meer middelen in gepompt. Het beleid was in de periode 1945-1960 ook nog niet zo breed en pretentievool 'modificerend', dus niet zo op sociale gedragsverandering gericht. Er waren nog niet op brede schaal oproepen om zuiniger te stoken, het milieu niet te schaden door vervuiling, een mestboekhouding te voeren, auto's verplicht te laten keuren, enz. Het beleid was eerder 'codificerend'.

Maar intussen is beleid verbreed naar vele terreinen, is de modificerende pretentie verbreed en is afweging meer gewenst vanwege de beperking van de financiële middelen, die we menen te zien of die overheidsbestuurders zich opleggen of opgelegd krijgen. De door toenemende opleiding mondiger wordende samenleving gelooft intussen ook minder in ficties, in beleidsmatig 'bellen blazen' rond werkgelegenheid scheppende plannen die aan veel te hoog gegrepen ambitie ten gronde gaan, bijvoorbeeld. Dat wil zeggen, dat meer gevraagd wordt (en

soms moet worden) naar beleid dat ook daadwerkelijk het beoogde effect heeft, anders kan het beter achterwege blijven.

In zo'n verschuivende context kan gevraagd worden naar milieu-effectrapportages en de effecten van beleid voor emancipatie. Er kan een roep om effectrapportages ontstaan.

Vraagstelling in een effectrapportage

Effectrapportages zijn een vorm van ex ante evaluatie: het vooraf, vóór de besluitvorming plaatsvindt, bekijken van bepaalde effecten van een beleidsvoorstel of enkele voorstellen. We schreven dit al. Effectrapportages zijn gericht op vragen als:

- welke zijn de doelstellingen achter een maatregel en wat de gewenste of ongewenste effecten van een eventuele maatregel?;
- is een maatregel uitvoerbaar, wat kost het en draagt die bij aan het realiseren van de doelstellingen?;
- sporen uitvoering, kosten en baten, en effecten en neveneffecten met de verwachtingen?;

Bij een effectrapportage wordt gerapporteerd over beoogde en soms over niet-beoogde effecten, dus neveneffecten.

Sommige effectrapportages zijn geïnstitutionaliseerd in regelgeving, andere niet. Zo zijn er voor de milieu-effectrapportage spelregels in wetgeving opgenomen, maar voor de economische effectrapportage niet.

De meeste effectrapportages zijn *sectoraal*. *Eén effectrapportage is integraal*. Die staat in de 'Aanwijzingen voor de regelgeving'. Deze is het gevolg van werk van de commissie-Geelhoed, over deregulering. Volgens die aanwijzingen moet het bereiken van effecten van regelgeving tevoren in kaart zijn gebracht en moet nagegaan worden of wetgeving wel het geëigende middel is om de doelstellingen te bereiken.

Onder 'integraal' versta ik bij effectrapportages dat alle sectoren die ermee te maken hebben evenwichtig behandeld worden. In een integrale effectrapportage worden de verschillende belangen die door de voorgenomen activiteit beïnvloed kunnen worden, in één rapportage betrokken.

In verband met effectrapportages wordt ook wel eens het woord 'consequentie-analyse' gebruikt (Becker, 1983). Is dat terecht? Niet helemaal want een consequentie-analyse is een inventarisatie van effecten terwijl een effectrapportage nadrukkelijk ook een beoordeling of waardering van de gevolgen van een maatregel inhoudt.

Historie van effectrapportages

Effectrapportages komen voor in het buitenland, bijvoorbeeld in de Verenigde Staten. Ook in Duitsland komt dit fenomeen voor, bijvoorbeeld in de vorm van 'Der Praxistest von Gesetzentwürfen'. In Nederland kennen we effectrapportages ook al langer. Denk onder meer aan

- de milieu-effectrapportage (mer),
- de bestuurseffectrapportage (ber),
- de emancipatie-effectrapportage (eer),
- de veiligheidseffectrapportage (ver),
- de economische effectrapportage,
- de gezondheidseffectrapportage (ger),
- de criminaliteitseffectrapportage (cer).

Opvallend is dat op deze lijst van voor 2000 een *generatie-effectrapportage* nog ontbreekt. Blijkbaar wordt er niet nagegaan welke problemen, bijvoorbeeld de schuldenlast, doorgeschoven worden naar toekomstige generaties. Een aantal jaren terug werd door een raadslid uit de gemeente Venlo gepleit voor een *jeugdeffectrapportage*.

Een ander nieuw idee is een *archeologie-effectrapportage*. Bij de aanleg van industrieterreinen en grote andere infrastructurele ingrepen kan men een dergelijke effectrapportage verplicht stellen. Nuttig of niet? Waar? Onder welke condities? Welke argumenten voor en welke tegen?

Sommige effectrapportages zijn bekender dan andere. Sectorale effectrapportages zijn veel bekender dan integrale. Het meest bekend is de (sectorale) milieu-effectrapportage geworden.

Wie heeft er in 1978 gedacht aan een verkeerseffectrapportage? Toch kwam die toen ook al voor. Zo ik zelf als medewerker van het Nijmeegse Instituut voor Toegepaste Sociologie in opdracht van de Directie Gelderland van Rijkswaterstaat een onderzoek uitgevoerd naar de vraag welk effect de aanleg van de A73 had op de recreatiedruk van het Hatertse en Overasseltse Vennengebied bij Nijmegen. Dat was, hoewel het nooit officieel werd uitgesproken, een onderzoek om milieu- en recreatiemensen die zich verzetten tegen de weg en bepaalde kunstwerken daarin, tunnels, wind uit de zeilen te halen.

Niet alle typen effectrapportages zijn geïnstitutionaliseerd maar de bekendste wel. Met de mer is dat het geval. Bij de ber is het niet zover gekomen. Met de ber, de bestuurseffectrapportage, is het nooit wat geworden. Of nieuwe effectrapportages zullen 'doorbreken' moet worden afgewacht.

Oorsprong van effectrapportages in macroperspectief

Waar komen rapportages uit voort?

Een eerste reden is ongetwijfeld bezorgdheid dat gewenste effecten van beleid niet gehaald worden of er neveneffecten optreden. Dat blijkt bijvoorbeeld naar voren te komen uit effectrapportages over de uitbreiding van vliegvelden. Wat de werkgelegenheidseffecten van een uitbreiding zijn, blijkt tot veel verschil van mening te leiden (zie ook de dissertatie van Kramer). Wilt u meer voorbeelden lees dan C. Schuyt's boekje *'Tegendraadse werkingen'*.

Een tweede reden voor effectrapportages is de noodzaak van nauwgezetere afweging van belangen en het dreigend in de knel komen van bepaalde belangen. Als het milieubelang verwaarloosd dreigt te worden, wordt een effectrapportage vooraf bepleit. Als bij voorbeeld de benzineprijs in Nederland sterk stijgt ten opzichte van het buitenland is dit voor pomphouders in de grensgebieden goed voelbaar en wordt gepleit voor een *grenslandeffectrapportage*. De bedoeling is te bereiken dat de rijksoverheid zich voorafgaand aan het nemen van maatregelen meer bewust wordt van wat die maatregelen voor een grensgebied kunnen betekenen. Minister van Buitenlandse Zaken Frans Timmermans (woonplaats Heerlen) besloot in 2013 om zo een toepassing te verkennen.

Dan zijn er ten derde de belangengroepen die pleiten voor een rapportage. De emancipatieraad bepleitte een emancipatie-effectrapportage. Effectrapportages zijn in de belangengroependemocratie verzeild geraakt. Er zijn intussen allerlei rapportages. Elk belang heeft zijn rapportage: ber, mer, ver, eer, enz.. De ene rapportage lokt ook de andere uit. Er lijkt sprake van een soort wapenwedloop aan de gang. We zagen dit bij de discussie over de uitbreiding van Schiphol in '95-'96.

d Ook vanuit de catacomben van de rijksdienst komen we geestdriftige pleidooien tegen voor nieuwe effectrapportages. Blijkbaar is een effectrapportage een momenteel in de tijdgeest passend instrument dat aanspreekt omdat het nader zicht kan bieden op gewenste richtingen die met beleid worden ingeslagen. Wat leert ons een *wapenbeheersingseffectrapportage*? Welke kant gaat het op als een onderdeel van het gezondheidszorgbeleid die en die effecten heeft?

Effectrapportages vertegenwoordigen, zo lijkt het, elk een belang. De bestuurseffectrapportage komt voort uit de Raad voor het Binnenlands Bestuur (Rbb), die, ik vertaal vrij, vóór de besluitvorming en invoering van maatregelen plaatsvond, uitgezocht wilde zien wat de consequenties van rijksmaatregelen zijn voor medeoverheden. De emancipatie-effectrapportage is bijvoorbeeld bedacht door de Emancipatieraad, ook een adviesraad. En de MER wordt geassocieerd met milieugroepen.

Effectrapportages in meso- en micro-perspectief

Dat effectrapportages nodig zijn, kan men beoordelen vanuit macro-perspectief door te kijken naar verschuivingen in overheidsbeleid. Daarop gingen we reeds in. Men kan ook kijken naar de meso- en micro-verklaringen, dus bij de vraag naar een effectrapportage door belangengroepen en bij een concrete roep om een rapportage. Dat brengt ons bij vier factoren, vier verklaringen.

a. Meer willen weten door meten

Het bepleiten van effectrapportages kan te maken hebben met kenmerken van het beleid. De effecten van bepaald beleid worden eerst na jaren zichtbaar. Doel van de effectrapportage is dan om een bepaald effect van beleid *eerder* in beeld te krijgen om te zien wat de kans op het bereiken van een gewenst of ongewenst effect is. Een voorbeeld vormen emancipatiebevorderende maatregelen. Effecten van emancipatiebeleid zijn meestal inderdaad niet binnen een-twee jaar zichtbaar. Het is begrijpelijk dat daarom zo'n rapportage wordt gevraagd. De Emancipatieraad heeft daarbij ongetwijfeld ook de versterking van de eigen positie op het oog gehad. Daarin zijn diverse kabinetten echter niet meegegaan. Vandaar dat er niet zoveel van de emancipatie-effectrapportage terecht gekomen is.

b. Nieuw belang zien

Een tweede verklaring voor het ontstaan van effectrapportages is dat in de belangenafweging nieuwe belangen naar voren gebracht worden, die men *preciezer* in beeld wil krijgen. Een toenemend belang voor veiligheidsaspecten lokt zo een veiligheidseffectrapportage uit.

c. Reactie op eerder weggedrukt belang

Een derde reden voor de opmars kan zijn: het tegengaan van het wegdrukken van een bepaald belang. De mer is daar een voorbeeld van. Milieugroepen wilden een middel tegen het wegdrukken van 'het milieu' in de belangenafweging.

Om effectrapportages wordt natuurlijk gevraagd als mensen denken dat een gewenst belang *verwaarloosd* wordt in de belangenafweging bij een beleidsvoornemen, in de ontwerpfasen van beleid. Verdedigers van de effectrapportage denken dat de effecten van het realiseren van het beleidsvoornemen negatief, of te negatief, of niet positief genoeg zullen zijn. Ze rekenen erop dat een effectrapportage dit tijdig, voor het nemen van beslissingen aan het licht brengt. Een effectrapportage is een (quasi-) *neutraal* instrument, een instrument dat het belang als het ware uit de hoek van de belangengroep tilt. Verdedigers verwachten van een effectrapportage over een mogelijk beleidsvoornemen a dat als blijkt dat het belang x in de knel komt, alsnog een aanpassing van het voornemen a plaatsvindt.

d. *De ene rapportage lokt de andere uit*

Een vierde reden is dat de ene effectrapportage *de andere uitlokt*, als een soort tegengas. Een voorbeeld is de economische effectrapportage rond een vliegveld, die het gemeentebestuur van Rotterdam eind jaren tachtig vroeg bij de discussie over vliegveld Zestienhoven.

Effectrapportages uitvoeren

Effectrapportages kennen methodologische aspecten. Effecten vooraf in kaart brengen is betrekkelijk lastig want het voorspellen van effecten, zeker als ze betrekking hebben op menselijk gedrag of gedrag van organisaties, is niet eenvoudig. Uit wetsevaluatie-onderzoek blijkt dat ook.

Gezondheidseffectscreening

Vanuit het Ministerie van Volksgezondheid, Welzijn en Sport is in 1996 een inventarisatie gemaakt van bestaande methoden en instrumenten voor effecttoetsing van beleidsvoornemens. In het rapport '*Gezondheidseffectscreening*' is een overzicht te vinden. Daarin worden genoemd: de milieu-effectrapportage, de emancipatie-effectrapportage, de bestuurs-effectrapportage, de veiligheidseffectrapportage, de criminaliteitseffectrapportage, de wetgevingstoets en de MDW-operatie (marktwerking, deregulering en wetgevingskwaliteit). De vergelijking wijst uit dat er twee typen bestaan.

De eerste soort is gebaseerd op de veronderstelling dat oorzaak-gevolgrelaties bekend en te kwantificeren zijn. Volgens een vast stappenplan kan dan berekend worden wat de gevolgen van alternatieve beleidsopties zijn. Dit heet in het rapport het rationeel model. Op milieugebied blijkt dit frequent aan te treffen.

Achter de tweede soort gaat de veronderstelling schuil dat oorzaak-gevolgrelaties niet goed bekend en te kwantificeren zijn. Bij dit type wordt via onderhandeling en overleg tussen de betrokken actoren bezien welke beleids optie naar verwachting de gunstigste effecten heeft. Omdat de oorzaak-gevolgrelaties niet bekend zijn is de procedure met nogal wat onzekerheden omgeven. Maar ze hoeft niet minder effectief te zien. Dit type is meer incrementeel.

Een toetsing van beleidsvoorstellen op gezondheidseffecten is volgens de minister van VWS in '*Volkgezondheidsbeleid 1995-1998*' mogelijk, ook al zijn veel oorzaak-gevolgrelaties 'niet altijd volledig bekend'. Voor beleidsinitiatieven waarbij veel epidemiologische kennis beschikbaar is, kan het rationele model van effectrapportage wel van toepassing zijn. Vraagstellingen waarbij dergelijke kennis afwezig is, zijn via een incrementeel model te benaderen, dus via overleg en onderhandeling. Veel procedures kunnen elementen van beide hebben.

De toetsing van gezondheidseffecten is niet geheel nieuw omdat op het gebied van milieu en verkeersveiligheid effecten op de gezondheid al langer in kaart worden gebracht en bij de besluitvorming betrokken worden. Gezondheidseffectscreening heeft 'als doel om een dergelijke gezondheidstoets op een meer systematische en tegelijkertijd meer selectieve wijze in te zetten' (TK 1995-1996, 24 126, nr. 14: 2). De minister besluit om het instrumentarium voor deze toets te laten uitwerken en experimenteel te laten toepassen. De minister van VWS, mw. E. Borst-Eilers, beschouwt zichzelf als initiatiefnemer van de toets.

Europatoets en grenslandtoets

Europarlementariër Ria Oomen (CDA) bepleitte in mei 1998 en ook al daarvoor een *Europatoets*. Deze zou een nationaal wetsontwerp op het gebied van belastingwijziging, sociale zekerheid en dergelijke moeten vergezellen. De bedoeling was de effecten voor 'grensoverschrijdende' pendelaars (uit het grensgebied) duidelijk te maken. Zo'n Europatoets

zou niet alleen in Nederland door de Nederlandse wetgever maar ook in andere landen uit de Europese Unie moeten worden uitgevoerd.

Een variant op de Europatoets is de *grenslandtoets*. De grenslandtoets beoogt negatieve effecten van voorstellen op het gebied van nationale regelgeving voor grensregio's vóór besluitvorming plaatsvindt op te sporen: een *grensland-effectrapportage* dus.

De Grondwet schiet niet tekort met mogelijkheden voor grensregio's voor een eigen 'buitenlands beleid'. Invoering van een *grenslandtoets* bij nationale wetgeving heeft geen zin, omdat die te vertragend werkt. Dat antwoorden minister Dijkstal en staatssecretaris Van de Vondevoort van Binnenlandse Zaken medio februari 1998 op vragen van de D66-ers Wessels en Scheltema over het nieuwjaarsessay '*Limburg en de Grondwet*' van commissaris der koningin Van Voorst tot Voorst.

De Limburgse gouverneur schreef daarin dat de huidige staatkundige structuur, zoals vastgesteld in de Grondwet, grensregio's te weinig bevoegdheden geeft om 'kleine buitenlandse zaken' zelf te regelen. Ook meende hij dat los van de Grondwet politiek Den Haag zijn houding ten opzichte van grensregio's moet veranderen, meer in het bijzonder verruimen.

Dijkstal c.s. vinden niet dat de Grondwet grensregio's te weinig ruimte biedt. Volgens hen geeft Van Voorst tot Voorst daar ook geen concrete voorbeelden van en bepleit hij ook helemaal geen wijziging van de Grondwet. Volgens Van Voorst tot Voorst is die immers niet nodig 'wanneer de rijksoverheid voldoende aandacht heeft voor buitenlandse zaken met de directe burens en indien de wetgever voldoende ruimte geeft aan decentrale overheden'.

De minister en staatssecretaris vinden dat de bepleite grenslandtoets voor nationale wetgeving, die bedoeld is om negatieve effecten in grensregio's te voorkomen, alleen maar vertragend zal werken. Bovendien zijn negatieve effecten pas in de praktijk merkbaar en niet vooraf.

De bewindslieden zijn ten slotte niet van plan om eventuele knelpunten bij grensoverschrijdende samenwerking op een rijtje te zetten. Gezien de vele verschillen tussen Nederlandse, Duitse en Belgische wetgeving is dit onbegonnen werk. De meeste klachten betreffen ruimtelijke ordening, milieu, sociale wetgeving en media/telecommunicatie. Afstemming van wetgeving is 'een proces van lange adem'. In de praktijk komt men vaak tot 'bevredigende oplossingen', aldus de bewindslieden.

De bewindslieden overschatten dit wellicht. De Euregio Maas Rijn bijvoorbeeld is rond 2010 een lamme eend. De Euregio Maas Rijn was bedoeld als denktank. Ideeën moest die spuien, zodat de bestuurders worden gevoed door het veld. Het dagelijks bestuur van deze Euregio wordt gevormd door de Limburgse commissaris der koningin en zijn buitenlandse tegenhangers.

De grensregio's gijzelen ook zichzelf. De Euregioraad Maas Rijn had te kampen met te weinig aangedragen agendapunten en de opkomst op vergaderingen was bedroevend. De belangstelling voor de raad is gering. De denktankfunctie kon niet waargemaakt worden. Ideeën: rond de nul komma nul. Daar kwam bij dat brainstormen mocht, maar besluiten nemen, ho maar. Bestuurders laten het dan al gauw afweten. Geen wonder dat er geen kip naar een vergadering komt. Zijn er dan geen knelpunten? Die zijn er wel maar wie ze aandraagt, moet er ook rekening mee houden dat hij ze gezamenlijk met anderen moet oplossen en budget moet leveren. Dan durven lokale en provinciebestuurders aan alle zijden van de grens niet. De Euregio had geen (groot) eigen budget.

De commissaris heeft met zijn nieuwjaarsessay niet alleen het hart gelucht maar ook eens van zich afgebeten, zonder tanden overigens, richting een derde: de nationale boeman.

De Europatoets en grenslandtoets komen samen in het volgende. Steunenberg e.a. (2012) hebben de mogelijkheden in kaart gebracht voor het uitvoeren van een *grenseffectenanalyse* voordat een EU-richtlijn wordt omgezet in nationale wet- en regelgeving. Zij verkenden ook of het mogelijk is te monitoren of tijdens en na omzetting grenseffecten zich alsnog voordoen en te bezien hoe deze opgelost kunnen worden. Een dergelijke analyse moet helpen voorkomen dat er in de toekomst ongewenste grenseffecten ontstaan. Minister van Buitenlandse Zaken Timmermans gaat er, zoals gesteld, naar kijken. Dat zegde hij in 2013 toe.

Waardering van effectrapportages

Om tot een waardering van effectrapportages te komen, kan men verschillende aspecten onderscheiden.

a De wapenwedloop.

Sommigen signaleren een wildgroei van effectrapportages, omdat de ene rapportage de andere uitlokt.

De uitbreiding van Schiphol illustreert dit. De mer voor de vijfde landingsbaan kon de toets der kritiek niet doorstaan, en daarom was een aanvullende mer nodig. De aanvullende mer was voor de voorstanders van de baan weer aanleiding om een economische effectrapportage te laten opstellen.

We zien een schakeling van effectrapportages, een proliferatie, een soort wapenwedloop. Meer integraliteit in ex ante evaluatie kan de versnippering van effectrapportages wellicht stoppen.

b Effectrapportages lokken nadere beoordeling uit.

Er is in het geval van Schiphol een zware wissel getrokken op het integratievermogen van de overheid. Maar niet alleen dat. Het parlement heeft in geval van Schiphol zelf een secundaire analyse van de effectrapportages laten uitvoeren. Rond de Betuweroute is er ook al contra-expertise ingeroepen.

Er is blijkbaar behoefte aan nadere weging van effectrapportages. Er wordt iemand voorgeschakeld, wat het gevaar geeft van toenemende stroperigheid.

c Wat is de technische kwaliteit van effectrapportages?

Hierover is een onderzoek gepubliceerd door De Valk en De Vries (1994). Daaruit blijkt dat diverse mer's kwalitatief goed zijn maar hier en daar vallen ook wel steekjes. De Valk (1997) meent dat er soms wel veel teveel alternatieven worden doorgerekend waardoor door de bomen het bos niet meer is te zien.

d Hebben effectrapportages invloed?

Is het een veelbelovend instrument? De mer is uit het perspectief van 'beleidsgericht leren' positief (Ten Heuvelhof en Nauta, 1997). Veel gebruikers melden dat ze in allerlei opzichten wat gehad hebben aan een mer-rapport. Het heeft hun inzicht verruimd, en hun begrip voor diverse bezwaren tegen een alternatief vergroot. De Valk (1997) analyseerde 41 besluiten waarbij een mer werd gemaakt en vergeleek die met 41 besluiten zonder mer. Hij is minder positief dan Ten Heuvelhof en Nauta. Vooral bij overheden zelf, die een mer laten verrichten werkt de MER door. Andere initiatiefnemers laten zich er minder aan gelegen liggen.

DEEL VI: LEREN VAN EVALUEREN



EVALUEREN = SAMEN LEREN



17 DE EVALUATOREN: WAT ZE KUNNEN LEREN VAN KRITIEK

Is er ook kritiek mogelijk op evaluatieonderzoek waar evaluatoren wat van kunnen leren? Daarop moet worden ingegaan omdat als evaluatoren iets kunnen leren dat weer positief kan inwerken op het 'leren van evalueren' in organisaties.

Vooreerst, waarop komt weinig of geen kritiek? Op enkele belangrijke aspecten van evaluatie.

1. Opvallend is dat de *onafhankelijkheid* van de evaluator zelden kritiek krijgt. Onafhankelijkheid is overigens wel belangrijk, ook als een visitatiecommissie aantreedt of een parlementaire (evaluatie)commissie. Leden van een parlementaire commissie moeten hun lidmaatschap van een bepaalde partij als het ware terzijde schuiven en ze moeten gaande het onderzoek spelregels in acht nemen, zoals niet lekken (Van der Zwan en Noordegraaf, 1998:71).
2. De *onderzoeksaanpak* moet ten allen tijd helder gestructureerd zijn. De keuze van de opzet is belangrijk omdat dan moet blijken om welk beleid het gaat: om beleid in een een-actor situatie of in een multi-actorennetwerk?

Waarop wel kritiek?

1 Te weinig bezinning op de opdracht

De eerste kritische waarneming behelst dat onderzoekers *vaak (te) laat bij de evaluatieopdracht betrokken worden*. Deze ervaring baseer ik op tal van opdrachten die ik zelf heb uitgevoerd en kennisname van studies van anderen.

Waarom? Het proces van opdrachtverlening is vaak sequentieel. Het besluit tot evaluatie is gevallen, er zijn contouren aangegeven over het wat en waarom en hoe van het evaluatieonderzoek en het budget ligt vast. Dan komt de evaluator pas aan bod. Als een externe evaluator gezocht wordt, mag die dan met medestrevers een offerte uitbrengen? Met die offerte gaat hij 'teruggeven' (herformuleren in eigen woorden) aan de opdrachtgever hoe het onderzoek wordt opgevat. Dat is een rem op durf en creativiteit. Juist in de startfase zou eens vrijmoedig gesproken moeten worden, ook aan de kant van de opdrachtverleners over vragen: wat willen we nou echt met het onderzoek?; wat niet?; wat moet het voor soort informatie opleveren? De entreebesprekingen tussen opdrachtgever en onderzoeker(s) en eventueel ook de betrokkenen uit het beleidsnetwerk zijn van groot belang. Dit wordt vaak onderschat.

2 Begrip van doelbereiking schiet tekort

Een tweede kritiek raakt de verklaring van output en/of effecten (outcomes).

Tal van evaluaties kijken naar de bereikte prestaties of resultaten, zelfs naar effecten en naar de processen ernaar toe. Maar ze 'leggen geen relatie met de daarvoor verantwoordelijke mechanismen'. Ze gaan niet terug naar het 'waarom' van de effecten.

'De achterliggende beleids- en interventietheorie(en) worden niet geïncorporeerd, gereconstrueerd noch getoetst'

aldus Leeuw en Van Gils (2010: 170) op basis van hun analyse van een steekproef van 30 evaluaties uit een totale populatie van 100 op het terrein van het ministerie van Buitenlandse Zaken. Er ontbreekt iets wat lijkt op – streng gesteld – een *oorzakelijkheidsanalyse*. Als beleid effect heeft of geen effect moet duidelijk worden of dat ligt aan de beleidsinhoud, de in- of

uitvoering of aan iets anders. En als beleid de oorzaak is, hoe kan het dan gebeuren dat de doelstellingen niet gehaald worden? Die vraag moet beantwoord worden.

De implicatie daarvan is duidelijk. De onderzoekers leren de lezers nogal eens te *weinig* over *het waarom* van het wel of niet bereiken van bepaalde resultaten. Wat nog wel gebeurt, is dat aangegeven wordt of de bereikte resultaten en effecten eventueel overeenstemmen met de beleidsintenties en afspraken.

Evaluaties volstaan vaak met de doelbereiking. Weinig gebruiksgerichte evaluatiestudies zijn *echte* effectiviteitsanalyses. Waarom? Omdat de onderzoekers niet gevraagd wordt te zoeken naar de 'mechanismen', omdat het budget voor de analyse beperkt is en omdat er altijd weer haast is. Onderzoekers worden in feite nogal eens te veel opgejaagd. Dat betekent anders gesteld dat de opdrachtgevers niet echt interesse hebben voor effectiviteitsanalyses.

Veel analyses zijn bovendien slechts procesevaluaties.

Deze aanpak van doelbereikingsgerichtheid heeft gevolgen voor de beleidsrelevantie. Onderzoekers die een bestaande beleidstheorie niet ter discussie kunnen of mogen stellen en ook niet zoeken naar een alternatief of een dynamisering van de beleidstheorie, bereiken dat de beleidvoerders padafhankelijk blijven opereren. Ze blijven denken in termen van het beleid als potentieel succesvol terwijl het beleid misschien rigoreus op de schop moet (zie ook Van Twist, 1993; Heffen, 2008).

Een voorbeeld. Een provinciebestuur zette een omgevingsbeleid op dat moest leiden tot het voldoen aan milieunormen door agrarisch ondernemers. De stankoverlast van bedrijven van agrarische ondernemers moest verminderen en de ammoniakcijfers (bij voorbeeld als gevolg van varkenshouderij) moesten omlaag. Dat leidde tot een provinciaal voornemen om voor bepaalde landbouwbedrijven geen bedrijfsuitbreiding toe te staan. En andere bedrijven zouden moeten worden verplaatst. Daarvoor was een bepaald budget beschikbaar. Na enkele jaren bleek dat amper een varkenshouderij verplaatst was naar een gebied waar de ondernemer binnen aanvaardbare stanknormen bleef. Het beleid was dus volkomen vastgelopen. De doelstellingen werden niet bereikt. Waarom? De centrale veronderstelling achter het verplaatsingsbeleid bleek niet juist. De overgrote meerderheid van agrarische ondernemers was niet bereid om opnieuw fors in een elders gevestigd bedrijf te investeren. De meesten die in aanmerking kwamen, waren al ouder dan 50 jaar en hadden bovendien geen bedrijfsopvolger. Dus zouden de schulden bij een bedrijfsverplaatsing tot ver na hun 65^{ste} levensjaar om hun nek blijven hangen. Ze redeneerden: 'we zitten het wel uit op de bestaande lokatie, ook als geen verdere uitbreiding meer mogelijk is'. Eigenlijk zou in dit geval het roer fors om moeten: een herziening van de achterliggende beleidstheorie diende zich aan. Kwam die er ook? Nee. De evaluator wachtte af of het provinciebestuur het beleid wil stoppen.

3 De keuze van de centrale maatstaf

Een derde kritiekpunt slaat op het ontkennen van een beleidsnetwerk.

De keuze voor een effectiviteitsevaluatie of een zwakkere variant daarvan impliceert de keuze van doelstellingen die achter beleid schuil gaan van *een centrale overheidsactor*. Dat is een gerechtvaardigde aanpak in geval deze overheid *volledige* doorzettingsmacht heeft. Dan stelt de centrale actor beleid vast en voert ook zelf uit. Bij veel beleid is echter de doorzettingsmacht niet slechts in één hand; er is sprake van *afhankelijkheden* tussen

overheden en misschien ook wel in bredere zin (Abma en In 't Veld, 2001; Bekkers en Lips, 1998).

Neem het landschapsbehoud in Zuid-Limburg. Op een conferentie bleek dat heel veel partijen bij elkaar waren met elk eigen taken, bevoegdheden en middelen; de rijksoverheid, het waterschap, de provincie, de dienst landelijk gebied, de gemeenten, landbouwers, recreatieondernemers (pensionhouders, campinghouders, eigenaar van een skipiste, etc.), chemische bedrijven, de regionale VVS, enzovoorts. In dit geval moeten doelen verknoopt worden en middelen gebundeld. We zitten daarmee bij netwerkmanagement.

Verdedigers van netwerkmanagement noemen het uitgaan van gerichte doelstellingen van één overheidsactor een eenzijdige en verkeerde *top-down* benadering (Klijn en Koppenjan, 2001: 189). In een netwerkcontext moet juist worden uitgegaan van een benadering *van onderop*, door de denkwijzen en oriëntaties van *alle* relevante actoren uit het beleidsnetwerk mee te nemen.

Wat levert dat op? Erkend moet worden dat zich in het multi-actoren netwerk vaak dynamiek voordoet. Dat doelstellingen stabiel blijven is een misverstand, stellen zij (Klijn en Koppenjan, 2001). Er is sprake van: 'de onhoudbaarheid van een eenduidig inhoudelijk ijkpunt'. Bij een evaluatie van beleid dat in een netwerkcontext van van elkaar afhankelijke actoren tot stand kwam, denk maar eens aan de evaluatie van de vernattingspolitiek van De Peel, moeten de doelstellingen van alle actoren meegenomen worden. Al de actoren moeten in feite bij de opdracht tot evaluatie betrokken worden. Zij moeten immers ook de resultaten relevant vinden en benutten.

4 De reikwijdte van het onderzoek

Een vierde kritiekpunt: veel evaluatiestudies hebben een *beperkte geldigheid*. De bevindingen slaan alleen maar op wat is onderzocht en zijn ook begrensd in de tijd. Dat valt in elk geval op te maken uit een studie naar evaluaties op het terrein van het ministerie van Buitenlandse Zaken (Leeuw en Van Gils, 2010: 170). Deze evaluaties zeggen bij wijze van spreken iets over het beleidsprogramma in jaar 1 en eventueel 1 en 2 in de plaatsen a tot en met h. Maar ze hebben niet veel reikwijdte naar jaar 3 en de plaatsen i, j en k. Van generaliseerbaarheid is niet of nauwelijks sprake.

Dat kan te rechtvaardigen zijn omdat *gebruiksgerichte evaluatie* niet tot doel heeft om de kennis in een wetenschapsgebied uit te breiden maar om bij te dragen aan beter beleid en/of betere besluitvorming. Toch zou meer generaliseerbaarheid hier en daar mogelijk zijn. Denk maar eens aan gebiedsgericht beleid. Het is wel zo nuttig als een analyse van de politiek in De Biesbosch-Noordwaard, Kerngebied Vechtplassen en de Gelderse Poort inzichten oplevert die ook elders van nut zijn (zie Cörvers, Glasbergen en Korsten, 2003).

Zoeken naar *synthese* dient zich als opportuniteit aan. Want het is vaak niet zo dat wat overheid x laat onderzoeken zo uniek is dat een andere overheid y dit onderwerp nog nooit bij de kop heeft gepakt. Wie kijkt naar studies van lokale rekenkamers en rekenkamercommissies ontdekt dat veel van deze organisaties in een bestek van een jaar of vijf een aantal zelfde of sterk vergelijkbare studies ondernemen. Blijkbaar kijkt men bij het opvoeren van mogelijke onderzoekthema's naar wat andere rekenkamers onderzochten.

5 Veldkennis te gering?

Uit het voorgaande vloeit voort dat veel evaluatieonderzoekers wel wat weten van onderzoek verrichten maar gemiddeld genomen zelf *niet uitblinken in kennis van het terrein dat ze onderzoeken* (Bongers en Den Hertog, 2007). Veld- of sector kennis zit niet in het

onderzoeksteam dat evalueert. De evaluatoren kijken naar doelstellingen maar laten zich *niet* in met de beleidsinhoud. Die wordt als onomstreden vertrekpunt aanvaard.

Dat heeft het voordeel dat de onderzoekers *enige afstand houden* tot de sectorcultuur waarin 'heilige huisjes' en taboes bestaan. Maar daar staat een nadeel tegenover. Dat zou eigenlijk gecorrigeerd moeten worden. Onderzoekers moeten thuis zijn of verder thuis geraken in de materie, zeker als het om een complexe casus gaat, zeggen Van der Zwan en Noordegraaf (1998: 71) die het Gemeentelijk Vervoerbedrijf uit Amsterdam hebben doorgelicht. Een ander nadeel is dat wie zich niet grondig verdiept in de beleidsinhoud ook *moelijk met radicale suggesties* kan komen om beleid over een andere boeg te gooien. Hier dreigt dus iets te ontstaan dat eerder padafhankelijkheid is genoemd. Onderzoekers blijven werken binnen het bestaande beleidspadigma (lees: 'frame').

Tineke Abma (2001) is een van de evaluatoren die meent dat de onderzoeker een multiële evaluatie kan kiezen. Dat impliceert dat de beleidsinhoud zelf ter discussie kan komen. Dan moet je dus over materiekennis beschikken.

6 Te weinig nieuws?

Een zesde kritiekpunt op beleidsevaluaties impliceert dat gebruikers van evaluaties wel eens aangeven dat ze *niet veel nieuws* geleerd hebben van een evaluatiestudie (Leeuw en Van Gils, 2010: 172).

Dat is begrijpelijk als evaluatoren zich niet inlaten met de beleidsinhoud (zie kritiek nr. 5). Evaluatoren die zich niet inlaten met de beleidsinhoud, of die zelfs maar beperkt begrijpen, neigen ertoe om padafhankelijk te blijven denken en dus slechts te komen met suggesties in de vorm van 'kleine stappen'.

Hier is wat aan te doen door enkele maatregelen.

1. Voor de start van een evaluatie grondiger *nadenken over de evaluatieopzet*. Moet de evaluatie een *multiële of multiperspectivische evaluatie* zijn vanuit meerdere perspectieven in plaats van alleen een doelbereikings- of een effectiviteitsevaluatie? Wat worden maatstaven ter beoordeling van beleid: alleen effectiviteit vanuit een overheid gezien? Is de evaluatie een evaluatie van beleid, dat in een netwerkcontext tot stand kwam, waarbij de actoren van elkaar afhankelijk zijn omdat geen actor de volledige doorzettingsmacht heeft?
2. Een tweede suggestie houdt in om *het prospectieve* explicieter onderdeel te laten zijn van de evaluatie. Een evaluator die gevraagd wordt om een wet te evalueren, stel een mestbeperkende wet voor varkenshouders, kan pogen de bevindingen te doordenken vanuit enkele scenario's die in het land de ronde doen, zoals het scenario van radicale deregulering. Die suggestie sluit aan op denken in termen van een multiperspectivische evaluatie (Korsten, 1988; Ringeling, 1993; Abma, 2001: 314). Mijn ervaring bij de (vroegere) Sociale Verzekeringsraad (SVR) is geweest dat als dit niet bij de opdracht al is meegegeven evaluatoren hiertoe niet geneigd zijn. Die weerzin kan gevoed zijn door gebrekkige kennis van het maatschappelijke debat rond een thema. Men is niet in staat om scenario's te formuleren of denkt teveel in politiek vaarwater te belanden. Men heeft de durf van het Rathenau-instituut niet.
3. Een derde suggestie is om voor het verschijnen van een definitief rapport een of meer *expertmeetings* te organiseren waarbij juist betrokkenen uit een sector evaluatieresultaten of stellingen voorgelegd krijgen om op te reageren. Dergelijke meetings kunnen een zeefwerking hebben omdat ze uitfilteren wat in de ogen van

experts en veldkenners belangrijker of minder belangrijk is. Ze kunnen ook bijdragen aan het begrijpen van resultaten, argumentaties toevoegen of corrigeren, en additionele aanbevelingen opleveren.

7 Evaluaties te weinig prospectief

Nog een zevende kritiekpunt, dat aansluit op de twee genoemde. Menige beleidsevaluatie is *teveel retrospectief* en *te weinig prospectief*. Evaluaties blijven, anders gesteld, teveel in het verleden hangen, in waarom het ging (doelstellingen) en hoe het gegaan is (uitvoering, prestaties). Ze bieden geen of een te geringe denkact over wat er thans en de komende periode aan knelpunten speelt en wat de oplossingen daarvoor zijn. Het ontbreekt bij evaluatie dus vaak aan beantwoording van de vraag 'hoe verder'? Terwijl bestuurders na een evaluatie juist erop uit zijn om te weten 'hoe voorkomen we uitwassen'.

Denk maar aan onderzoek:

- naar de bankencrisis (commissie onder leiding van Jan de Wit) of
- naar fouten bij onderwijsinstelling InHolland (commissie-Leers),
- naar het gemeentelijk optreden bij de asbestaffaire in Utrecht (commissie-Jansen),
- naar het echec van schoolorganisatie Amarantis (rapport van de commissie onder leiding van Riet de Wit uit 2012 en van de vervolgc commissie-Halsema uit 2013) en
- naar privatisering en verzelfstandiging van overheidsdiensten (commissie-Kuiper).

Reguliere evaluaties, dus *niet* die van commissies welke na een crisis aantreden, zijn dus nogal eens te weinig prospectief.

De kans is groot dat visitatiecommissies ontkomen aan dit verwijt omdat daarin vaak ex-bestuurders naast evaluatoren zitten die weten dat bestuurlijke relevantie van belang is. Het is natuurlijk vaak ook een kwestie van opdrachtformulering.

Denk ook aan de suggestie om de bevindingen te plaatsen in multiperspectief perspectief.

Ex ante evaluatie-elementen koppelen aan de ex post evaluatie? Wie een geëvalueerde wet ineffectief acht, kan een nieuw wetsontwerp doordenken of ontwikkelen en via *spelsimulaties* 'uittesten' (Termeer e.a., 1995).

Wat is daar nog als reactie op mogelijk?

Er zijn nog diverse suggesties gedaan. Ik noem vier alternatieve typen evaluatie:

1. Responsieve evaluatie (Guba & Lincoln; Abma) en multiple evaluatie (Ringeling);
2. Argumentatieve evaluatie (Fischer, Pröpper);
3. Netwerkgericht evalueren (Klijn, Koppenjan);
4. Lerend evalueren (Edelenbos).

Ik licht deze toe en bekijk de eventuele meerwaarde.

Alternatief 1 - Responsieve evaluatie: rekening houden met belanghebbenden

Welk alternatief is er dat enigszins aan de kritiek op bestaande beleidsevaluatie tegemoet komt? In een historische schets van evalueren maken Egon Guba & Yvonna Lincoln (1989: 22-31) duidelijk dat er volgens hen *vier generaties evaluatie* bestaan hebben (zie ook Abma, 1995, 1997, 2000) en dat de *vierde generatie*, de responsieve evaluatie, een alternatief is voor de andere drie. Welke generaties zouden er dan zijn?

1. De eerste generatie evalueren is te beschouwen als 'meten'. De beleidsonderzoeker profileert zich als een technicus en richt zich op kwantitatieve gegevens. Tegenwoordig wordt dit wel aangeduid met monitoring (aldus Bongers en Den Hertog, 2007: 32).

2. De tweede generatie evalueren richt zich op *'beschrijven'*. De evaluator analyseert de sterkten en zwakten van beleid en betreft die op de gestelde doeleinden. De 'goal achievement' komt in beeld.
3. De derde generatie evaluaties richt zich op *beoordelen*. De evaluator geeft een soort finaal oordeel als een rechter.

Guba & Lincoln (1989) zijn van mening dat er in deze generaties beleidsevaluatie enkele gemeenschappelijke tekortkomingen zijn aangetroffen.

1. Ten eerste richten de evaluatoren zich in alle drie de generaties evaluatie op de *officiële doelstellingen van het bestuur*. Het is het bestuur dat de opdracht verleent tot evaluatie en de resultaten worden ook weer uitgebracht aan de (machtige) bestuurders alsof er geen belanghebbenden (met minder macht) bestaan die interesse zouden hebben in of een belang bij een evaluatie. Waar staat 'bestuur' kan ook worden gelezen 'de leiding van organisaties'. Beleidsevaluaties zijn in feite *niet machtsneutraal*.
2. Ten tweede gaat er achter de drie generaties evaluatie een gemeenschappelijk geloof schuil, namelijk in *maatschappelijke waardenconsensus*. Terwijl de maatschappij in de kern pluralistisch is geworden op waardegebied. De vraag is daarom *welke waarden* domineren in evaluatieonderzoek. Spelen waarden van andere belanghebbenden dan overheidsbestuurders ook een rol in de evaluatie of zijn deze belanghebbenden alleen geschikt om als informant te fungeren? De vraag stellen is deze beantwoorden.

Beide auteurs, Guba & Lincoln, pleiten voor een vierde generatie evaluatie (*'fourth generation evaluation'*). Ze noemen die *'evaluatie door onderhandelen'*.

Evaluatoren zouden volgens die aanpak moeten kunnen omgaan met divergerende waardeoriëntaties en dus met *meerdere maatstaven* dan alleen effectiviteit en doelmatigheid. Zeker in omstandigheden dat sprake is van pluriformiteit, en dus van controverses. Tineke Abma (1997, 2000) en Sandra Kensen sluiten hierop in hun werk aan.

Tineke Abma betoogt dat sprake is van eenzijdige rationaliteit in de vorm van instrumentaliteit in het denken volgens de klassieke beleidsevaluatie (volgens Hoogerwerf). Er wordt geen rekening gehouden met perspectieven van andere belanghebbenden dan het overheidsbestuur (Abma, 2001: 313).

Hoe ziet de gewenste responsieve evaluatie er verder uit?

1. De vierde generatie evaluatie moet een *responsieve evaluatie* zijn, volgens Guba & Lincoln (1989). Beide auteurs spreken van *'responsive focusing'*. De vragen in het evaluatieonderzoek moeten in *samenspraak* met belanghebbenden ('stakeholders') tot stand komen. De evaluatiecriteria worden ook ontleend aan de issues waar de belanghebbenden tegen aanlopen, hun 'concerns' en hun 'claims'.
2. De vierde generatie evaluatie wordt ook gekenmerkt door een *specifieke methodologie*. Die heeft te maken met een *bepaalde werkelijkheidsopvatting*. Volgens beide auteurs bestaat er *geen* objectieve werkelijkheid maar is het maar 'hoe je kijkt' naar de werkelijkheid. Er bestaan meerdere werkelijkheden ('sociale constructies genoemd'). Een evaluatie moet aan deze meervoudige wijze van kijken en van de werkelijkheid (de werkelijkheidsconstructies) waarnemen. Hoe dan te werk te gaan? Door te *'Verstehen'*: door te proberen om te begrijpen 'door je in te leven' in de positie van belanghebbenden en door met hen in gesprek te gaan. De werkelijkheidsaanspraken

zouden op elkaar betrokken moeten worden, wat zou kunnen bijdragen aan meer wederzijds begrip en inzicht in waarover overeenstemming bestaat. Het gaat om zoeken naar consensus tussen de deelnemers aan de evaluatie (Nelissen, De Goede en Van Twist, 2004: 219).

De Rotterdamse hoogleraar Arthur Ringeling (1993) erkent de zin van wat hij noemt een *multiple evaluatie* waarmee niet gezegd is dat zijn aanpak geheel overeenkomt met die van Abma. Volgens critici is deze 'vierde generatie evalueren' echter toch teveel een poging om belanghebbenden te laten deelnemen aan beleidsvoorbereiding. Zoals de Twentse hoogleraar Andries Hoogerwerf eens zei: '*Het is geen op waarneming van effecten gerichte beleidsevaluatie meer maar participatie van een onderzoeker in ontwerpprocessen*'. Het is een actiegericht ontwerpaanpak in een context van waardenpluriformiteit.

Toch zijn de opmerkingen van Guba & Lincoln nuttig. Want de neiging bestaat om een beleid te ontwerpen dat vervolgens (bijna) mechanisch 'uitgerold' wordt over de samenleving. Dergelijk beleid, waarin een overheid zich als monocentrisch besturend orgaan situeert, werkt in veel gevallen niet; zeker niet in een netwerkcontext waarin van elkaar afhankelijke actoren zijn opgenomen. Dus beleidsevaluatie zal soms dood kunnen lopen als de netwerkcontext ontkend wordt.

Een waarschuwing is mogelijk? Responsieve evaluatie valt te proberen in een netwerkcontext maar *pas op voor rolvermenging*. De evaluator dreigt in een rol te komen van zowel procesregisseur, die partijen bij elkaar brengt in een beleidsontwerp of uitvoeringsproces, als evaluator. Deze responsieve evaluatie is daarom niet zonder problemen. Bijgevolg wordt die ook *niet* vaak toegepast.

Op naar een volgend alternatief.

Alternatief 2 - Evaluatie als argumentatie

Frank Fischer (1995) is de man die in '*Evaluating Public Policy*' pleitte voor een verbreding van het gangbare evaluatieonderzoek. De oordeelsvorming moet zijn inziens in evaluatiestudies breder zijn dan dat het alleen zou gaan om effectiviteit of doelmatigheid van beleid. Evalueren moeten zich ook richten op *de betekenis van beleid voor de maatschappij als geheel*.

Fischer pleit voor een *waardengerichte evaluatie*. Evaluatoren moeten minder technocratisch te werk gaan voornamelijk of alleen door te letten op gegeven beleidsdoelstellingen en daarop aansluitende middelen. Ze evalueren niet of de 'goede dingen gedaan worden' maar meer of 'de dingen goed worden uitgevoerd'. Evaluatoren moeten zijns inzien echter het fundamentele debat bevorderen en *daarbij kan horen* dat ter discussie wordt gesteld of 'de goede dingen gedaan worden'.

Dat betekent dat de beleidswetenschap meer moet zijn dan een rationeel-instrumentalistische wetenschap. Hij bepleit een argumentatieve wending ('the argumentative turn'). Het zou bij evaluatie niet moeten gaan om het beoordelen van instrumenten en het vinden van de goede instrumenten maar om *overtuigende argumentaties*. En *dienstbaarheid aan de democratie*. Dit laatste betekent dat *debatanalyse* in beeld komt (zie ook Hoppe en Peters, 1998)

Is dit alternatief levenskrachtig gebleken, ook ten behoeve van gebruiksgerichte evaluaties?

1. Hoppe en Peters (1998) voelden zich duidelijk geïnspireerd. Zij spreken van de onafwendbaarheid van de argumentatieve wending. Ze menen dat de machthebbers de waarheid verteld moet worden. Hier is weinig tegen in te brengen. Ze stellen ook: '*Politiek als evenwichtskunst vergt argumentatieve beleidsanalyse en netwerkanalyse*'.

2. Het analyseren van politieke en maatschappelijke debatten komt weer in beeld.
3. Igor Mayer en Jacques Geurts benadrukken overigens de instrumentele mogelijkheden van de argumentatieve beleidsanalyse (en responsieve evaluatie). Ze noemen een aantal *participatieve methoden*, zoals beleidsexercities (Hoppe en Peterse, 1998).
4. Hoe verdedigbaar ook het werk van Fischer is, in de praktijk blijkt het schema van Fischer met de vier niveaus van beoordeling lastig toepasbaar (Hoppe: 'men gooit de niveaus door elkaar'). Voor een toepassing verwijzen we naar een analyse van Brandenburg (in Hoppe en Peters, 1998).

Conclusie: De aanpak van Fischer is te complex voor 'gewone' gebruiksgesichte evaluaties. Hoewel het nagaan van 'of het beleid het goede beleid is' nuttig is en ook of er sprake is van bredere effecten van beleid op samenleving en democratie. Die verbredingsdiscussie kan meegenomen worden in de hier bepleite entree-onderhandelingen voor aanvang van het evaluatieonderzoek.

Igno Pröpper (2001: 145-165) heeft aan evalueren als argumentatie een eigen uitwerking gegeven. Hij meent dat evalueren zich niet alleen moet richten op effectiviteit. Evalueren moet zich richten op meer dan instrumentele rationaliteit en ook aandacht hebben voor normatieve, interactieve en expressieve zaken. Hij beoogt via inhoudelijke en procedurele analyses besluit- en oordeelsvorming te beoordelen met behulp van criteria als overtuigingskracht ('overtuigt het beleid?') en deugdelijkheid.

Alternatief 3 - Netwerkgericht evalueren

Het hiervoor genoemde alternatief van responsief evalueren en van evalueren als argumentatie is een vorm van verbrede evaluatie ten opzichte van effectiviteitsevaluatie (Abma, 2001: 314). Toch ben ik er hiermee niet. Want is een overheidsbestuur altijd de centrale instantie in de sturing? Een andere brede evaluatie-aanpak is de aanpak waarbij de procesgang in en uitkomsten van een beleidsnetwerk (dus in een multi-actorensetting) wordt beoordeeld. Anders gesteld, het gaat om evaluatie in een *'governance'-setting*. Het governance-begrip verwijst naar *beleidsnetwerken* (Hajer e.a., 2004). Iets waarvoor Tineke Abma met de door haar bepleite responsieve evaluatie overigens oog heeft.

Ook in dit geval komen weer, net als in de kijk van Fischer en Pröpper, *andere beoordelingsmaatstaven* naar voren dan effectiviteit of doelmatigheid, namelijk onder meer openheid in het overleg en betrouwbaarheid van overlegpartners (Klijn en Koppenjan, 2001; Cörvers, Glasbergen en Korsten, 2003; De Rynck e.a., 2010).

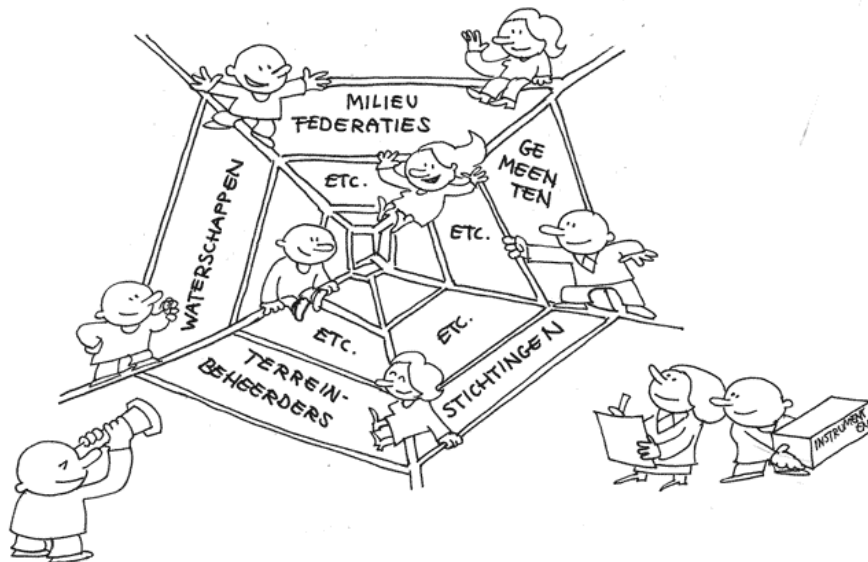
Bestaan beleidsnetwerken?

Zeker en vast bestaan beleidsnetwerken want het netwerkperspectief benadrukt de informele, gedecentraliseerde en horizontale verhoudingen die kunnen leiden tot bij voorbeeld *milieuconvenanten*. Denk ook aan *strategische groenprojecten* in de Gelderse Poort, het *gebiedsgerichte project* Kerngebied Vechtplassen, het *diagonale project* Biesbosch-Noordwaard en de *diagonale benadering* in het beekdalengebied van de Beerze en de Reusel (Cörvers, Glasbergen en Korsten, 2003).

Maar niet alle overheidsbeleid voltrekt zich in beleidsnetwerk. Ik noem twee begrenzingsen.

1. Bepaalde problemen zijn *te groot* om in een nationaal, regionaal of lokaal beleidsnetwerk te behandelen, zoals bijvoorbeeld voedselveiligheid of het

- klimaatprobleem (Hajer, e.a., 2004: 34). Voedselveiligheid moet ook Europees bekeken worden. Denk aan de gekke koeienziekte/BSE;
2. Of de problemen zijn juist *te klein*, zoals de keuze van een straatnamenplan. Een straatnamenplan is een routinezaak voor de ambtenaren.



Evident is dat beleidsnetwerken bestaan van overheidsactores en andere organisaties die *van elkaar afhankelijk zijn*. Geen actor beschikt in een netwerk over alle taken, bevoegdheden en middelen en geen enkele actor heeft dus alleen alle doorzettingsmacht (Mayntz, 1999; Hajer, Van Tatenhove en Laurent, 2004).

Denk maar eens aan het behoud en de ontwikkeling van een nationaal landschap Zuid-Limburg dat kwaliteiten heeft om – volgens sommigen – op de lijst van werelderfgoederen te staan.

Hier bewegen zich een provinciebestuur, waterschap, tal van gemeenten, enkele departementen, woningcorporaties, recreatieondernemers, de land- en tuinbouwbond, landbouwers, etc. Actoren zijn om iets voor elkaar te krijgen dus *tot elkaar veroordeeld*. Ze moeten taken, doelstellingen en middelen ‘vermengen’ of ‘verknopen’. Het provinciebestuur zal het landschap kwalificeren in het kader van een provinciaal omgevingsplan maar daarmee worden er nog geen maatschappelijke of bestuurlijke initiatieven genomen. Willen er projecten komen voor biologische landbouw, voor rondleidingen, voor fiets- en wandelroutes, voor wijnbouw en bezoek aan een wijngaard, etc. dan zullen actores elkaar in gemeenschappelijke initiatieven en projecten moeten vinden.

De actores in een netwerk, die beschikken over taken en bevoegdheden en/of middelen, zullen moeten overleggen en het eens moeten worden over wat het probleem is, of de problemen zijn en hoe die gepercipieerd worden, over een gedeelde probleemaanpak, een kaderstelling en kanalisatie van initiatieven. Meestal zullen dit overheids- of semi-overheidsactoren zijn maar in het kader van de ontwikkeling van Brainport Eindhoven wordt gesproken van *triple helix*, een verbinding tussen overheden, onderwijsinstellingen en bedrijfsleven.

Toegesplitst op het vorige voorbeeld, een landschap is er niet om slechts te behouden en conserveren, er moet ook activiteit zijn (misschien wel een verdienmodel). Het denken over beleidsnetwerken benadrukt de *informele, gedecentraliseerde en horizontale verhoudingen* tussen actoren in een netwerk, bij voorbeeld in een landschapsnetwerk of huisvestingsnetwerk.

Hajer e.a. (2004) benadrukken dat er meerdere vormen van 'governance' zijn. Een vorm is *multi-level governance*, een type sturing dat beschouwd kan worden als interbestuurlijke (overheids)sturing. Een andere vorm is de *deliberative governance*; deze vorm legt accent op de kwaliteit van de participatie van burgers en van het beraad. Deze 'deliberative governance', die de 'argumentative turn' (Fischer, Hoppe) als noodzakelijk verwerkt heeft, lijkt opener naar marktpartijen en de 'civil society' dan de 'multi-level governance'.

Gebiedsgericht beleid wordt wel gezien als een illustratie van het denken in termen van beleidsnetwerken. Hier vallen ook verschillende aanpakken te onderscheiden. Hajer, Van Tatenhove en Laurent (2004) spreken hier vanuit een historisch perspectief zelfs al van *drie generaties gebiedsgericht beleid*.

Netwerkstrategieën

Er bestaan twee netwerkstrategieën om een netwerk verder te brengen:

1. strategieën van netwerkconstituering, en
2. netwerkmanagementstrategieën (Kickert, Klijn en Koppenjan, 1997).

De vraag welke actrices met elkaar rond de tafel moeten en op welk moment, is er een van netwerkconstitutie.

De tweede strategie, netwerkmanagement, uit zich in vragen als hoe actrices via het wegwerken van barrières (fixaties) en het scheppen van positieve condities in een netwerk komen tot een gemeenschappelijke koers, bij voorbeeld in de vorm van gemeenschappelijke beeldvorming, die uitmondt in doelverrijking en middelenbundeling.

Zijn netwerken zelforganiserend?

Dat wordt wel gesteld (Hajer e.a., 2004: 16) maar dat getuigt mijn inziens van overoptimisme. Er is doorgaans in netwerken werk aan de winkel voor een procesregisseur. De bal moet in het spel gebracht worden, daarin gehouden worden en blijven rollen. Het kan ook wat gewichtiger gesteld worden. Procesmanagement moet de interacties in de 'beleidsspelen' die in het netwerk plaatsvinden, bevorderen (De Bruijn en Ten Heuvelhof, 2001; De Bruijn e.a., 2002). Mijns inziens gaat het vooral om sturingspogingen om de percepties van de actrices bij elkaar te brengen en over elkaar te leggen. De actrices hebben ieder voor zich wel een zekere ruimte van handelen, verrichten ook wel zelfstandig bepaalde taken (zoals het waterschap in een gebied dat gekwalificeerd is als nationaal landschap) maar slagen er niet in om het probleem of bepaalde uitdagingen alleen op te lossen. Ze moeten zich op elkaar kunnen en willen oriënteren en zijn in feite gedwongen om de eigen positie te relativiseren en zich in de ander te verplaatsen. Om het eens te worden. Hier moet een *procesregisseur* helpen. Dat proces van bij elkaar brengen van actrices, van begripsvorming voor elkaars positie en gedeeld bewustzijn moet uitmonden in arrangementen tussen actrices, zoals een convenant, om verder te kunnen. En conflicten moeten worden uitgepraat of getemperd.

Hoe komt je tot een gedeeld beeld in een netwerk?

Misschien door 'reframing', dus door anders naar een vraagstuk te kijken en verbeeldingskracht te tonen (dus door een ander verhaal). Dit wordt keer op keer door netwerkspecialisten en anderen benadrukt (Korsten, 1988; Schön & Rein, 1994; Klijn en Koppenjan, 2001: 183).

Netwerkmanagement is al met al geen eenvoudige kost. Het vereist veel competenties, zoals op het vlak van kennis van een netwerk, kennis van wet- en regelgeving en oog voor belangen, en onderhandelingsvaardigheid. Het ontwikkelen en tonen van verbeeldingskracht is geen overbodigheid.

Procesmanager

Wat is een *procesmanager* voor een type manager? Niet iemand die proclameert en verordonneert of geldt als centrale bestuurder maar eerder iemand die (tijdelijk) stimuleert, faciliteert, verbindt en bemiddelt. Dat kan een van de actoren uit het netwerk zijn of een externe. De rol van regisseur is een tijdelijke. Wie het is, kan afhangen van het vraagstuk, initiatief of project (Klijn, Koppenjan en Termeer, 1995; Van Eeten en Termeer, 1996a, b; De Bruijn en Ten Heuvelhof, 2001).

Wie zien dat in dit soort interacties in netwerken *het proces de inhoud stuurt*. Als het proces erkenning krijgt en partijen niet uitstappen en het vertrouwen groot is en blijft, kan dit de voorwaarden scheppen of bevorderen om het ergens over eens te worden. Is het vertrouwen niet groot (genoeg) dan ontstaat er een kans op een '*dialogo van doven*', op *stroperigheid* en *proceduralisme*. Dat impliceert dan dat de aandacht voor interactie de aandacht voor de inhoud kan overstijgen, dus dat de inhoud wordt weggedrukt.

Het is mogelijk dat het proces de inhoud stuurt maar dat de inhoud geen stand houdt. Als partijen het over iets eens zijn, wil dat nog niet zeggen dat het werkt omdat voor sommige zaken andere kennis van belang is dan een onderhandelingsresultaat ('negotiated knowledge') biedt. Een bekend voorbeeld betreft een onderdeel van de ruimteveer Challenger. Dit ruimtevaartuig was veilig volgens de procesbesprekingen, zeg maar het procesmanagement. Toch deugde een bepaalde ring niet en ontplofte de Challenger. Bij hoog technologische vraagstukken kun je dus niet volstaan met procesmanagement-oplossingen, bij deliberatie.

Naar netwerkgerichte evaluatie

Wat betekent dit nu voor de beleidsevaluatie vanuit een netwerkperspectief: welke criteria zijn gewenst om uitkomsten te beoordelen? Hoe wordt onderscheiden tussen succes en falen?

- Het is *ongewenst* om het verloop en de uitkomsten van netwerkproductiviteit te beoordelen aan de hand van de doelstellingen van één actor, ook al is dat een actor in het overheidsbestel (Klijn en Koppenjan, 2001: 189). Er is *niet* slechts één algemeen belang. Er is niet slechts hiërarchische sturing. In een evaluatie van een beleidsnetwerk, bijvoorbeeld van een huisvestingsnetwerk, moet uitgegaan worden van de visie van al de in het netwerk opgenomen of toegelaten actoren. Actoren zijn autonoom en trekken zich dus maar beperkt wat aan van de ene actor die denkt dat de macht bij hem ligt. Dus luidt het advies: *ga in een beleidsevaluatie uit van meerdere maatstaven*. Dat zien we ook in het proefschrift van Cörvers over regionale netwerkprojecten op het vlak van natuurbeleid (ook Cörvers, Glasbergen en Korsten, 2003).
- Hou er rekening mee dat actoren in een beleidsnetwerk *doelzoekend gedrag* kennen (Herweijer, 1981). Soms weten 'ze' in de fase direct na netwerkconstitutie nog niet precies wat ze willen. Logisch, want alleen kunnen ze maar beperkt wat voor elkaar krijgen, zoals gesteld. Ze hebben andere organisaties nodig om samen dingen na te streven en mogelijk te maken en dat moet verkend worden. *Advies aan de evaluator*: kijk in een netwerkgerichte evaluatie daarom of actoren elkaar opzoeken, en of sprake is van vertrouwen en of actoren initiatief nemen (Klijn, 2002). Actoren kunnen zelf leerprocessen doormaken. Wat eerst niet lukt, gaat later misschien wel. Percepties van

problemen en uitdagingen zijn dynamisch; ze zijn in ontwikkeling. Door uit te gaan van geformuleerde doelstellingen van één actor wordt dit leren van andere actoren ontkend. De netwerkdenkers zijn op dit punt heel uitgesproken. *Verplaats je als evaluator in elk van de betrokken actoren.*

- Het is bovendien gewenst rekening te houden met *belangen* van de verschillende actoren in een netwerk. Leidt samenwerking tot win-win situaties, meer dan elk alleen zou kunnen bereiken?
- Niet uitgesloten is dat na verloop van tijd de netwerkactoren *samen* besluiten tot een doelstelling van het netwerk. Dat is dan zeker een prominent punt en dus in een evaluatie te betrekken. *Advies aan de evaluator: zoek ook naar gezamenlijke netwerkgerichte doelstellingen.* Hierop ga ik nader in.

De Rynck, Voets en Van Dooren (2010) hebben zich gezet aan een poging om prestaties van netwerken als geheel in een uitgewerkter analysekader te vatten. Zij volstaan dus niet met de evaluator te laten kijken naar hoe afzonderlijke actoren zich opstellen, dus naar wat afzonderlijke actoren procesmatig gezien doen.

De auteurs onderscheiden 1) productieprestaties, 2) procesprestaties en 3) regimeprestaties. Ik licht dit toe.

Figuur 8: Beoordelingskader voor netwerkprestaties vlg. De Rynck c.s.

Dimensies	Criteria	Kosten en baten
Productieprestaties	Doelbereiking en efficiëntie	Kosten en baten verbonden met productieprestaties
Procesprestaties	Legitimiteit Verantwoording - verantwoording afleggen - verantwoordelijk gehouden instemming - issues waarover instemming gevraagd wordt - mechanismen voor verlenen instemming - status van instemming	Kosten en baten verbonden met procesprestaties
Regimeprestaties	Lidmaatschap Netwerkinstitutionalisering - netwerkorganisatie Kwaliteit van relaties - dichtheid - veelvormigheid	Kosten en baten verbonden met regimeprestaties

Productieprestaties op het eerste niveau verwijzen naar output van het netwerk.

‘Beleidsnetwerken zijn alleen succesvol als ze effectief zijn’, schrijven de auteurs. Dat wil zeggen dat ze ‘een meerwaarde’ bieden ten opzichte van wat individuele organisaties kunnen en willen. Zonder een zekere effectiviteit verliest een netwerk het vertrouwen van de deelnemers en wordt het netwerk irrelevant. Wat is dan effectiviteit? Effectiviteit heeft van doen met doeleinden. Maar welke?

Op het niveau van de individuele organisaties die deel uitmaken van een netwerk gaat het om de tevredenheid over het netwerk. De evaluator kan de actoren daar naar vragen.

Doelbereiking op het niveau van het netwerk (!) is nagenoeg bereikt als organisaties van mening zijn dat wat het netwerk bereikt aan productiviteit hun doelstellingen dient. Als organisatie de betekenis van het netwerk zien en erkennen kan dat de netwerkproductiviteit verder ten goede komen.

Het tweede *niveau* van beoordeling van prestaties in en van een netwerk betreft *procesprestaties*. Procesprestaties in en van een netwerk verwijzen naar democratische kwaliteit. Netwerken kunnen door de kwaliteit van de interacties in processen bijdragen aan

het opbouwen van een maatschappelijk en bestuurlijk draagvlak voor beleid. Hierbij worden drie elementen onderscheiden: legitimiteit, verantwoording en goedkeuring of instemming.

Het derde *niveau van analyse* verwijst naar *regimeprestaties*. Het gaat er hier om of en in welke mate het netwerk de capaciteit van de publieke sfeer versterkt: maakt het netwerk de maatschappelijke infrastructuur sterker?

Drie criteria ter beoordeling zijn hier te benoemen volgens de auteurs: de aard van het 'lidmaatschap' van het netwerk, de duurzaamheid van de deelname en de kwaliteit van de relaties.

Een lidmaatschap impliceert dat een netwerk toegang biedt tot hulpbronnen die voor een probleemaanpak van betekenis zijn, zoals bevoegdheden, informatie, productiecapaciteit, legitimiteit, financiële middelen. Hoe meer relevante leden, hoe meer hulpbronnen ingebracht worden, of ontstaan, of gemobiliseerd kunnen worden. Bij teveel deelnemers kunnen de coördinatiekosten (of transactiekosten) echter weer gaan oplopen.

Een tweede regimecriterium slaat op de structuur van het netwerk en de manier van organiseren van het netwerk. Stabiliteit van het netwerk kan de capaciteit van de overheid vergroten. Stabiliteit en management van het netwerk kunnen een indirecte prestatie-indicator vormen.

Het derde criterium om de regimeprestaties in het zicht te krijgen betreft de kwaliteit van de relaties tussen actoren in het netwerk. Aantal relaties, dichtheid van relaties en veelvormigheid zijn van belang, aldus De Rynck c.s. Hoe groter de relatiedichtheid en hoe veelvormiger de relaties, hoe hoger de regimeprestaties van het netwerk (De Rynck c.s., 2010: 43). Hiermee komen we weer dicht bij het denken van auteurs als Klijn c.s.

Conclusie: we zien met een netwerkgerichte evaluatie dus duidelijk een verbrede evaluatie. Deze vorm van evaluatie komt nog maar schaars voor omdat deze evaluatie relatief duur is. Immers, de evaluatoren moeten zich verplaatsen in afzonderlijke actoren uit het netwerk. Bovendien, zonder een beetje doorleefde kennis van van het netwerkdenken kan een evaluator er maar beter niet aan beginnen. Er worden dus ook eisen gesteld aan de evaluator. Is een medewerker die juist heel erg het netwerk kent dan geschikt? Wel als het ook om iemand gaat met competenties op het vlak van evaluatie.

Vingeroefeningen

Is er een voorbeeld te vinden van een evaluatie op basis van de 'network approach' in evaluatie? Ron Cörvers (2001) heeft in zijn proefschrift een aantal gebiedsgebonden netwerkprojecten met een ruimtelijke en natuurcomponent beoordeeld. Deze netwerkprojecten werden betiteld als *dialogisch beleid*. Interactie was cruciaal. Uit die evaluatie blijkt onder meer dat in diverse projecten, zoals het project de Gelderse Poort, het Vechtplassengebied en de Biesbosch, voor 2001 sprake was van een *heel trage voortgang*; er verschenen plannen op plannen, dat wel. Integrale oplossingen ontbraken vrijwel volledig en van gemeenschappelijke beeldvorming was evenmin sprake. Het bundelen van overheidsmiddelen vond slechts sporadisch plaats. Investeringsmiddelen uit de private sector werden nauwelijks aangesproken. Win-winuitkomsten waren meer ideologische schijn dan realiteit. Dus teleurstellende bevindingen maar er waren ook opmerkelijke uitkomsten.

Als er een gemeenschappelijke doelstelling centraal staat, betekent dit meestal toch dat het beleid van een 'hogere' overheid betreft. *Hiërarchische verhoudingen* blijken nog steeds een rol te spelen. Ze blijken hardnekkig. Van nevenschikking tussen overheden bleek niet echt sprake.

Wat ook heel opmerkelijk was, was dat *externe omstandigheden* een cruciale rol speelden bij de voortgang in gebiedsgerichte projecten. Als er een crisis optreedt, zoals overstromingen, blijkt er ineens productie (lees voortgang) mogelijk in het netwerk, zoals dijkversterking (Cörvers, Glasbergen en Korsten, 2003).

Conclusie: een evaluator moet er altijd op bedacht zijn dat een beleidsnetwerk in een overwegend verticale sturingscontext kan functioneren of weer getrokken kan worden. In zoverre begrijpelijk omdat elke overheidsactor weer functioneert in een democratische context; GS moet terug naar Provinciale Staten, etc. Die contexten kunnen de vrijheid van handelen van dagelijks bestuurders blijkbaar danig verlammen. Dan kan de verticale sturing het op bepaalde momenten gaan winnen. Succes is bij netwerkmanagement dus zeker geen vanzelfsprekendheid.

Kritiek op netwerkgericht evalueren

Er is ook wel kritiek geuit op de netwerkbenadering bij evaluatie (Pröpfer, 1996; Klijn en Koppenjan, 1997).

1. Een kritiekpunt van iemand als van de Utrechtse milieukundige Pieter Glasbergen is dat degenen die netwerkproductiviteit beoordelen teveel aandacht besteden aan de proceskant en te weinig aan wat het netwerk aan productie voortbrengt. Zoals aangegeven maakte promovendus Cörvers (van de promotores Glasbergen en Korsten) zich hieraan niet schuldig.
2. Een tweede kritiekpunt is dat heel goed in de gaten gehouden moet worden hoe de uitkomsten van een netwerkproductieproces zich verhouden tot de verantwoording in politiek-democratische organen waarmee elk van de overheidsactoren te maken heeft. De leden van deze organen kunnen het gevoel hebben dat gepoogd wordt hen een netwerkresultaat 'door de strot te duwen' (Hajer e.a., 2004).

Ondanks deze kritiek is er geen alternatief voor netwerkgericht evalueren. Een evaluator kan niet doen alsof er geen netwerkcontext is als de verantwoordelijkheden versnipperd zijn, de afhankelijkheden tussen actoren groot is en geen enkele actor de volledige doorzettingsmacht heeft om de eigen beleidsvoorkeur er doorheen te duwen. Dan moet je netwerkgericht evalueren.

Toch is er nog een ander concept gelanceerd, en wel door enkele auteurs die vertrouwd zijn met zowel netwerkdenken als responsief evalueren. Dat voert ons naar alternatief 4. We zullen zien dat dit op zich interessante type in onze ogen niet probleemloos is. Nog niet proberen te doen dus in een context van gebruikgericht evalueren.

Alternatief 4 - Lerend evalueren: leren in een beleidsnetwerk

In een vierde alternatief - *lerend evalueren* genoemd - wordt teruggegrepen op Guba & Lincoln: dat wil zeggen op proberen de belanghebbenden te begrijpen. En op de netwerkbenadering.

Onder het motto van 'al evaluerend leren' hebben Jurian Edelenbos en Arwin van Buuren (2005) zich gericht op het verder brengen van evalueren in prospectief opzicht. Het gaat hen om evalueren als leerproces in een netwerkcontext van van elkaar afhankelijke actoren.

Zij laten het inventariseren van succes of falen en het beoordelen als onderdeel van een gebruikgerichte evaluatie niet vallen maar doen wel een poging om verder te komen.

Wat is de kern? Evaluator en geëvalueerde moeten volgens hen *gezamenlijk een leerproces organiseren*. Dat stelt uiteenlopende eisen aan evaluatoren en genereert *rolconflicten* voor hen.

Deze aanpak is tot stand gekomen na een doorlichting van het Stimuleringsprogramma Burger en Milieubeleid van het toenmalige ministerie van VROM.

Hun inzichten sluiten aan op die van Guba en Lincoln (fourth generation evaluation; 1989), Abma (1996) en Edelenbos en Van Eeten (2001). Wat betekent dat?

1. Ze betrekken bij hun aanpak het inzicht dat als je klassiek evaluerend te werk gaat de geëvalueerden of andere gebruikers van een rapport zich niet actief opstellen. Ze zijn net receptief. Ze zoeken niet zelf actief naar wat te doen met de resultaten van een evaluatie. De beide auteurs menen dat het in het belang van het beleid is ('de zaak zelf') dat actoren juist wel *participeren*.
2. De evaluator wordt geen controleur meer maar *samenspeler*. Hij gaat zich verplaatsen in de positie van elk van de 'partijen' in een netwerk. Want netwerkconstitutie is nodig (goede partijen aan tafel); en het netwerk moet productief worden (gemeenschappelijke beeldvorming en doelverrijking).
3. De evaluator tracht zijn inzichten *over te brengen* op en *uit te wisselen* met de geëvalueerden om zo tot gezamenlijke kennis en betekenisvolle inzichten te komen. Daarom wordt ook wel gesproken van een '*deliberatieve evaluatie*' (House & Howe, 2000).
4. De evaluator is in hun optiek van een 'lerende evaluatie' ook een *verandermanager*. Hij neemt verantwoordelijkheid om het programma mee te verbeteren. Deze evaluator wordt een 'key person' genoemd die bijdraagt aan 'a more sophisticated reality' (Guba & Lincoln, 1989: 262).

Conclusie

Evaluatoren krijgen waardering en kritiek. Tot de waardering behoort dat er steeds meer evaluatiekennis bestaat in de vorm van tijdschriften en handboeken. Evaluatie is een vakgebied. Wie voldoende kennis heeft van dit gebied kan ook op een aanvaardbare manier gebruiksgesichte evaluaties verrichten.

Tot de kritiek behoort onder meer:

1. dat de evaluatoren soms ver af kunnen raken van de ideale onderzoeksopzet met nulmeting en volgende metingen; dan zijn 'remedial' oplossingen nodig; ;
2. dat evaluatoren wel weet hebben van evaluatie maar te weinig veldkennis hebben, dus van het gebied dat ze evalueren;
3. dat ze te weinig prospectief en teveel retrospectief gericht zijn;
4. dat ze met hun aanbevelingen teveel binnen het bestaande beleidsparadigma blijven (waardoor aanpassingen na evaluatie weer tot beleidsfalen kunnen leiden);
5. dat ze te weinig oog hebben voor de belangen van degenen die het beleid 'ondergaan' of anderszins betrokken zijn en teveel gericht zijn op de beleid voerende overheid (unicentrisme);
6. dat ze te weinig oog hebben voor het feit dat een overheid vaak geen doorzettingsmacht heeft en afhankelijk is van andere overheden.

Daarom zijn er remedies bedacht. Evaluaties kunnen argumentatief versterkt worden (Pröpfer), meer responsief zijn (Abma) of waar opportuun meer netwerkgericht (Klijn). Evaluatoren doen er goed aan te vermijden dat evalueren zo opgerekt wordt dat het overgaat

in beleid ontwerpen en evaluatoren in rolconflicten terecht komen. Edelenbos en Van Buuren hebben gepoogd creatief om te gaan met een opdracht om een lerende evaluatieaanpak te introduceren. Ze verdienen lof daarvoor maar in werkelijkheid levert hun aanpak toch serieuze problemen op. De betrokkenen krijgen te maken met een evaluator die verschillende rollen vervult. Niet steeds is voor hen duidelijk in welke rol de evaluator 'nu weer zit', hetzij in de rol van begeleider en adviseur, hetzij in die van beoordelaar.

18 DE AMBTENAREN EN BESTUURDERS: WAT ZE DOEN MET EVALUATIES

Evaluëren heeft mede tot doel om tot verantwoording te komen en ervan te leren. Hoe kan dat leren worden gezien?

Leren kan worden geoperationaliseerd in verstrekking van informatie over effectiviteit en doelmatigheid van beleid (de verantwoordingsfunctie) en gebruik voor verbetering van beleid en/of de beleidsuitvoering (Korsten, 1983; Leeuw e.a., 2009; Mark & Henry, 2004; Leeuw en Van Gils, 2010).

Drie manieren om gebruik te analyseren

Er zijn in principe drie manieren om gebruik (of benutting) of non-gebruik en de vormen van gebruik te bestuderen (Caplan, 1979; Van de Vall, 1980; Korsten, 1983; Askim, 2007, 2008; Mark & Henry, 2008).

1. Door uit te gaan van een bepaalde *evaluatiestudie* en te zien welke tocht een rapport door de overheid maakt, of het op de goede plekken terecht komt of niet, of het gelezen wordt of niet, of er reflectie op uitkomsten plaatsvindt en of mogelijk zelfs conclusies en aanbevelingen overgenomen worden of niet. Hier dient zich een fasenperspectief aan. Van de Vall heeft deze aanpak aanbevolen en toegepast.
2. Door uit te gaan van *beleidsvoerders* en vanuit hun opdrachten en werkzaamheden te kijken naar hun informatiebronnen, en naar de momenten en omstandigheden van en voor benutting.
3. Door uit te gaan van de *interactie* tussen onderzoekers en beleidsvoerders. Deze aanpak kan vooral toepassing vinden als een beleid in een netwerkcontext plaatsvindt, een netwerk van van elkaar afhankelijke partners die elk de doorzettingsmacht missen om alles te bepalen. Wat doet een beleidsevaluatie in een netwerkcontext?

Deze benaderingen kunnen ook complementair worden toegepast. Dan geven ze steeds een andere doorkijk die samen een beeld biedt.

Grofweg leidt de eerste aanpak, uitgaan van onderzoek, tot de vaststelling dat onderzoek in een organisatie kan stagneren. De fase-optiek leert dat er hier en daar wel kennis genomen wordt van resultaten uit een evaluatieonderzoek maar *de routing* laat vervolgens zien dat de onderzoeksbevindingen als het ware verbleken; de samenvatting van het onderzoek wordt steeds korter tot die bijna nietszeggend wordt.

Als onderzoek laag in de organisatie is ingebracht en/of het beleid waarop de evaluatie slaat niet hoog op de politiek-bestuurlijke agenda staat dan kan de beleidsevaluatie blijven 'hangen' in de organisatie. Dan maakt een rapport wellicht nog wel een *incubatietijd* door.

Mijn onderzoek wijst uit dat beleidsevaluaties vaak op *langere termijn* en *indirect* gebruikt worden. Indirect omdat ook andere actoren zich uitspreken over een onderwerp en eventueel een onderzoek ook van commentaar voorzien. Het evaluatierapport leidt als het ware een *zwervend bestaan* en kan ineens weer opduiken en relevantie krijgen door een incident of doordat om wat voor reden dan ook een bezinning op beleid aan de orde is.

De tweede aanpak, door uit te gaan van beleidsvoerders, maakt markanter dan de eerste aanpak, duidelijk dat beleidsvoerders bezig zijn en dat het verschijnen van een beleidsevaluatie een kleine eruptie is in een druk bestaan van ambtenaren waarin van alles gebeurt en nodig

wordt geacht. Ambtenaren benutten voor hun werk vaak tal van bronnen, waaronder politieke standpunten in de vorm van moties, standpunten uit het betreffende netwerk, protestacties, enz. Echter, als het onderzoek deugt en markante uitkomsten bevat die de media 'inknallen', kan dit leiden tot vragen vanuit een volksvertegenwoordiging en krijgt een issue ineens een lift op de agenda. Studies over benchmarking halen meestal niet de voorpagina's van dagbladen en halen niet tv-rubrieken maar kunnen wel een rol spelen in verantwoordingsprocessen of dienen als rechtvaardiging voor 'een route die men toch al wilde inslaan'.

De als derde genoemde aanpak om gebruik van evaluatiestudies te analyseren door uit te gaan van interactie tussen ambtenaren en tussen bestuurders en ambtenaren focust op wat er overleg en debat plaatsvindt op een beleidsveld.

Neem aan dat in een netwerk tal van actoren opereren die elk slechts beperkt beschikken over taken, bevoegdheden en middelen. Elk krijgt dan alleen weinig voor elkaar en heeft dus andere partners nodig. Het gaat dan om doorbreking van het 'eigen belang'-denken en het doorbreken van fixaties om uit te komen bij het realiseren van verbindingen van doelstellingen en het bundelen van financiële en andere middelen. Beleidsevaluaties kunnen daarin een rol spelen. Denk maar aan een rapportage van het toezicht op de mijnen over de beperking van gasboringen in Groningen na de dag na dag opgetreden trillingen in de Groningse aardbodem (in februari 2013). Toen deze beperking in de omvang van gewonnen gas als idee werd geopperd, kwam dit rapport ineens weer bovendrijven. Het werd 'een wapen' in de discussie. De interactie-aanpak maakt duidelijk dat evaluaties een ammunitie- of legitimatiefunctie kunnen vervullen.

Gebruiksfuncties van beleidsevaluaties

Beleidsevaluaties kunnen verschillende gebruiksfuncties vervullen (Weiss, 1980; Korsten, 1983). Ik noem er een paar. Ze kunnen

1. een *agenderingsfunctie* vervullen, dus een onderwerp opbrengen,
2. een *reflectiefunctie*, want bezinning op de beleidskoers of middelen bevorderen,
3. een *legitimatiefunctie* vervullen: een bevestiging geven van een ingeslagen route,
4. een *'enlightenment'*-functie: in het algemeen verhelderend werken,
5. een *ammunitiefunctie* vervullen, dus een middel zijn voor een bepaalde actor om andere organisaties te bestoken,
6. een functie voor *'blame avoidance'*: tegelijk met de verantwoording over beleid wordt de eigen stoep schoongeveegd.

Wat leert onderzoek hierover?

Een onderzoek naar leren van evalueren bij het ministerie van Buitenlandse Zaken laat zien dat verschillende vormen van gebruik van evaluatieonderzoek vielen aan te treffen? Inderdaad, zo bleek op basis van analyse van dertig evaluaties (Leeuw e.a., 2009). Welke vormen?

- 1) Ten eerste bleek beleidsevaluatie een *intern communicatiemiddel* ter ondersteuning van de overdracht van functie. Waarom? Omdat beleidsevaluaties vaak het enige middel zijn voor het bieden van een overzicht van (delen van) het beleid. Dit is een functie waarop de leiding binnen de organisatie niet stuurt. Het is eigenlijk een nevenfunctie van onderzoek.
- 2) Ten tweede vervult evaluatieonderzoek een *interne promotiefunctie*. Ook in dit geval gaat het om een nevenfunctie. Het onderzoek biedt een mogelijkheid om jezelf als initiatiefnemend ambtenaar of de eigen afdeling meer in de schijnwerpers te zetten.

3) Ten derde viel de *legitimatiefunctie* te traceren. Een onderzoek dient ter ondersteuning van beleid of als basis voor overleg met externe partijen die subsidie ontvangen of regelingen uitvoeren.

4) Ten vierde is er blijkens het onderzoek bij Buitenlandse Zaken de *verantwoordingsfunctie*: een beleidsevaluatie is een middel om beleid systematisch vast te leggen zodat verschillende partijen zich eraan kunnen binden of gebonden kunnen worden. Evaluaties worden gebruikt ter verantwoording en als zodanig, al of niet voorzien van een reactie, toegestuurd aan de Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (IOB) en aan de Tweede Kamer.

5) 'Veel evaluaties' kennen nog een ander gebruik: het *wijzigingseffect*. Evaluaties worden aangewend om zich op het geëvalueerde beleid te bezinnen of het te herzien. Dat houdt niet in dat de *'policy change'* volledig kan worden teruggevoerd op de beleidsevaluaties maar de samenhang is er zeker, tenminste in de hoek van het directoraat-generaal Internationale Samenwerking (DGIS), aldus Leeuw e.a. (2009: 10) . Soms leidt evaluatie tot beperkte aanpassingen of blijft actie achterwege. En soms lopen beleidsveranderingen vooruit op conclusies en aanbevelingen van de evaluatie. Dat komt omdat ambtenaren tussentijds geïnformeerd worden en zo informatie mee kunnen nemen in de beleidsvoorbereiding.

Wordt evaluatie gebruikt om bestuurders af te rekenen en zelfs naar huis te sturen? De gehele literatuur over beleidsevaluatie overziende, kan niet anders gesteld worden: *bestuurders worden zelden op basis van alleen een evaluatierapport de laan uitgestuurd* (Korsten, 2010). Hoe dat kan?

Omdat evaluaties vaak complicerende factoren aan het licht brengen, een beleid vaak al langere tijd loopt waardoor een bestuur niet zelf verantwoordelijk gehouden kan worden en de 'sorry-democratie', een term van Ed van Thijn, zijn werk doet. Over de *'blame avoidance'*-functie, dus over de stoep schoon vegen met een beleidsevaluatie, lezen we bij Leeuw c.s. (2009) overigens niets.

Patroon van gebruik

Van gebruiksfuncties naar patronen of vormen van gebruik is maar een kleine stap.

Figuur 9: Veel voorkomende patronen van gebruik van beleidsevaluaties

<i>Patroon van gebruik of doorwerking</i>	<i>Omschrijving: evaluatie leidt tot... of draagt bij aan</i>	<i>Voorbeeld</i>	<i>Typering van gebruik</i>	<i>Kanttekening</i>
Instrumenteel gebruik	Door evaluatie beleidsverandering en aanpassing gedrag van organisaties in uitvoering	Beeldende Kunstenaars-regeling (BKR)	Gegevensbron voor beleid	Veel in vorm van 'bijstelling'
Conceptueel gebruik	Door evaluatie verandering van referentiekader bij betrokkenen	Gelijkheid van kansen-studies en onderwijsbeleid	Inspiratiebron voor beleid in vorm van begrippen, theorieën	Onderschatte vorm van gebruik
Agenderend gebruik	Door evaluatie komt een nieuw onderwerp of thema op agenda	Onderzoek naar SNS levert zich op fraude op	Inspiratiebron voor beleid	Kan gaan om bijvangst van evaluatie
Politiek-strategisch gebruik	Door evaluatie pogen een of meer partijen de machtspositie te versterken, belangen te dienen of doelstellingen te realiseren	Evaluatie van geluidsnormen op Schiphol	Legitimatie van eigen positie (o.a. Milieudefensie)	In netwerk-contexten

Bronnen: O.a. Korsten (1983); Bekkers, Fenger & Homburg (2005); Bekkers (2007: 321).

Aanzienlijk gebruik

Hoe omvangrijk is het gebruik van uitkomsten van evaluaties?

Bij evaluaties die betrekking hebben op affaires, fiasco's of crises worden rapporten doorgaans *veel* gebruikt.

Vergelijk de rapportages in de Ceteco-affaire (provincie Zuid-Holland), de te 'ondernemingsgerichte' bestuurscultuur (cie-Van den Heuvel; Noord-Holland), de asbest-affaire (cie-Jansen; Utrecht), de vuurwerkcramp (cie-Oosting; Enschede), de cafébrand (cie-Polak; Volendam), de bankencrisis (De Wit/Halsema).

Hoe is het gebruik van uitkomsten indien er weinig of geen ophef over beleid is ontstaan? Van een eenvoudig rechtstreeks, lineair en oorzakelijk verband tussen bevindingen uit beleidsevaluaties en beleidswijzigingen is over het algemeen *niet* te spreken. Ideeën voor beleidswijziging zijn vaak op *meerdere* informatiebronnen, discussies en bezinningsmomenten terug te voeren (Weiss, 1980; Korsten, 1983).

Maar Leeuw e.a. (2009: 11) melden over gebruik van beleidsevaluaties bij Buitenlandse Zaken wel iets opmerkelijks:

'Niettemin is er over het algemeen sprake van een aanzienlijke mate van overeenstemming tussen de inhoud en teneur van evaluaties en die van het beleidsdebat en de beleidswijzigingen. De meest bescheiden conclusie is dat in elk geval een deel van de door evaluaties gesignaleerde problemen en voorgestelde oplossingen op de beleidsagenda staat'.

Deze uitkomst heeft de onderzoekers van gebruik van de evaluaties zelf verrast. Zij veronderstelden aanvankelijk dat in de diplomatieke cultuur van Buitenlandse Zaken, met vrij veel doorstroming van personeel, weinig of zelfs zo goed als geen benutting zou optreden. Een analyse van 30 evaluatiestudies liet echter wat anders zien. Veel evaluaties gaan zelfs gepaard met een *'herziening van het beoordeelde beleid'* (Leeuw en Van Gils, 2010: 168, 169).

Welke kennis meest gebruikt: doelbereikingsanalyses of procesevaluaties?

Het type kennis dat evaluaties oplevert, maakt uit voor gebruik van evaluaties en het bereiken van leereffecten. Bepaalde kennis werkt dus *wel* door, andere *minder*. Welke werkt dan door? Een vraag in dit verband luidt: werken doelbereikingsevaluaties meer door dan procesevaluaties of omgekeerd?

Het viel op dat bij Buitenlandse Zaken zowel doelbereikingsevaluaties als procesevaluaties werden verricht. Een vergelijking is dus mogelijk.

Meer belang van doelbereikingsevaluaties

Doelbereikingsevaluaties bleken nadrukkelijk *meer* van belang dan procesevaluaties. Ze werden meer gebruikt. Waarom? Doelbereikingsevaluaties leveren informatie over de vraag of beleid is uitgevoerd zoals de bedoeling is en of de doelstellingen zijn gehaald. Dergelijk onderzoek geeft een reconstructie van de beleidsintenties en die informatie blijkt nuttig want die wordt vaak betrokken bij de bezinning op de vraag of beleidsvoerders moeten doorgaan met het bestaande beleid of een aanpassing moeten plegen in de vorm van een andere doelgerichtheid of krimp aan middelen of iets anders. Dergelijk onderzoek kan ook informatie geven over hoe gedrag te beïnvloeden is en biedt misschien zelfs wel generaliseerbare kennis (Leeuw e.a., 2009: 11).

Minder belang van procesevaluaties

Procesevaluaties hebben ook hun nut, dat zeker. Er gaat gebruikswaarde van uit. Maar procesevaluaties bieden *minder* relevante informatie als de bezinning op beleid aan de orde is. Procesevaluaties bieden in feite minder aanleiding en gelegenheid tot leren, stellen Leeuw e.a. (2009: 11).

Wat werkt in op gebruik of benutting?

Op benutting kunnen in principe tal van factoren inwerken (Weiss, 1972, 1980; Korsten, 1983; Mark & Henry, 2004; Askim, 2007, 2008; Askim, Johnson & Christophersen, 2008). Ik maak een indeling in drie categorieën factoren.

1. aan de kant van de onderzoeker en het onderzoek (ook wel '*research constraints theories*' genoemd),
2. aan de beleidskant en de beleidsvoerders (ook wel '*interaction theories*' genoemd), en
3. op het vlak van interactie tussen onderzoekers en beleidsvoerders (ook wel '*policy constraints theories*' genoemd).

Deze indeling is een zinnige. Veel onderzoek heeft aangetoond dat factoren in *elk* van deze clusters van betekenis kunnen zijn en vaak ook zijn (Caplan, 1979; Weiss, 1972, 1974, 1979, 1980; Korsten, 1983; Van Oijen, Van Westerlaak en Snellen, 1984).

Ad 1 Aan de kant van de '*research constraints theories*' valt te denken aan de volgende variabelen die negatief of positief kunnen inwerken op de mate van gebruik van beleidsevaluaties door bestuurders en ambtenaren en eventueel andere partners in een beleidsnetwerk:

- een probleemstelling die vaag of onduidelijk is (negatieve invloed op gebruik of benutting),
- een methode van onderzoek die niet goed aansluit op de vraagstelling (negatief),
- een onderzoek dat methodisch-technisch een zwak (negatief) of juist een sterk onderzoek is (positief),
- de toegankelijkheid van de taal waarin gerapporteerd is (ontoegankelijke taal is negatief),
- of een samenvatting met aanbevelingen is opgeleverd (aanbevelingen zijn positief)
- of er steun is bij de doelgroep voor aanbevelingen (steun is positief),
- of sprake is van een uitgekende disseminatiestrategie van resultaten uit het onderzoek (positief),
- en hoe het zit met de kwaliteit en reputatie van de onderzoekers, de evaluatoren (goede kwaliteit is positief).

Te veronderstellen is dat gebruik van resultaten uit beleidsevaluaties positief beïnvloed wordt als

- de probleemstelling onomstreden is en erkend wordt in een departement en in het veld als 'relevant',
- als een onderzoek in algemene zin deugdelijk is,
- als de onderzoekers kwalitatief goed zijn en zelfs gereputeerd zijn,
- als het verslag toegankelijk is geschreven,
- als sprake is van de aanwezigheid van duidelijke en argumentatief onderbouwde conclusies met aanbevelingen die daarop aansluiten, en

- als een multimediale disseminatiestrategie gevolgd wordt waarin ook plaats is voor popularisering van uitkomsten, en
- als onderzoekers gevoel hebben voor wat speelt in een sector en het discours in de sector kennen.

Dat is nogal een lijst.

Veel open deuren?

Zien we tekorten in gebruik van beleidsevaluaties op het vlak van Buitenlandse Zaken?

Inderdaad omdat de informatie uit evaluaties een *begrensde geldigheid* hebben (Leeuw e.a., 2009). Maar dit resultaat wordt niet direct opgevoerd als variabele die verklaart waarom bepaalde resultaten uit beleidsevaluaties minder of niet benut worden.

Uit onderzoek naar beleidsevaluaties op het terrein van het ministerie van Buitenlandse Zaken blijkt nog een specifieke, afwijkende bevinding: dat extra inspanningen voor verspreiding van de uitkomsten *niet* tot meer gebruik of tot gebruik in bredere kring leiden (Leeuw e.a., 2009). Hoe kan dat? Dat komt voor een groot deel omdat het onderzoek de doelgroep bij Buitenlandse Zaken blijkbaar toch wel bereikt, zo blijkt.

We weten ook dat onderzoek *zelden alleen bepalend* is voor wat met beleid gebeurt in een proces van beleidstabieleit of verandering. Een uitzondering vormen technische analyses onder crisissomstandigheden.

Vaak moet *parallele informatie* naar voren komen die niet contrasteert met de beleidsevaluatie, willen resultaten benut worden. Als alles dezelfde kant opwijst, en dus onomstreden is, is dat positief voor de waarschijnlijkheid van gebruik van resultaten uit een beleidsevaluatie.

Ad 2 Het tweede cluster gebruik beïnvloedende factoren betreft de ‘*policy constraint variables*’. Het gaat hier om knelpunten in de hoek van beleidsvoerders. Denk aan de vraag: wat maakt aan beleidszijde de kans groot of juist klein dat de bevindingen uit een beleidsevaluatie benut worden? In principe valt hier te denken aan

- het moment waarop het onderzoek wordt gepubliceerd,
- het niveau in de organisatie waarop het onderzoeksrapport wordt ingebracht (laag of hoog?) en de routing,
- de begeleidende rol van het onderzoek door ambtenaren (is sprake van beleidsrelevante conclusies?),
- de vraag of parallel verschijnende informatie of gebeurtenissen in dezelfde richting wijzen als de onderzoekers doen,
- en de kwaliteiten van ambtenaren om de onderzoeksuitkomsten te kunnen begrijpen en interpreteren;
- daarnaast werkt gebruiksbevorderend als er een bezinningsgroep uit het beleidsnetwerk wordt ingesteld die de studie tot zich neemt en zich bezint op de vraag of en hoe hieraan een vervolg in beleid te geven.

Te veronderstellen is dat onderzoek beter benut wordt

- als het qua timing op een geschikt moment verschijnt (denk aan de drie stromentheorie van Kingdon) en dus niet te laat,
- als het onderzoek direct al hoog wordt ingebracht en dus geen lange route van onderop door de organisatie richting ambtelijke top en minister hoeft door te maken waarbij slijtage en tijdverlies optreedt,
- als de beleidsvoerders het onderzoek begrijpen en kunnen interpreteren,

- als er ook in cultureel opzicht ontvankelijk voor zijn, en
- als andere informatie in dezelfde richting wijst.

Is het mogelijk dat er op het terrein van Buitenlandse Zaken met ongeschoolden op het vlak van beleidsevaluatie *evaluatievijandigheid* optrad en waargenomen is? Leeuw c.s. (2009: 12) hebben die vijandigheid bij Buitenlandse Zaken *niet* waargenomen terwijl ze er wel naar keken. 'Medewerkers zijn vooral geïnteresseerd in bruikbare resultaten', schrijven ze.

Maar van een *leergierige cultuur* blijkt bij Buitenlandse Zaken geen sprake: medewerkers zoeken doorgaans *niet* actief zelf naar uitkomsten uit beleidsevaluaties. Op de een of andere manier moeten beleidsevaluatieve bevindingen makkelijk tot hen komen, bij voorbeeld langs publicitaire weg, eventueel doordat een uitkomst sterk de aandacht trekt, ophef veroorzaakt of een rel veroorzaakt. Onderzoek tot zich nemen concurreert blijkbaar met andere informatiestromen en beleidsvraagstukken die de aandacht vragen.

Wie resultaten uit evaluaties op de agenda wil krijgen, moet dus kennis hebben van manieren die ertoe leiden dat de aandacht van relevante beleidvoerders getrokken wordt. Een interview in een dagblad of op radio en tv strekt tot aanbeveling. Zoals een in Haagse kringen zeer bekende topambtenaar het wel eens zei:

'Als je als onderzoeker geen ophef veroorzaakt, heb je geen bijzonder onderzoek gedaan en niet veel te melden'.

De onderzoekers hebben bij Buitenlandse Zaken, ondanks positieve uitkomsten over benutting, nog *geen echt lerende organisatie* aangetroffen. Er ontbreekt een visie op wat leren in de organisatie kan betekenen. Dat is een belangrijke waarneming voor Leeuw c.s. Er wordt wel getracht met behulp van een kennisagenda, die stapje voor stapje wordt geaccepteerd, een *leercultuur* te vestigen. Dat is positief.

Ad 3 Tenslotte is er de *relationele kant* van onderzoek die van betekenis kan zijn voor het begrijpen van de mate van gebruik (of benutting) van onderzoek in beleid. Hierbij is te denken aan de aanwezigheid van een begeleidingscommissie, de samenstelling daarvan en het functioneren.

Een dergelijke commissie bevat vaak vertegenwoordigers van de opdrachtgever maar ook representanten van organisaties uit het beleidsveld, gebruikers van kennis en mensen met methodologische kennis. Een begeleidingscommissie kan meedenken over de kwaliteit van het onderzoek en de legitimatie in het veld, de timing van openbaarmaking van resultaten bevorderen, de opdrachtgever waarschuwen voor zwakheden in het onderzoek en die helpen wegnemen, meedenken met onderzoekers over analyses, conclusies en daarop aansluitende aanbevelingen.

Wellicht kunnen ook onderzoeksbevindingen de gebruikers bereiken vóór het onderzoeksrapport al verschenen is. De interactie tussen onderzoekers en beleidvoerders kan bevorderd worden door de aanbieding van het rapport aan relevante betrokkenen, de organisatie van een debat of commentaar, of door een conferentie.

Naar een lerende cultuur

Wie van de drie clusters van verklaringen? Welke van deze drie clusters factoren die van invloed kunnen zijn op een bepaalde mate van gebruik van beleidsevaluatie, van non-gebruik tot maximaal gebruik, is doorslaggevend?

De afgelopen jaren is vaak gedacht dat als je nu maar aan de kant van het *management van onderzoek* veel doet, veel bereikt kan worden ten positieve voor het gebruik van evaluaties in overheidscontexten. Denk aan een goede opzet van het onderzoek en de communicatie van onderzoeksresultaten richting doelgroep.

Maar de relatie van kennis tot de 'policy' is ook heel belangrijk (cf Leeuw en Van Gils, 2010: 167). De vraagzijde van onderzoek, daar waar de opdracht verstrekt wordt en de gebruikers zich aandienen, is voor het scheppen van condities voor gebruik wel degelijk van belang. Dat is uit een perspectief van het bevorderen van een cultuur van leren van evaluaties *heel belangrijk* om te weten. Het gebruik of zelfs het effect van evaluaties hangt namelijk ook sterk af van inhoudelijke kenmerken van evaluaties, zoals *het type kennis* dat ze opleveren, en de *beleidscontext* waarin deze evaluaties tot stand komen. Daarbij is te denken aan de mate waarin de actoren in een context al ontvankelijk zijn voor 'beleidsevaluatie als leren'. Die factor is belangrijker dan de marketing of disseminatiestrategie van de evaluatoren, menen enkele onderzoekers (Johnson e.a., 2009; Leeuw en Van Gils, 2010: 168).

Mijn conclusie is dat alle drie de clusters van invloedsfactoren in principe van belang zijn voor het begrijpen van gebruik van evaluaties en ook invloed hebben op de mate van gebruik van beleidsevaluatie. Deze mening is ook gestoeld op eigen ervaringen met evaluaties in departementen, provincies en gemeenten (Korsten, 1983; Van Oijen e.a., 1984; Mulder e.a., 1991; Pröpper, 1993, 2001). Wie resultaten van evaluatie kaal presenteert in jargon moet niet verbaast zijn als een rapport niet of slecht gelezen wordt. Wie met een persbericht en persconferentie niet in dagbladen weet door te dringen heeft mogelijk niet veel bijzonders te melden.

Hoe nu na het voorgaande een lerende cultuur te bevorderen?

Leeuw c.s. (2009) doen suggesties. En ik deed zelf onderzoek naar '*De lerende organisatie in een ambtelijke cultuur*' (Gerrichhauzen, Korsten en Heijen, 2002). Ik richtte me daarbij op een provincie en onderscheidt twee soorten leren: institutioneel leren of organisatieleren en individueel leren.

Ik deel de suggesties in twee categorieën in:

1. suggesties op macroniveau (*organisatieleren*: voor een organisatie als geheel) en
2. suggesties op microniveau (*individueel leren*: per evaluatieonderzoek).

Ik kom allereerst tot de volgende reeks macro-aanbevelingen voor een op leren gerichte cultuur.

1. *Visie*: Formuleer als politiek bestuur een kernachtige, algemene visie op een lerende cultuur in de organisatie en in netwerken (suggestie van Leeuw e.a., 2009). Schep in een visiedocument ook meer duidelijkheid over de noodzaak van evalueren en wat te verstaan is onder leren (en gebruik) van evaluaties (Leeuw e.a., 2009). Toelichting: een visiedocument is een kader waarbinnen ambtenaren kunnen werken.
2. *Evaluatie-agenda*: Een expliciete up to date evaluatie-agenda is nuttig. Op basis daarvan wordt bij voorbeeld om de drie maanden evaluatieprojecten opgevoerd, nagelopen, afgevoerd en besproken. Dat heeft zin omdat daarmee een reflectie kan plaatsvinden op meta-niveau over nut en noodzaak, omvang, prioriteit, budgettaire implicaties en kwaliteit van evaluaties van de organisatie als geheel en zijn onderdelen. Dubbelingen met de rekenkamer kunnen eruit gehaald worden. Bestuurders dienen deel te nemen aan een discussie over een evaluatieagenda vanwege hun verantwoordelijkheid, hun sturingsmogelijkheden en om te voorkomen dat iets dat nodig is geparkeerd wordt

(lees; weggeschoven wordt). De kans op de bestuurlijke uitroep 'evaluaties interesseren me niet' moet nihil zijn.

3. *Waarbij wel of geen evaluatie:* Tot een lerende cultuur behoort het zo tijdig mogelijk vanuit bestuurlijk perspectief formuleren bij welke beleids- en organisatieonderwerpen *in elk geval* beleidsevaluatie gewenst is. Toelichting: hoe eerder ambtenaren of netwerkpartners op de hoogte zijn, des te vroeger kan er worden nagedacht over de evaluatieopzet: wel of geen nulmeting vooraf?
4. *Voorkom negatieve neveneffecten van evalueren:* Voorkomen moet worden dat het aantal evaluaties van een politiek bestuurder informeel gezien wordt als een palet waarop de bestuurder afgerekend kan worden. Voorkom het motto: hoe minder je met evaluatieonderzoek te maken hebt, hoe beter het is want des te kleiner is de kans dat volksvertegenwoordigers je te grazen nemen. Het optreden van dit verschijnsel zou tegen een lerende cultuur zijn. Ook moet selectief 'shoppen' in evaluatierapporten (cherry picking) zoveel mogelijk voorkomen worden. Dit is – om een zwaar woord te gebruiken - een kwestie van bestuurlijke ethiek.
5. *Bevorder ex ante evaluaties:* Als beleid heel vaak achteraf niet de gewenste effecten of gewenste mate van effectiviteit heeft, zoals blijkt, vraag je dan als bestuur af wat de waarde is van veronderstellingen, doelstellingen en verwachtingen. Verricht vaker ex ante evaluatie want dit kan ertoe leiden dat van beleid wordt afgezien of het beleid vooraf nog bijgesteld kan worden vanuit meer gefundeerde en realistische veronderstellingen, doelstellingen of middelenkeuze. Besef dat politici vaak meer behoefte hebben aan 'onzekere daden dan aan zekere uitkomsten' en dus ex ante evaluaties niet met gejuich ontvangen.
6. *Creëer en behoud overzicht:* Ga beleidsevaluatie niet 'dicht regelen' (bureaucratiseren) maar zorg in een organisatie wel voor overzicht van wat aan evaluaties loopt, op welk beleid het betrekking heeft, wie het uitvoert en wat het kost. Volksvertegenwoordigers gaan er toch naar vragen.
7. *Kennisbevordering:* Bevorder in de organisatie de kennis van beleidsevaluatie in al zijn aspecten en dus ook van methoden en technieken van beleidsevaluatieonderzoek (suggestie van Leeuw e.a., 2009). Reserveer budget. Toelichting: alleen als er kennis is van beleidsevaluatie en men ermee enigszins vertrouwd is, zal later gebruik van resultaten plaatsvinden (Friedrich, 1970).
8. *De kennisschatkamer:* Zorg voor digitale (webgebaseerde) en andere voorzieningen om gebruik van kennis en leren te bevorderen, ook van beleidsevaluaties (suggestie van Leeuw e.a., 2009). Maak alle beleidsevaluaties vrij toegankelijk (open source), ook voor burgers. En bouw aan ervaringen. Daartoe behoort onder meer: Zorg ervoor dat het zicht op het proces of de output de blik op effecten (outcomes) niet vertroebelt (Bemelmans-Vidéc en De Vries, 2010: 134). Het komt voor dat beleidsprestaties aangezien worden voor effecten. Wie wil bevorderen dat een schouwburg bijdraagt aan verheffing en welzijn van de bevolking moet in een evaluatie *niet* volstaan met de stoelbezetting per uitvoeringen aan te geven. Want als het dan slechts gaat om populaire zanguitvoering en een beperkt maar trouw publiek blijven de programma-omvang, de diversiteit in het programma-aanbod en de spreiding in schouwburgparticipatie door bezoekers buiten beeld. Wat 'verheffing' is zal operationaal moeten worden.

9. *Maak van evaluatie geen 'verplicht nummer'*. Prestatiemeting bij de overheid is niet meer dan meting van wat tot stand is gekomen en dat zegt vaak nog niets over het behalen van beoogde doelstellingen en dus van gewenste effecten. Zorg er bovendien voor dat degenen die beoordeeld worden ('stakeholders') ook participeren in de opzet van de evaluatie. Werk niet over hun hoofden heen. Toon enige durf. En ga op zoek naar goede maatstaven en zinvolle indicatoren (Bemelmans-Vidéc en De Vries, 2010: 134).
10. *'State of the art'*: Doe aan synthese van onderzoek op bepaalde terreinen in de vorm van een 'state of the art'. Maak dus regelmatig de balans op (Leeuw e.a., 2009). Toelichting: wie bij voorbeeld vindt dat met gebiedsgericht beleid nieuwe wegen ingeslagen moeten worden, moet zorgen dat hij de recente relevante studies kent en heeft samengevat in een paper. Daaruit kan blijken of een beleidsevaluatie nog wel nodig is.
11. *Druk*: Zorg in het algemeen voor druk op de ketel bij evaluaties. Lichte druk ontstaat als bestuurders zelf interesse tonen in beleidsevaluaties. Toepassing van het horizonprincipe ('sunsets'; eindtermijnen in beleid) leidt tot enorme druk want beleid stopt na zekere tijd tenzij evaluatie aanleiding geeft tot beleidsvoortzetting in enigerlei vorm.
12. *Verantwoording over gebruik*: Zorg dat er verantwoording wordt afgelegd over de aard en mate van gebruik van resultaten van evaluatieonderzoek (suggestie van Leeuw e.a., 2009). Een bestuur kan standaard een rapport toesturen aan het controlerend bestuurlijk orgaan en elk rapport laten vergezellen door een beleidsreactie op het rapport waarin aangegeven wordt welke lessen het dagelijks bestuur trekt uit de onderzoeksresultaten.
13. *Een kennis- en discussieplatform*: Een platform kan zeker drie doelen dienen – a Het is geschikt om goed onderzoek naar voren te halen en dat biedt een oriëntatiemogelijkheid; b Er valt in een platform indirect te sturen op de kwaliteit van evaluatieonderzoek (ook Slob, 2004); c Te bevorderen is dat medewerkers zich intern en extern durven te presenteren met uitkomsten van beleidsevaluatieonderzoek. Dat leidt tot kennisuitwisseling en geeft zelfvertrouwen.
14. *Zorg voor culturele verversing*. Als iets een gewoonte wordt, gaat het frisse eraf en dreigt ritualisering. Zorg dus voor variatie en verrassing in rapportagevormen en bespreek- en discussievormen. Experimenteer met internetdiscussies. Belangrijke resultaten van evaluaties kunnen in een rapport gegoten worden maar bij voorbeeld ook in een tv- of videoformat (vol beelden, interviews).
15. *Koppeling aan hrm-beleid*: Neem bijdragen aan de bevordering van een leercultuur mee in functioneringsgesprekken en beloningsbeleid.
16. *Evaluëren leidt tot verantwoorden maar niet tot afrekenen*. Hou er rekening mee dat evaluaties zelden ertoe leiden dat schuldigen worden aangewezen. Daarvoor zijn meerdere argumenten aan te wijzen: het probleem van 'many eyes' (vele criteria waarop te beoordelen is) en 'many hands' (velen hebben een aandeel gehad in processen van besluit tot effect). Dat draagt bij aan de 'sorry-democratie' bij: 'spreek mij er niet op aan want ik heb het beleid niet voorgesteld en ook niet op de rails gezet'.

Aanbevelingen op microniveau, per beleidsevaluatie.

1. *Maatgevoel*: Het is wenselijk om geen evaluatie te verrichten die zelf duurder wordt dan het beleid zelf is. Ook als de rol van een overheid slechts die is van bescheiden cofinancier om een initiatief van derden (dat vooral door andere organisaties gedragen wordt) te financieren, kan het van wijsheid getuigen vooral geen evaluatie te verrichten of het initiatief daartoe aan derden te laten. Ook moet worden afgezien van evaluatie als een evaluatie qua timing of politieke relevantie van geen betekenis kan zijn.
2. *Crisis*: Geef direct een opdracht tot evaluatie in geval van een crisis, affaire, calamiteit, fiasco of catastrofe. Uit een oogpunt van: verantwoording en om te laten zien dat niets geheim gehouden wordt.
3. *Er zal evaluatie plaatsvinden, hanteer dan vervolgens een evaluatieladder*: Als een evaluatie nodig of wenselijk is, maak onderscheid in een lichte of meer zware vorm van evaluatie (de ladder), en in interne en externe evaluatie. Tot zware evaluatie behoren die evaluaties waarbij sprake is van een duidelijke onderzoeksopzet met onafhankelijke beoordelaars en met beoordelingscriteria.
4. *Netwerkcontext*: Evaluaties kunnen plaatsvinden in een situatie dat een beleidsvoerder alle doorzettingsmacht heeft of juist afhankelijk is. Zeker bij beleid in een netwerkcontext (van van elkaar afhankelijke actoren) moet duidelijk worden of de beleidsevaluatie in opdracht van de netwerkpartners geschiedt en ook het gebruik van de resultaten onderwerp van aandacht bij de netwerkpartners zal zijn. Als netwerkpartners later de resultaten zouden moeten gebruiken, moeten ze ook in de opdracht gekend zijn en in de evaluatie zelf aan bod komen.
5. *Kader voor evaluatie*: Voor de opdrachtverlening aan de orde is, moet duidelijk worden wat waarom hoe (in grote trekken) aan evaluatie wordt onderworpen en of doelstellingen achter een beleid uitgangspunt zijn voor de beoordeling of andere maatstaven gelden. In deze fase is bezinning heel belangrijk. Wil men een doelstellingsvrije evaluatie? Wil men een early warning evaluation?
6. *Voorafgaand aan de opdracht*: Al in de fase vóór de opdrachtverstrekking moet nagedacht worden over het soort beleidsevaluatie. Heb dus oog voor het type kennis dat evaluaties opleveren in relatie tot het gebruik ervan en de behoefte aan kennis (Leeuw e.a., 2009). Toelichting: De keuze tussen een ex ante of ex post evaluatie? In geval van een ex post evaluatie: doelbereikingsanalyses blijken waardevoller dan loutere procesevaluaties maar daarmee is niet gezegd dat dit de enige relevante variabele is voor de keuze.
7. *Sparren*: Het is nuttig als bij een opdrachtverlening betrokken ambtenaren tijdens de formulering van de opdracht of beoordeling van offertes een 'sparring partner' vinden in een interne unit voor beleidsevaluatie die expertise heeft op het vlak van beleidsevaluatie in al zijn facetten..
8. *De positie van potentiële gebruikers*: Ga ervan uit dat potentiële gebruikers geen actief zoekende houding aannemen om de resultaten van onderzoek te gebruiken voor bezinning op het onderzoek. Beleid veranderen kost namelijk moeite, geeft werk en dat wordt gemeden. Om gebruik te bevorderen moet daarom sprake zijn van grote druk.
9. *Begeleidingscommissie bij een evaluatie*: Een dergelijke commissie is doorgaans nuttig omdat daarin al bezinning met onderzoekers kan plaatsvinden op de opzet en uitvoering van het onderzoek, op wat wel of niet beleidsrelevant is, op de formulering

van conclusies en aansluitende aanbevelingen, en op het moment van verschijnen van een leesbaar rapport.

10. *Bezinningsgroep*: Komt per beleidsevaluatie tot een zorgvuldig samengestelde bezinningsgroep die nagaat wat de beleidsevaluatie inhoudt en kan betekenen en zou moeten betekenen voor organisatie en/of beleid en/of het netwerk.
11. *Voorkom rituele dansen*: Quasi-gebruik en serieuze bezinning op evaluatieresultaten kunnen door elkaar gaan lopen als standaard na een evaluatie een conferentie plaatsvindt waarop een onderzoek besproken wordt.
12. *Politieke context*. Evalueren geschiedt altijd tegen de achtergrond van het bestaan van een politieke coalitie en een oppositie. Erken dat. Weet dat kritische resultaten over een beleid genegeerd kunnen worden door coalitiepartners of opgepakt kunnen worden door een oppositiepartij. Maar weet ook dat een evaluatie zelden of nooit alleen door de oppositie gebruikt wordt. Veel representanten hebben gemeenschappelijk dat niet tot hun stijl en competentie behoort om regelmatig evaluatierapporten te lezen. Vrees als ambtenaar dus niet te gauw dat de bestuurder gevloerd wordt.
13. *Ben niet bang voor dissensus over resultaten van evaluatieonderzoek*. Onderzoek moet deugdelijk zijn. Dat streven moet voorop staan. Maar als onderzoekers iets gemist hebben of ergens een vinger niet goed achter kregen, kan dat in een volgend stadium na het uitkomen van de rapportage toch bevorderend werken op het debat. In het debat kunnen nog openingen bestaan. Een onderzoeker moet, anders gesteld, ook niet alles proberen dicht te timmeren (verg. Leeuw en Van Gils, 2010: 171).

19 DE POLITICI: WAT ZE DOEN MET EVALUATIES

Hoe verhouden politici zich tot evaluaties? Dat is de vraag in dit onderdeel. Eerst een casus.

Casus: Provinciaal reconstructiebeleid

Denkbaar is een beleidsevaluatie van wat de *Ecologische Hoofd Structuur* (EHS) in een provincie gebracht heeft? Of, wat is van het provinciale *Reconstructiebeleid* op landbouwgebied terecht gekomen?

De 'reconstructie' van de landbouwgebieden ontstond in 1997, na de varkenspestcrisis. Er kwam een Reconstructiewet. De provincie Overijssel stelde een reconstructieplan op, dat ervoor zorgde dat de ruimtelijke structuur voor de intensieve veehouderij is verbeterd. Er zijn verschillende zones gecreëerd, met ruimte voor natuur, wonen én intensieve veehouderij. Voor de intensieve veehouderij werden landbouwontwikkelingsgebieden (LOG's) aangewezen. Veehouders uit kwetsbare gebieden konden daar naartoe 'verhuizen'. Dat bood voor hen meer perspectief. Tegelijkertijd worden de kwetsbare natuurgebieden beter beschermd tegen ammoniakuitstoot. Inmiddels zijn de nodige verplaatsingen van bedrijven voor een groot deel afgerond.

Actualisatie Omgevingsvisie

Naar aanleiding van de evaluatie van het reconstructiebeleid hebben Provinciale Staten besloten dat de LOG's belangrijk blijven voor de toekomst van de intensieve veehouderij in Overijssel. De ontwikkeling van de intensieve veehouderij dient vooral plaats te vinden op bestaande agrarische locaties in de LOG's en alleen in bijzondere gevallen op nieuwvestigingslocaties. De provinciale Omgevingsvisie zal op dit punt aangepast worden.

Nut en grenzen van de LOG's

In 2011 is het nieuwe beleid in Overijssel geëvalueerd. Daar kwam uit dat de tijd rijp is om eens kritisch te kijken naar de grenzen van de LOG's. Sommige LOG's zouden wellicht kunnen worden geschrapt en andere LOG's zouden mogelijk kleiner kunnen. Tegelijkertijd zou gekeken moeten worden hoe de intensieve veehouderij in Overijssel sneller verduurzaamd kan worden.

De vraag die hier centraal staat, luidt: wat doen politici met deze en andere beleidsevaluaties?; hoe gaan ze daarmee om? Dan hebben we het dus onder andere over

- de algemene interesse van politici in opdrachten tot beleidsevaluatie;
- of beleidsevaluaties aansluiten op de controletaak van volksvertegenwoordigingen;
- hoe politici omgaan met opdrachtverlening voor een evaluatie;
- of politici kennis nemen van uitkomsten van evaluaties;
- of ze verantwoord gebruik maken van uitkomsten of ze misbruiken?

Onder politici versta ik hier leden van vertegenwoordigende organen, zoals van het parlement, provinciale staten en gemeenteraden. Leden van een dagelijks bestuur, zoals van een regering of college duid ik aan als bestuurders.

Stelling

Ik ga uit van *een stelling*:

Politici zullen gemiddeld genomen *veel belangstelling* hebben voor beleidsevaluaties en daartoe ook opdracht geven *want* als volksvertegenwoordiging behoren ze het dagelijks

bestuur zonder last en ruggespraak te controleren. Beleidsevaluaties verschaffen juist de informatie om dat controlerend werk te doen. Bovendien kunnen politici aan beleidsevaluaties inzichten ontlenuen waardoor de volksvertegenwoordiging kaderstellend kan opereren naar een dagelijks bestuur. Politici kunnen scoren met evaluaties. Ze passen bij hun doelstellingen.

Deze stelling bevat diverse items die zich lenen voor bespiegeling:

- Hebben politici echt belangstelling voor beleidsevaluaties?
- Verrichten politici vol verve hun controlerende taak, en passen evaluaties daarin? Of zitten ze vooral de bestaande coalitie in het zadel te houden of juist oppositie te voeren tegen een coalitie?
- Willen politici wel kaders stellen, bij voorbeeld in de zin van een oproep om iets te doen of te laten?
- Zijn politici bereid en in staat om de essenties van beleidsevaluatie tot zich te nemen?

Waarom de stelling niet opgaat?

Direct is al zichtbaar dat de opsteller van de stelling wel *erg optimistisch* moet zijn.

Waarom?

- Politici moeten al *veel stukken lezen*. Het is maar zeer de vraag of ze ook nog beleidsevaluatie-rapporten willen bekijken of de samenvatting daarvan.
- Bovendien kennen fracties *taakverdelingen*. Dat betekent direct al dat een verslag van een beleidsevaluatie over het algemeen nooit door alle fractieleden goed gelezen wordt maar naar een iemand.
- En, de roep om beleidsevaluatie ontstaat eerst later. Beleidsevaluatie is *niet het eerste middel* waarnaar politici grijpen. Waarom? Omdat het vertraagt. Politici kunnen intussen niet verder met het voorstellen van beleidsingrepen. Echter, andere politici kunnen juist werk maken van vertraging door een onderzoek te bepleiten.

Ik moet de belangstelling voor beleidsevaluatie dus direct al van een relativering voorzien.

Wat pro de stelling te zeggen is?

Twee argumenten;

- Subnationaal is sprake van *dualisering*. Controlerende organisaties, zoals een gemeenteraad, worden geacht een eigen rol te vervullen in de controle. Wie de controle verwaarloost, kan dat op zijn brood krijgen. Politici zouden zich dus eigenlijk niet moeten permitteren om beleidsevaluaties links te laten liggen.
- In geval van een evaluatie van een crisis is de kans groot dat de stelling opgaat. En daar staat nog iets anders tegenover. De *facebookrellen in Haren* ('Project X', 21 sept. 2012) zijn door de commissie-Cohen doorgelicht. De vier rapporten hebben de grote belangstelling gehad van de burgemeester van Haren en van vele andere burgemeesters, de politiechef, de officier van justitie, de voorzitter van de veiligheidsregio en van de gemeenteraad van Haren. Burgemeester Bats trad op basis van de rapporten af.

De volgende stap die nu door mij gezet wordt, is het bezien van enkele andere aspecten van de relatie 'evaluatie – politiek'.

Bij welke type volksvertegenwoordigers evaluatie past

1) Stelling: Beleidsevaluatie past bij de controletaak van een (controlerende) volksvertegenwoordiger

Politici zijn er in soorten en maten. Ik heb wel eens een onderscheid gemaakt in typen volksvertegenwoordigers. Dan blijkt dat we de volgende typen kennen.

De vragensteller. Dit type pakt een maatschappelijk probleem op, verzamelt er informatie over en stelt schriftelijke vragen aan een dagelijks bestuur of dit bestuur met een wantoestand of ander aspect bekend is en zo ja, wat eraan gedaan wordt.

De ombudsman-volksvertegenwoordiger. Dit type houdt een spreekuur en voert gesprekken om suggesties vervolgens te vertalen in discussies met medefractieleden of in een bepaalde commissie.

De stukjesschrijver. Type Bolkestein. Deze volksvertegenwoordiger etaleert kwaliteiten buiten overheidsgebouwen door met een bijdrage de opiniepagina's van dagbladen te zoeken en zo zichzelf en/of de partij en fractie te profileren. Dit type wordt gewaardeerd maar kan ook jaloezie opwekken.

De netwerker. Deze volksvertegenwoordiger beweegt zich breed maatschappelijk, zoekt verbindingen en kijkt waar die productief kunnen worden gemaakt. Ook dit type doet veel buiten de overheidsgebouwen en wordt gewaardeerd omdat hij 'veel mensen kent' en veel voorkeurstemmen krijgt. In de kranten kom je deze persoon niet veel tegen, op recepties des te meer.

De vergadertijger. Dit type beweegt zich vooral in de 'zaal van de volksvertegenwoordiging'. Handig met moties. Kan meerderheden smeden. Heeft veel contact met leden van andere fracties.

De bestuurder. Dit type is erg intern gericht, was eerder al bestuurder of loopt zich warm voor een nieuwe legislatuur. Kan een fractieleider zijn die de boel bij elkaar moet houden.

Het nagaan of beleid de gewenste effecten heeft, hoort in principe bij de taak van een volksvertegenwoordiging, maar wie deze indeling in typen volksvertegenwoordigers overziet, ziet nog *niet* direct dat iedereen interesse heeft in beleidsevaluaties. De netwerker zal een rapport niet veel kunnen schelen. Maar de stukjesschrijver zal er juist wel belangstelling voor hebben want in elk rapport zit vast een stukje. En de vergadertijger kan een motie smeden. Niet iedereen heeft dus vanzelfsprekend belangstelling.

De belangstelling zal natuurlijk ook mede afhangen van de uitkomst van een evaluatieonderzoek.

2) Politici zijn niet vanzelfsprekend pro beleidsevaluatie: evaluatie is een laatste stap

Politici hebben te maken met vormen van opschaling. Een kwestie moet eerst aangekaart zijn in de vorm van schriftelijke vragen. Als politici klachten ontvangen over beleid, gaan ze doorgaans vragen stellen aan een bewindspersoon of college. 'Of het de minister bekend is dat'. Zijn de antwoorden onbevredigend dan kan een mondelinge discussie volgen in een Algemeen Overleg (AO).

Zijn de antwoorden bekend en niet voldoende, dan kan een mondelinge bevraging van een bestuurder volgen en eventueel een interpellatie. Beleidsevaluatie wordt dus niet zo maar omarmd. Het is een latere stap.

Omgaan met een opdracht

3) *Een opdracht voor beleidsevaluatie: politici op subnationaal niveau dringen zelden aan*
Een ander aspect van hoe politici omgaan met een evaluatie betreft de kwestie 'of ze aandringen op een evaluatieopdracht'. Zijn politici eigenlijk wel geïnteresseerd in opdrachten voor beleidsevaluaties? Daarnaar is geen systematisch onderzoek verricht maar er is wel het een en ander over bekend.

We zien nogal eens dat als Provinciale Staten gevraagd worden onderwerpen te noemen ten behoeve van een long list van wat geschikt is voor evaluatieonderzoek het nog wel wil lukken, maar een short list is al veel moeilijker. Een zelfevaluatie-rapport van de Commissie voor Beleidsevaluatie van de provincie Noord-Brabant '*Terugblikken met het oog op de toekomst*' (2011) laat dat zien.

Over het algemeen ligt de interesse bij subnationale politici *niet direct bij beleidsevaluaties*. Om de volgende redenen.

- a) Politici zijn verkozen op een inhoudelijk program. Men wil iets realiseren. Beleidsevaluaties houden daarvan af.
- b) Beleidsevaluaties duren een tijd en vragen budget. Misschien krijg je in de periode dat je volksvertegenwoordiger bent de uitkomsten niet eens meer onder ogen.
- c) Beleidsevaluatie kunnen politiek bedreigend zijn, althans voor coalitiepartijen. Ze kunnen de hechtheid van een coalitie aantasten.
- d) Veel politici hebben geen kennis van wat beleidsevaluatie inhoudt en kunnen er geen facetten van onderscheiden.

Van alle moties die in gemeenteraden aangenomen worden, heeft slechts een heel kleine minderheid dan ook betrekking op beleidsevaluatie. Maar de interesse bij leden van de Tweede Kamer, zeg maar beroeps politici, is doorgaans groter. Zij beschikken ook over medewerkers die rapporten kunnen nalezen.

4) *Omgaan met opdrachten: waar je zit maakt uit voor waar je staat*

Als er dan toch een beleidsevaluatie moet komen, hoe willen politici een beleidsevaluatie dan inrichten?

Politici *staan zelden neutraal in een opdracht*; het coalitiebelang is nooit ver weg. Hoe zo?

Coalitiepartijen willen vaak niet meewerken aan een evaluatie omdat de opdracht wordt gezien als een ondergraving van de coalitie, als een onderdeel van het politiek steekspel. Het gevolg kan zijn dat veel oppositiepartijen een onderzoek van de grond proberen te trekken maar dat dit niet lukt. De discussie wordt dan niet in deze termen gevoerd maar wordt technisch gevoerd. Dan wordt bij voorbeeld vanuit coalitiepartijen gezegd dat 'een onderzoek nog te vroeg komt' terwijl men iets anders bedoelt.

De kans op dit gedrag is mogelijk wat groter bij tal van gemeenteraden en provinciale staten. Nader te onderzoeken.

Fractieleden in de Tweede Kamer stellen zich vaak wat *meer onafhankelijk* op en kijken naar de zaak zelf en het voortraject.

- Dat kunnen ze doen omdat een kwestie soms al lang speelt en er onder een vorige coalitie beleid ontstond. Partijen staan dan wat vrijer ten opzichte van het onderwerp.

- Een tweede reden daarvoor is dat wat niet deugt, later weer terugkeert op de agenda. Dan is het maar het beste om tijdig informatie te hebben. Dat streven past bovendien bij de controlerende rol van de Tweede Kamer.

Maar er is meer over te zeggen. Laat ik een volgende casus opvoeren.

Casus: politieke strijd over de vraagstelling

Het hondenbeleid in de gemeente Gulpen-Wittem wordt begin 2014 uitgebreid geëvalueerd. Daarbij kunnen hondenbezitters hun zegje doen, maar ook andere betrokkenen - zoals boeren en natuurorganisaties - worden bij het onderzoek betrokken. Het gaat niet alleen om de hondenbelasting maar ook om zoiets als het wel of niet 'aanlijnen' van honden en opruimen van fecaliën.

De oppositiepartijen GroenLinks, Balans, Brants, Groen! en PvdA hebben een motie ingediend waarin zij het college vragen om een enquête te houden onder hondenbezitters. Bij het vaststellen van het huidige beleid is uitgerekend die groep overgeslagen, constateren zij.

In dat beleid is onder meer vastgelegd dat honden ook op bepaalde plaatsen in het buitengebied aan de lijn moeten lopen. Met name hondenbezitters uit het dorp Mechelen ergeren zich groen en geel aan die maatregel.

De coalitiepartijen voelen niets voor een aparte enquête onder hondenbezitters. Ze hechten eraan dat alle betrokkenen straks hun mening kunnen geven over het huidige beleid. 'We vragen toch ook niet aan toeristen wat ze van onze toeristenbelasting vinden', aldus een raadslid van een coalitiepartij.

Beleidsvaluatie kan een object van strijd zijn binnen een gemeenteraad. Fracties proberen een evaluatieopdracht een bepaalde kant op te sturen of dit juist tegen te houden.

De beleidsvaluatie ligt er. En dan?

5) Politici lezen niet alles: 'information overload'

Voor politici is omgaan met schaarse tijd en inschatting wat van betekenis een belangrijke competentie. Het is qua tijdsbesteding onmogelijk voor een (sub)nationale volksvertegenwoordiger om alles te bekijken, te lezen en er ook nog een mening over te vormen. Er ontstaan daarom prioriteiten en taakverdelingen. Derhalve is het onjuist om te veronderstellen dat elke politicus alle evaluaties zou kunnen en moeten bekijken.

6) Scannen als patroon

Politici lezen rapporten die ze ontvangen gemiddeld genomen nooit helemaal maar ze lezen, als ze al lezen, gericht: wat is de kern en wat moet ik hiermee? Heeft dit politieke gevolgen en zo ja, welke eventueel? Betekent het iets voor de coalitie? Politici die *pas lid* zijn van een volksvertegenwoordiging en *ambitieu*s zijn neigen vaker tot het lezen van rapporten en tot meer lezen (Askim, 2007).

7) Begeleide evaluaties worden meer bekeken

Nemen *bepaalde* politici dan wel kennis van evaluatiestudies? En, worden ze ook gelezen of besproken?

Als een regering of college een evaluatie *van een brief voorziet met bezinning op de rapportinhoud* is de kans groot dat het presidium van de Tweede Kamer, van provinciale staten of van de gemeenteraad een studie ook agendeert en reactie uitgelokt wordt. Rapporten die zonder commentaar doorgezonden worden naar een controlerend orgaan, worden zelden behandeld. Fracties beschouwen die doorgaans als niet politiek relevant.

8) Manieren van lezen

Politiek relevante rapporten voorzien van een 'oplegnotitie' worden dus wel gelezen, maar wat is lezen? Rapporten zijn grofweg op tien manieren te lezen. Ik noem enkele mogelijkheden:

1. de omslag bekijken en het rapport weggooien ('niet voor mij van belang');
2. de omslag bekijken en het rapport aan een ander geven voor wie het(meer) relevant is of lijkt;
3. de omslag bekijken en het rapport opbergen voor gebruik op het goede moment;
4. de samenvatting even 'scannen' op inhoud en betekenis;
5. de samenvatting scannen en de conclusies en aanbevelingen bekijken; daarna weer weg leggen;
6. het rapport doorbladeren en per hoofdstuk even bezien wat de kern is;
7. het hele rapport grondig lezen en de aanbevelingen bezien;
8. het rapport lezen, de kwaliteit beoordelen, en het rapport gaan bespreken in eigen kring;
9. naar aanleiding van het rapport een inbreng voorbereiden (samenvatting maken, speech opstellen, motie formuleren etc.);
10. enz.

Het is niet op voorhand een gegeven dat leden van oppositiepartijen evaluatierapporten meer benutten om een dagelijks bestuur te bestoken dan coalitiepartijen (Askim, 2007). Ook kan niet worden gezegd dat ze diepgaander kennis nemen van resultaten. Maar *jonge* volksvertegenwoordigers hebben iets 'gretigers' dan routiniers volgens de Noorse onderzoeker Askim.

De aard van de kwestie doet ertoe

9) Gebruik van evaluaties van een affaire of crisis

Het gebruik van resultaten van evaluatiestudies is groot als een evaluatie plaatsvindt van een crisis en de betreffende overheid een rol heeft in de 'kwestie' en een speciale commissie door een vertegenwoordigend orgaan aan het werk gezet is. Dat weet ik uit gesprekken met een aantal betrokken commissievoorzitters zoals Marten Oosting en Gerd Leers. En het is af te lezen uit reacties van volksvertegenwoordigers in de media en uit verslagen van vergaderingen van controlerende organen.

Aan welke crises is te denken? Bij voorbeeld aan de verzelfstandiging van het Gemeentelijke Vervoerbedrijf in Amsterdam en de financiële perikelen daaromheen (commissie-Van der Zwan), de vuurwerkramp in Enschede (commissie-Oosting), de cafébrand in Volendam (commissie-Polak/Versteden), de Herculesramp in Eindhoven, de Schipholbrand in Schiphol, de ontploffing bij Chemo Pack in Moerdijk, de asbestaffaire in Utrecht (commissie-Jansen), aardbevingen als gevolg van aardgasboringen in Groningen (rapport van Toezicht op de mijnen), de bankencrisis (commissie-De Wit), de SNS-bank, examenvraagstukken bij InHolland (commissie-Leers), financiële perikelen bij onderwijsinstelling Amarantis (commissie-De Wit; commissie-Halsema), facebookrellen in Haren (commissie-Cohen), etc.

Deze speciale commissies worden geacht ook aandacht te besteden aan de rol van overheden zelf, in casu een of meer dagelijks besturen. Een crisis leidt vaak tot speciale debatten die gewijd zijn aan de crisis.

Adviezen

10) *Evaluaties van een affaire of crisis kennen als positief punt prospectie*

Rapporten die gaan over evaluaties van een crisis worden gebruikt omdat ze bijna altijd prospectieve beschouwingen kennen. Ze gaan over 'wat is nodig opdat voorkomen wordt wat gebeurd is?' Een snelle analyse van debatten die gewijd zijn aan evaluaties van crisis maakt duidelijk dat de discussie gaat over 'hoe ver heeft het zo kunnen komen?' en over 'hoe bereiken we dat dit nooit meer gebeurt?' en over 'wat moeten we nog meer doen?' ('wat wordt nu het nieuwe beleid?')

11) *Evaluaties van prioritaire kwesties worden benut*

Politici stellen op voorhand (meer) belang in een evaluatiestudie als een kwestie prioriteit heeft, het financiële belang groot is en op voorhand al sprake lijkt te zijn geweest van een beleidsfiasco. Denk aan een commissie grote projecten (commissie-Duivesteijn).

12) *Bepaalde controlerende organen verrichten zelf een evaluatie van prioritaire kwesties of crises*

Als een kwestie belangrijk is, zal op landelijk niveau een vertegenwoordigend orgaan zelf onderzoek doen. De Tweede Kamer besluit dan tot een onderzoek door een eigen eenheid, of tot een parlementair onderzoek en stelt een commissie in. Levert dit onderzoek niet genoeg op dan volgt een parlementaire enquête, eveneens onder leiding van een commissie. De Tweede Kamer werkt dus trapsgewijze. Een onderzoekscommissie beschikt dan over een aanzienlijk budget, trekt een deskundig secretariaat aan, besteedt een deel van het onderzoek uit en gaat eerst daarna bepaalde personen horen in hoorzittingen. Partijbelangen worden zoveel mogelijk uitgebannen. De commissie opereert volstrekt onafhankelijk. Voor participatie in ene commissie gelden strikte regels en daar hoort een ontwikkeld moreel besef bij (do's en don'ts).

13) *Ongeprogrammeerd onderzoek*

De grote betrokkenheid van een volksvertegenwoordiging en volksvertegenwoordigers bij evaluaties van crises impliceert dat een dagelijks bestuur (college, kabinet) vooral niet alles moet (laten) onderzoeken.

Het belangrijkste en meest gebruikte evaluatieonderzoek, opgevat als onderzoek dat tot groot en breed politiek en maatschappelijk gebruik leidt, is *ongeprogrammeerd onderzoek*, dus evaluatie die niet voorkomt op een vooraf door de volksvertegenwoordiging geaccordeerde lijst van onderwerpen voor evaluatieonderzoek. Logisch want alleen al alle crisisgerelateerd onderzoek is ongeprogrammeerd.

14) *Gevaar van overreactie*

Evaluaties van crises met prospectieve onderdelen kunnen ook averechts werken want tot een te snelle en daadkrachtig ogende reactie van controlerende organen leiden. Soms kan een dagelijks bestuurder zich niet hieraan onttrekken. Dit bleek ten aanzien van de vuurwerkramp in Enschede. Een aantal maanden nadat 'de rook was opgetrokken' en de debatten hadden plaatsgevonden moest een minister (Donner) corrigerend optreden en eerdere voortvarendheid enigszins corrigeren. Hij had te snel toezeggingen gedaan.

Advies: Hebben deze inzichten implicatie voor 'evalueren als leren'? Ja, programmeer niet alles.

Conclusies

Dit stuk ging over 'the politics of evaluation', over de verhouding tussen politiek en beleidsevaluaties.

Beleidsvaluatie passen op het oog heel goed bij het werk van een volksvertegenwoordiging. Want een dergelijk orgaan moet controleren, en sinds de wettelijke dualisering is dat een

belangrijker taak geworden. Het verwaarlozen van de controlerende taak is maatschappelijk schadelijk want daardoor kan ondoelmatigheid optreden. Als politici niet hebben opgelet, kan het aanzien van de politiek schade oplopen. Politici doen er dus goed aan om aan te dringen op beleidsevaluaties, vooral van beleid waarvan men vermoedt dat het niet de doelstellingen bereikt of wel maar op een ondoelmatige manier. Of waarvan men ongewenste neveneffecten denkt te zien. Die evaluaties vegen de stoep ook schoon van volksvertegenwoordigers. Zij kunnen claimen tijdig gewaarschuwd te hebben.

Toch is er ook een andere kant. Beleidsevaluatie en politici hebben *geen* vanzelfsprekende en gemakkelijke verhouding tot elkaar. De coalitie moet in stand gehouden worden en dat werkt in op opdrachten. Soms houden coalitiepartijen evaluaties af omdat men vermoedt dat de oppositie een evaluatie alleen gebruikt om de coalitie te ondergraven en het dus niet alleen maar om de zaak zelf gaat die aan evaluatie wordt onderworpen. Bovendien, binnen fracties bestaan stijlen van vertegenwoordiging en taakverdelingen. Beleidsevaluaties passen meer bij een bepaalde stijl dan een andere. Gebruik van rapporten wordt voorts geremd door taakverdelingen in portefeuilles, en het feit dat vertegenwoordigers overladen worden met informatie. Tal van beleidsevaluaties worden daarom door een groot aantal volksvertegenwoordigers niet of matig gelezen.

Evaluaties worden wel gebruikt als het gaat om een grote zaak, als vele fracties een opdracht steunen en een commissie uit de volksvertegenwoordiging zelf het onderzoek verricht met steun van ambtenaren en anderen. Dat kan een Tweede Kamer, samengesteld uit fulltime volksvertegenwoordigers, echter makkelijker doen dan provinciale staten of een gemeenteraad, die bestaan uit leden met een andere hoofdbetrekking. Provinciale staten en gemeenteraden kiezen er daarom vaak voor om het dagelijks bestuur een evaluatie te laten verrichten, die dan uitbesteed wordt. De figuur van een gemengde commissie, van externe leden en volksvertegenwoordigers, wordt ook wel toegepast.

20 LEREN DOOR EVALUEREN EN WAT LEREN BELEMMERT

Het uitvoeren van een beleidsevaluatie kan ertoe leiden dat kanttekeningen of zelfs vraagtekens worden geplaatst bij de resultaten en effecten van een beleidsprogramma. In hoeverre zijn de doelstellingen niet gehaald en waar ligt dat aan? Valt daar nog wat aan te doen? Zo ja, wat? Misschien wordt het beleidsprogramma kritisch tegen het licht gehouden of gaat het op de schop of wordt het gestopt? Daarmee komen we bij leren door evalueren (Bekkers, 2007). Wat betekent het om te spreken over 'leren'? En wat belemmert leren?

Omschrijving van beleidsleren

Leren verwijst naar verbetermanagement. Leren is te omschrijven als

'een proces waarbij actoren trachten hun handelingen te verbeteren op grond van kennis, ervaring en inzicht' (Bekkers, 2007: 330).

Beleidsleren is te omschrijven als

een proces van ontwikkeling en aanpassing dat gericht is op het systematisch verbeteren van het functioneren van organisaties die een rol spelen in beleidsprocessen, zoals de vorming en uitvoering van beleid, en vooral van het beleid zelf.

Bij beleidsleren zijn dus enkele ambities aan de orde:

- Systematisch handelen;
- Een verbeterambitie;
- Een verbeterambitie bij de organisatie, die gaat veranderen;
- En een verbeterpoging gericht op het beleid zelf.

De *essentie* bij beleidsleren is dat beleid gericht verbeterd wordt, zowel in denken als doen, aldus In 't Veld en Van der Knaap (1994; ook Bekkers, 2007: 330).

Wanneer dringt beleidsleren zich op? Wanneer vindt het plaats? Als de veronderstellingen achter beleid op gespannen voet staan met de beleidsdoelstellingen en bereikte resultaten (Bekkers, 2007: 334).

Een voorbeeld. Wielrenners dachten vroeger vaak dat als ze maar veel kilometers traiden en uren maakten, hun uithoudingsvermogen groter zou worden en hun conditie beter. Dan werden ze beresterk. Michel Boogerd trainde wel eens 300 kilometer op een dag. Tot ontdekt werd dat degenen die een week voor een grote koers rust namen daar ook veel baat bij konden hebben. Dat bracht het begrip 'dosering' aan het licht en overtraint zijn. Teveel trainingsarbeid verrichten kan ertoe leiden dat men niet meer 'vooruit te branden is'. Dat inzicht van doseren duidt op 'leren'.

Aspecten van leren

Enkele aspecten van leren zijn nader te bezien.

1) *Leren als vermogen, richting of resultaat?*

Is beleidsleren doenlijk of komt het echte leren pas door en na bij voorbeeld een crisis? Ik onderken drie posities:

- competenties hebben om te verbeteren (leren als leerhouding),
- leren als een richting inslaan om te verbeteren,
- leren als een daadwerkelijk verbeterresultaat.

Er wordt wel opgemerkt dat 'echt verbeteren' een overtrokken ambitie achter beleidsleren is; het is al heel wat als leren *het vermogen* impliceert om zich aan te passen. Dan zou het dus al wat zijn als medewerkers over competenties beschikken om zich aan te passen. Denk aan het verrichten van alarmoefeningen. Terwijl anderen toch ook graag iets verder willen gaan en iets willen zeggen over *de richting* van de verandering en weer anderen zelfs over *het bereiken van een resultaat dat op leren duidt*.

Dat verbeteren in termen van resultaten kan verschillende vormen aannemen. Het verbeteren, kan bij voorbeeld

- doordat de doelstellingen worden aangepast (meer haalbaar)
- of de instrumentenmix wordt veranderd
- of in bredere zin de veronderstellingen over een verschijnsel waarop het beleid zich richt, veranderen;
- leren kan ook bestaan uit gericht afwachten (spelen met tijd)
- of het gedeeltelijk of integraal stoppen van beleid.

2) *Leren is contextgebonden*

Een tweede aspect aan leren slaat op de contextgebondenheid. Bij wijze van spreken: het belang van boorputten slaan voor het vinden van water betekent in de Sahara iets anders dan op de Veluwe.

We weten inmiddels dat leren contextgeboden processen zijn, dus worteling hebben in bepaalde lokale praktijken van actoren, eventueel in beleidsnetwerken. Wat je aan beleid kunt veranderen, heeft dus ook te maken met de beleidscontext. Is sprake van een wet die verstrengeld is met andere wetten dan is een verbeterpoging in radicale zin zeker niet eenvoudig. Het bestaande determineert gedeeltelijk de kans op het nieuwe.

3) *Leren is een permanent proces; met leren ben je niet zo maar 'klaar'*

Een derde aspect aan leren slaat op het eindige of oneindige. Leren is in beginsel een oneindig proces en ook een iteratief proces, iets van *vallen en opstaan*. Leren is niet iets van 'even doen' maar meer iets dat in beginsel niet ophoudt maar van min of meer permanente aard is. Er zijn weinig argumenten te geven om te zeggen: 'we schakelen het leren even uit'.

4) *Leren kan bewust en expliciet plaatsvinden*

Leren kan nadrukkelijk en zelfs ongevraagd plaatsvinden doordat een overheid voor het blok gezet wordt, bij voorbeeld door een enorme crisis. Stel een dijk breekt door en het toestromende water verwoest tal van huizen. Er is dan geen vrije keus. De dijk zal gedicht moeten worden.

Leren kan ook in een rustiger context plaatsvinden, zeg maar gaandeweg doordat tijdens de uitvoering begeleiding plaatsvindt. Denk aan ervaringen met pilots bij de invoering van dualisering van het gemeentebestuur. En we kennen leren door 'trial and error'.

De rol van evalueren in beleidsleren houdt in dat leren op basis van een beleidsevaluatie bewust en expliciet kan plaatsvinden (In 't Veld en Van der Knaap, 1994).

5) *Leren op niveaus*

Wie leert er? In de literatuur over leren wordt een onderscheid gemaakt in individueel leren, organisationeel leren en leren in een beleidssector, aldus Victor Bekkers (2007: 331).

Individueel leren verwijst naar leren van een individu op basis van eigen ervaring, opgedaan in of buiten een organisatie waartoe het individu behoort.

De cultuur van een organisatie zegt iets over de heersende werkwijzen, opvattingen en houdingen. Is sprake van openheid of van geslotenheid? Is er sprake van de mogelijkheid om alternatieve visies te uiten? Wordt aangemoedigd om beleidsevaluaties te verrichten? Organisationeel leren heeft betrekking op het scheppen van condities in een organisatie om zich aan te passen aan veranderende omgevingsomstandigheden. Bij voorbeeld om meer beleidsevaluaties te starten.

Een derde niveau van leren verwijst in het voetspoor van het werk van Paul Sabatier naar beleidssectoren. Zijn er heersende gewoonten in een sectoraal beleidsnetwerk?

Leergradaties

Een bekend onderscheid is dat in *leren van de eerste, tweede en derde orde*. Veranderingen worden ook wel als zodanig aangeduid. Ik licht dat toe.

Bij *eerste orde-leren* of *aanpassingen van de eerste orde* staan de doelstellingen en beleidsinstrumenten *niet* ter discussie. Getracht wordt de toepassing van instrumenten te verfijnen of de voorwaarden waaronder beleid is ingezet te verbeteren. Het gaat om beleid-bijstukjes-en-beetjes. Denk aan de uitvoering van beleid met meer of minder budget of het geven van voorbeelden voor hoe beleid opgepakt kan worden door een begeleidend instituut. Dergelijke aanpassingen kunnen aan de orde zijn als beleid nog niet lang wordt toegepast of een overheid zich verkeken heeft om de omvang van het aantal gebruikers van een regeling.

Veranderingen van de *tweede orde* hebben betrekking op de beleidsinstrumenten. De doelstellingen achter het beleid staan niet ter discussie. Maar er wordt wel gesleuteld aan de beleidsinstrumentenmix. Eventueel wordt een instrument toegevoegd. Zo wordt soms wel getracht om door middel van voorlichting een regeling nog eens goed onder de aandacht te brengen.

Veranderingen van de *derde orde* hebben juist *wel* betrekking op wijziging van doelstellingen.

Denk maar aan de discussie of beleid van het kabinet-Rutte II gericht moet zijn op het halen van een 3 procent(schuld)norm of juist de economie moet worden aangejaagd en dus een iets grotere schuld aanvaardbaar is.

In geval van derde orde leren kan sprake zijn van een *paradigmawisseling* (Bekkers, 2007: 332). Een beleid op een bepaald gebied wordt wel gezien als een dominant '*frame*'. Een alternatief frame kan dan proberen de hegemonie te doorbreken. Zo is de landbouw in agrarische gebieden wel als een rationele aanpak betiteld op basis van groot, groter, grootst. Maar een aanvechtend frame biedt meer ruimte voor betrokkenheid van agrariërs bij landschapsbehoud of biologische landbouw (Korsten en Leers, 2005).

Hogwood & Peters (1983), Brewer & deLeon (1983), Rose & Davies (1994) en Baumgartner & Jones (2002) hebben aangegeven dat beleidsveranderingen verschillende vormen kunnen aannemen. Ik noem:

- *Beleidsonderhoud* ('maintenance'). Het bestaande beleid wordt slechts beperkt veranderd door een wijziging van de middelen, inclusief het volume, en de doelgroepen waar het beleid betrekking op heeft. Kwestie van eerste orde leren. Weinig opzienbarend. Denk aan de Deltawerken die in een fase van onderhoud belandden.

- *Beleidsinnovatie* ('innovation'). Dit beleid verwijst naar nieuw beleid op een nieuw, nog onontgonnen beleidsterrein. Denk aan *randgroepjongerenbeleid* in 1950, *klimaatpolitiek* in 1950 of *hardhoutbeleid* van gemeenten in 1950. Dat beleid bestond toen nog niet. Van leren zal dan vaak nog geen sprake zijn geweest. Beleidsevaluatie zal zelden tot beleidsinnovatie in deze zin leiden.
- *Beleidsopvolging* ('succession'). Beleidsopvolging wordt in het Engels aangeduid als 'policy succession'. Er zijn verschillende vormen van bekend, zoals beleidsvervanging ('replacing'; oude schoenen inruilen voor nieuwe), beleidssplitsing ('policy split'; een beleid wordt uiteengelegd in afzonderlijke beleidsprogramma's, zoals op het gebied van milieubeleid geschiedde) of beleidsintegratie ('policy integration'; zoals ook op het gebied van milieubeleid geschiedde; Evaluatiecommissie Wet milieubeheer, 2003). Hier is beleidsevaluatie volop aan de orde.
- *Beleidsbeëindiging* ('policy termination'; Daniels, 1997). Een beleid wordt integraal stop gezet of gedeeltelijk ingekrompen. Denk aan het schrappen van de winkelsluitingswet of BKR-regeling voor kunstenaars. Een dergelijke termination kan optreden bij de volgende combinatie: als het ideologisch tij veranderd is, er bezuiniging aan de orde is en een beleidsevaluatie negatief uitpakt. Beleidsbeëindiging kan resulteren in een 'policy residue', een beleidsrest. Zo werd een tijd na het stoppen van het prijzenbeleid nog een afdeling Prijzen bij het departement van Economische Zaken aangetroffen. Een dergelijke kleine eenheid kan van nut zijn als het beleid later toch weer opgestart wordt. Een beleidsrest treft men wel aan bij NASA-achtige organisaties die een ruimtevaartprogramma hebben stopgezet.

Er komen ook *mengvormen* voor, zoals een mix van nieuw beleid en partiële beëindiging die overgaat in beleidsopvolging. Als het ware: je ruilt sportschoenen in tegen schoenen waarop je ook wel kunt lopen maar ook balletoefeningen kunt doen.

Figuur 10: Kans op beleidsevaluatie in relatie tot typen 'policy change'

Ideaaltype beleidsverandering	Frequentie van voorkomen	Toelichting	Kans op rol voor beleidsevaluatie	Aard van leren
Beleidsinnovatie	Komt weinig voor. Klimaatpolitiek in 1950. Techn. innovatiebeleid in 1960.	Beleid op <i>onontgonnen</i> terrein. Beleid zonder biografie.	<i>Ex ante</i> evaluaties; <i>Weak signal</i>	Leren is nauwelijks aan orde.
Beleidsonderhoud	Komt voor. Deltawerken.	Beleid met biografie: beheerfase	Heel beperkt ex post evaluatie	Eerste orde. Doelstellingen en instrumenten staan niet ter discussie. Na verloop van tijd radicaal tweede en derde orde leren. Voorbeeld: Deltacommissaris. Marktwerving
Beleidsopvolging	<i>Komt zeer veel voor</i>	'Replacing'; beleidsuitsplitsing; beleidsintegratie	<i>Grote kans op ex post evaluatie en toekomstperspectief</i>	Eerste, tweede of derde orde leren. Bijvoorbeeld Omgevingswet
Beleidsbeëindiging	Komt beperkt voor	Beleidsrest?	Onmisbaar	Derde orde: geheel andere kijk op een zaak

Waardoor deze veranderingen in beleid optreden, zegt de typologie van vier vormen van 'policy change' niet.

Vaak wordt aangenomen dat veranderingen optreden als '*geloofsovertuigingen*' die aan de basis van een beleid liggen, zich wijzigen. Dat betekent dat de veronderstellingen achter het beleid veranderen, met andere woorden 'de beleidstheorie' is bij bestuurders, ambtenaren, burgers, bedrijven en/of anderen aan erosie onderhevig geraakt (Leeuw, 1993; Parsons, 1995: 592; Van Heffen, 1996; Leeuw, 2008).

Hoe het leren plaatsvindt en hapert

Leren kan op meerdere manieren:

1. leren als cybernetisch leren,
2. leren als gemeenschappelijk leren (Bekkers, 2007: 334-337).

Ad 1. Allereerst is er de *cybernetische kijk* (Bekkers, 2007: 334). Dan gaat het – zoals Ig Snellen het wel eens heeft voorgesteld - om een thermostaat die de kamertemperatuur meet en bij een bepaalde waarde 'aanslaat' waardoor de verwarmingsapparatuur weer gaat branden om het gewenste peil te bereiken. Zo wordt het gezegd. Een besturend orgaan kan ook ongeveer zo werken. Dan heeft de overheid een *monitorsysteem* en dat systeem alarmeert als het ware als er iets in het bestuurd orgaan gebeurt waardoor iets onder een norm zakt (De Kool, 2007a, b en c).

Het simpele leren is dat van een zender naar een ontvanger waarbij de ontvanger als het ware niet terugpraat. Dat past in de cybernetische kijk. Leren is dan *overdracht van informatie* die leidt tot *absorptie* (Bekkers, 2007: 335). Dit leren kan haperen.

- Leren kan stagneren als *de zender* gebrekkig is. De zender heeft misschien een zwakke reputatie of brengt de boodschap niet goed.
- Er kan ook iets mis zijn met *het medium*, dus met de drager van de boodschap. Denk aan ruis op de lijn of een medium (zoals tekst, radio, tv, internet, etc.) dat niet het goede is. Lager opgeleiden zijn ontvankelijk voor beelden en niet alleen of voornamelijk voor geschreven tekst.
- Maar ook aan de kant van *de ontvanger* kunnen zwakheden zitten, zoals een niet goed afgestelde antenne, gebrekkige ontvankelijkheid, niet opletten, geen interesse, etc.

Binnen de cybernetische optiek bestaan twee leerniveaus: *single loop learning* en *double loop learning*. Dit zijn termen die ontleend zijn aan het werk van Chris Argyris. Wat wordt daarmee bedoeld?

Single loop learning impliceert dat er feedback op basis van vergelijking plaats vindt van output en outcomes (effecten) naar oorspronkelijke doelstellingen. Als er een afwijking optreedt die noopt tot en ook leidt tot een verbeteractie in de vorm van een interventie om alsnog de doelstellingen te bereiken, is sprake van *single loop learning* (Bekkers, 2007: 335). Het gaat dan bij de correctie om meer middelen (mensen, budget, etc.) of andere middelen, verbeterde handhaving en toezicht. Eerder is gesproken over eerste en tweede orde leren. Dat past hierin. Ook het beeld van de thermostaat past hierin.

'*Double loop learning*' graaft een paar spades dieper. Een thermostaat kan niet aan *double loop learning* doen. De veronderstellingen waar het beleid op gebaseerd is, komen ter discussie. Dan is daarom aan de orde om doelstellingen en middelen ter discussie te stellen. De

beleidsmaker gaat een nieuwe beleidstheorie ontwerpen. Victor Bekkers (2007: 335) spreekt van probleemherziening, betere informatie over het probleem en wat er achter schuil gaat en nieuwe normen.

Figuur 11: Voorkomen van leergradaties

Type leren	Implicatie	Frequentie van leren	Wanneer analyse	Wat doen
Single loop learning	Doelen en/of beleidsinstrumenten aanpassen	Komt voor maar padafhankelijkheid belemmert	Incidenten/ media/ geen effecten	Ingreep na verkiezingen?
Double loop learning	Veronderstellingen achter beleid ter discussie	Komt minder vaak voor	Na crisis, te laat dus	De crisis vóór zijn

Ad 2 Leren als gemeenschappelijke beeldvorming

‘Beleidsleren kan worden gezien als de uitdrukking van gemeenschappelijke beeldvorming over hoe de werkelijkheid in elkaar steekt, hetgeen leren bij uitstek tot een sociale activiteit maakt’, aldus Bekkers (2007: 335). Dat betekent het volgende.

Een overheidsorganisatie heeft zelden alleen alle doorzettingsmacht. Deze organisatie beschikt over taken, bevoegdheden en middelen maar ook andere organisaties beschikken daarover. Wie de verdroging of vernatting van De Peel wil aanpakken heeft te maken met zo een omstandigheid (Bekkers en Lips, 1998). Rond dit vraagstuk bevinden zich meerdere departementen, enkele provincies, waterschappen, gemeenten, grondeigenaren en pachters. Ze zijn tot elkaar veroordeeld om stappen te kunnen zetten. Ze moeten zich verenigen in een gezamenlijke aanpak, eigen posities niet willen maximaliseren. Dat betekent dat doelstellingen verenigd moeten worden en middelen bij elkaar gelegd.

Hier blijkt de noodzaak van *collectieve interactie*, die kan worden geïnterpreteerd als leren omdat niemand eenvoudig even de eigen positie kan maximaliseren. Deelnemers worden geacht van en met elkaar te leren, zo dat een *dialogo van doven* vermeden wordt (Van Eeten, 1999). Er is de noodzaak om een gemeenschappelijke taal of referentiekader te ontwikkelen om elkaar te verstaan en betekenissen toe te kennen aan fenomenen. In deze optiek is aan de orde om te komen tot gedeelde opvattingen en ervaringen in de vorm van informatie, kennis, inzichten, gedachten, opvattingen, beelden. Een dialoog is nodig. De zgn. ‘*Alderstafel*’ Schiphol, vernoemd naar oud-minister Hans Alders - die hiervan voorzitter is -, is als zodanig te beschouwen.

Casus Alderstafel

De Alderstafel Schiphol is ingesteld naar aanleiding van de evaluatie van het Schipholbeleid en het kabinetsstandpunt van april 2006 dat de beschikbare milieuruimte benut en kansen voor hinderbeperking verzilverd moeten worden. Later zijn de Alderstafels Eindhoven en Lelystad daar bij gekomen.

De Alderstafel Schiphol heeft in juni 2007 haar eerste advies uitgebracht over de ontwikkeling van Schiphol en de regio op korte termijn, tot en met 2010. Hierin stond dat er maximaal 480.000 vliegtuigbewegingen van en naar Schiphol plaats mochten vinden in 2010. Vanaf 2008 is een tweede overlegronde gestart die een advies voor de middellangetermijn aan het kabinet uit heeft gebracht, in oktober 2008. Aan de Alderstafel Schiphol overleggen de partijen over de implementatie en uitvoering van het advies, in het bijzonder over de uitwerking van het nieuwe geluidsstelsel en de hinderbeperkende maatregelen die afgesproken zijn.

Deelnemers aan de Alderstafel Schiphol

De volgende partijen nemen deel aan de Alderstafel Schiphol:

Bestuurlijke Regiegroep Schiphol, namens deze: Gemeente Amstelveen, Amsterdam, Castricum, Haarlemmermeer, Gemeente Uitgeest; Provincie Noord-Holland, Provincie Zuid-Holland.

Ministerie van Infrastructuur en Milieu.

Schiphol Group.

Luchtverkeersleiding Nederland.

KLM.

BARIN.

Bewonersvertegenwoordigers Commissie Regionaal Overleg luchthaven Schiphol (CROS).

Bewonersvertegenwoordigers Vereniging Gezamenlijke Platforms (VGP).

De VGP is in 2007 toegetreden naar aanleiding van een evaluatie van het proces rond het kortetermijnadvies.

De BARIN is in 2012 toegetreden naar aanleiding van een verzoek uit de Tweede Kamer waarop door de voorzitter een advies is uitgebracht over de vertegenwoordiging van de overige luchtvaartmaatschappijen.

Processen van collectief leren dienen zich vooral als noodzakelijk aan in geval van *'intractable problems'*, zeg maar weerbarstige problemen (of *'wicked problems'*) in een netwerkcontext. Denk aan het omgaan met geluidshinder op Schiphol en voetbalvandalisme (Abma en in 't Veld, 2001). *'Wicked problems'* zijn problemen waarover de kennis om die op te lossen beperkt is, de actualiteit niet beperkt is tot een discipline, kennis verspreid is en er niet vanzelfsprekend overeenstemming bereikt wordt tussen relevante actoren. Er bestaan hierbij volop kansen op blokkades en fixaties, en die moeten vermeden worden.

Belemmeringen voor leren

Een bekende theorie die iets zegt over belemmeringen over leren is groupthink (Bekkers, 2007). De psycholoog Irving Janis stond aan de wieg van deze theorie. Hij bestudeerde cases zoals Pearl Harbour. Ook andere cases zijn bekeken zoals Watergate.

Casus: Pearl Harbour. De Amerikaanse admiraal Kimmel voerde de Amerikaanse strijdmacht aan tegen de Japanners. Kimmel en zijn mensen lagen 'voor anker' in de haven van Pearl Harbour en waanden zich daar veilig. Ze hielden er geen rekening mee dat een onderzeeër de haven zou kunnen bereiken en hen zou kunnen bedreigen. Toch gebeurde dat. Kimmel c.s. negeerden de signalen die een medewerker. Die dacht dat de Japanners wel degelijk voor de haven lagen. Wat juist bleek.

Kimmel had vergeten tegenspraak te organiseren. De mensen om Kimmel heen waren loyaal maar tegelijk ook te volgzzaam. Ze hadden geen oog voor alternatieve informatie.

Kenmerken of symptomen van groepsdenken zijn ('t Hart, De Jong en Korsten, 1991: 28).

1 Zelfoverschatting

- Illusie van onkwetsbaarheid
- Superioriteit van de groep/zelfvertrouwen

2 Collectieve vernauwing

- Negeren nieuwe info
- Stereotypering anderen

3 Conformeringsdruk

- Zelfcensuur
- Illusie van unanimiteit

- Druk op dissidenten
- Mindguard/gedachtenhoeders/Ideeënpolitie

Figuur 12: Belemmeringen voor beleidsleren

Type belemmering	Toelichting	Hoe voorkomen	Uitwerking
<i>Tolerantie voor suboptimaliteit</i>	Bepaalde zwakten accepteren en door de vingers zien	Niet	
<i>Groepsdenken: waarom we ons confirmeren aan de groep en de leider</i>	<p>1 Kleine hechte groep, sterke leider</p> <p>2 Bestaand beeld dominant, vooral onder druk; blind voor tegenargumenten; halstarrigheid</p> <p>3 Afwijzen van afwijkingen, kritiek gesmoord; geen zoeken naar wat afwijking betekent</p> <p>4 Groepsdruk op volgzzaamheid</p> <p>Uitingsvorm: Zelfoverschatting leider; roekeloosheid; overoptimisme</p>	<p>1 Leider moet tegenspraak organiseren;</p> <p>2 Aanstellen van advocaat van de duivel (rouleren);</p> <p>3 Permanent denken in alternatieven toestaan;</p> <p>4 Besluitvorming onthaasten</p> <p>5 Concurrerende groepen instellen</p> <p>6 Experts van buitenaf</p> <p>7 Na (aanstaande) consensus second opinion</p>	<p>Janis; 't Hart, De Jong en Korsten, 1991.</p> <p>Voorbeeld: biometrisch paspoort</p>
<i>Padafhankelijkheid: op zelfde pad blijven lopen, niet in staat tot flexibiliteit</i>	Voorbeeld: Opstelling Russische militairen ten tijde van Cuba-crisis. Men blijft zich gedragen zoals altijd ook als omstandigheden ander gedrag vragen.	Paradigmaverandering durven doordenken. Tegenspraak organiseren.	Allison
<i>Niet toestaan van variëteit</i>	Denk aan communistische regimes	Experimenteren met beleid. Check op veronderstellingen.	Van Gunsteren

Het nadeel van deze theorie is dat die vooral betrekking heeft op besluitvorming onder druk in kleine groepen. Denk aan gijzelingsacties.

Conclusie

Voorgaand overzicht bevat het instrumentarium om in een organisatie na te gaan of van leren doro evaluatie sprake is en wat de aard daarvan is. Tal van begrippen passeerden de revue. Eerder is een 'concurrerend' overzicht verschaft. Men kan ook kijken naar gebruik en benutting van evaluaties op basis van tradities, zoals 'the utilization of research'.

DEEL VII: CAPITA SELECTA EN SAMENVATTING

21 DENKPISTES: BELEIDSEVALUATIE IN RELATIE TOT BENADERINGEN VAN BELEID

In de vorige paragrafen is veel over beleidsevaluatie duidelijk geworden. Dat er een klassieke aanpak van beleid evalueren bestaat, dat er tal van maatstaven bestaan voor het beoordelen van beleid, dat ex post evaluatie het meest voorkomt maar dat ex ante-evaluatie ook heel nuttig is. Ook bleek dat het opzetten van een ex post evaluatie een design vraagt dat niet altijd te realiseren is (wet van de evaluatieve gebreken). Maar met een beetje handigheid is toch ook hier en daar een reparatie mogelijk, bij voorbeeld in de vorm van een goal free evaluation (om op zoek te gaan naar onbedoelde effecten) of een early warning evaluation (als de evaluatieopdracht te vroeg komt en je op zoek gaat naar consistentie in de in- en uitvoering van beleid terwijl je effecten nog niet in het zicht krijgt).

Het opzetten van een pragmatisch ingestoken plan van aanpak voor een *gebruiksgerichte* beleidsevaluatie blijkt desalniettemin mogelijk. Ook bleek dat er nadere keuzen te maken zijn in de omvang van de evaluatie. Daarvoor werd een ladder gepresenteerd. Sommige evaluaties zijn lichter, andere zwaarder.

Soms moet men ook de ladder niet bestijgen. Voor sommig beleid hoeft helemaal geen evaluatie plaats te vinden omdat er al voldoende kennis over succes en falen bekend is of beleid eenvoudig beëindigd wordt.

Ook is aangegeven dat andere evaluatietypen in de literatuur voorkomen, dan de genoemde, zoals inputevaluatie, procesevaluatie, doelmatigheidsevaluatie, responsieve evaluatie, netwerkgerichte evaluatie.

Wie ex post gebruikgericht gaat evalueren, de in de praktijk veel voorkomende manier van evalueren, heeft meer keuzemogelijkheden. De evaluator kan ook uitgaan van *benaderingen van beleid* die een cluster aan criteria met zich brengen.

Het verbinden van benaderingen van beleid met beoordelingsmaatstaven in een beleidsevaluatie is een thema dat in de bestuurskunde vaker aan de orde is gesteld (Abma en In 't Veld, 2001; Bekkers en Ringeling, 2003; Bekkers, 2007/2012). Het is dus iets om niet te negeren.

Een bekende indeling in beleidsoptieken met gevolgen voor de beleidsevaluatie betreft de volgende:

- de rationele benadering van beleid en beleidsevaluatie,
- de politieke benadering,
- de culturele benadering, en
- de netwerkbenadering.

Deze indeling wordt toegelicht. Het is een indeling voor gevorderden op beleidsevaluatief vlak en fijnproevers. In de hectiek van alledag zullen beleidsvoerders van een centrale of decentrale overheid niet vaak hierbij stil staan. Daarom behandel ik die pas op het laatst in deze beschouwing.

Deze indeling is niet zonder betekenis. Want als een bepaalde overheid afhankelijk is van andere overheden en private organisaties en zich dus in een netwerkcontext bevindt en niet over alle doorzettingsmacht beschikt, moet een evaluator zich realiseren dat hij of zij niet kan volstaan met een rationele beleidsbenadering waarin de overheid als een monocentrische eenheid wordt gezien (alsof één overheid het alleen voor het zeggen heeft) maar dat ook het

netwerkperspectief in het oog gehouden moet worden (Klijn en Koppenjan, 2001; Klijn, 2005). Voor de opzet van een evaluatie en de vraag aan wie te rapporteren maakt dat uit.

Geeft een provinciebestuur een opdracht tot een evaluatie van beleid (dat zich voltrekt in een netwerkcontext) en de rol van andere actoren worden daarin niet nadrukkelijk betrokken dan zal deze evaluatie tamelijk veel aan relevantie missen; immers de provincie is slechts een actor en die heeft het ook nog niet voor het zeggen. Ook hier is weer van een dilemma sprake want het komt zelden voor dat een heel netwerk van betrokkenen opdracht geeft tot een evaluatie. Een netwerk heeft immers over het algemeen geen formele status in het openbaar bestuur.

In het navolgende volgt een toelichting op de benaderingen.

Figuur 13: Vier benaderingen van beleid en de plaats van beleidsevaluatie daarin

	Rationele benadering	Politieke benadering	Culturele benadering	Netwerkbenadering
Focus actoren	Vastgestelde doelstellingen	Belangen, positiebehoud	Oriëntaties en gewoonten	Afhankelijke partijen
Beleid	Rationeel plan: cognitie (uitdenken) en vaststellen	Beleid is uitkomst politieke strijd: meer interactief dan cognitief	Beleid is symbolische expressie en kent gemeenschappelijke taal	Beleid is creatieve verbinding tussen van elkaar afhankelijke partijen in netwerk
Beleidsinhoud	Onderbouwd doel-middelen schema	Beleid is niet alleen onderbouwd maar ook compromis	Beleid als gedeeld verhaal	Interactief beleid, gemeenschappelijke beeldvorming
Sturen	Monocentrische overheid	Belangenafweging, macht verwerven	Pluriformiteit. Communicatie als collectief leerproces	Netwerkconstitutie en netwerk-management: poging tot doelverrijking en middelenballing
Vertrekpunt Ex-post evaluatie	Beleidstheorie/ veronderstellingen achter beleid/ argumentatie	Verplaatsing in positie, doelstellingen en belangen van stakeholders	Communicatie en draagvlak door gedeelde beeldvorming	Eerst netwerk in kaart brengen. Daarna analyse van kijk op beleid en waardering door partijen
Sterkte	Democratisch vastgesteld en onderbouwd beleid is uitgangspunt	Aandacht voor koers en bereikte consensus	Burgers moeten beleid erkennen; intersubjectiviteit	Geschikt als een overheid afhankelijk is (netwerk). Oog voor zwak georganiseerde belangen
Zwakte	Verwaarlozingen belangen en gewoonten	Neiging minder naar consistentie in doel-middelenschema te kijken	Neiging doel-middelenschema + belangen en macht te verwaarlozen	Te sterke procesgerichtheid. Resultaat moet ook beoordeeld worden.
Leren	Corrigeren van doel-middelenschema	Anticiperen op invloedspogingen van machtigen	Beleidslerend debat	Leren in netwerken: fixaties doorbreken + dialoog van doven voorkomen; actoren verplaatsen zich in elkaar; nieuwe beeldvorming

1) In de *rationele benadering* van beleid, eerder in deze beschouwing ook wel de klassieke benadering van beleid (evaluatie) genoemd, gaat de aandacht uit naar de officiële doelstellingen van het beleid, die richting geven aan de inzet van de middelen.

Bij middelen valt te denken aan de zweep, de peen en de preek, volgens en vrijmoedige duiding door oud-minister Pieter Winsemius. Logisch eigenlijk om doelen uit te werken met middelen. Bij bepaalde doelstellingen passen wellicht – volgens een andere indeling – juridisch instrumenten (wetgeving met verboden of geboden) en niet de keuze van economische instrumenten (belasting of subsidie), communicatieve instrumenten (convenant, voorlichting,

propaganda, interactieve beeldsvorming) of voorzieningen. Maar misschien past ook juist wel een *beleidsmix* van instrumenten of wordt een andere keuze gemaakt.

Het succes van de inzet van middelen wordt in de rationele benadering van beleid en beleidsevaluatie bepaald door drie hoofdcriteria: effectiviteit, doelmatigheid en samenhang (Bekkers, 2007). Bij beleidsevaluatie vanuit de rationele benadering zoals bepleit en verdedigd door onder andere Andries Hoogerwerf en de Algemene Rekenkamer zijn deze criteria van belang.

Bij effectiviteit van beleid gaat het om de vraag of de bereikte effecten overeenkomen met de gewenste effecten (lees: beoogde doeleinden). Als je daar als evaluator een beeld van hebt, is vervolgens de vraag of de middelen ook hebben bijgedragen aan het bereiken van de doeleinden in de vorm van prestaties ('output') en waargenomen effecten ('outcomes').

Om een voorbeeld te geven uit het verkeersveiligheidsbeleid: draagt de keuze van een voorziening (het middel de aanleg van verkeersdrempels) ook bij aan het doel van zeg 10 procent minder verkeersslachtoffers in een bepaalde periode of doet zoneringsmaatregelen tot de rijsnelheid van 30 km. dat ook of meer?

Eenvoudig is werken vanuit deze rationele beleidsbenadering niet. Achter deze aanpak gaat namelijk de assumptie schuilt dat alle doelstellingen altijd op papier staan en helder (beter SMART) zijn geformuleerd in meetbare grootheden. Is dat ook zo? Niet altijd want in een politieke democratie met coalities van minderheden kan beleid een zaak van compromissen zijn. In een volksvertegenwoordiging aangenomen moties en amendementen maken beleid niet altijd doelgericht(er) of meer SMART maar leveren soms 'een lappendeken-effect' op. Soms zal het behoud van vaagheid en het vermijden van meetbaarheid juist tot overeenstemming leiden maar ten koste gaan van het gemak bij evaluatie.

In de verderop te behandelen politieke benadering wordt dit compromiszoekend gedrag politiek rationeel geacht. Vanuit de rationele benadering is het verstoren van een consistent bouwwerk van gefundeerde veronderstellingen in beleid echter een gruwel.

De Algemene Rekenkamer (AR) heeft geen boodschap aan de 'goal free evaluation' of aan een politieke benadering. Die kamer stelt: dit is democratisch vastgesteld beleid en we mogen ervan uitgaan dat een meting van effectiviteit binnen een politieke democratie normaal en respectabel geacht mag en zelfs moet worden. De AR houdt de spiegel voor en is niet vies van verdere rationalisering in de vorm van het alsnog formuleren van meetbare doeleinden, een betere onderbouwing van de beleidstheorie en/of een beter aansluitende keuze van beleidsinstrumenten. De Tweede Kamer erkent deze opstelling van de Algemene Rekenkamer want de Tweede Kamer behandelt de rapporten en wijst de aanpak zelden af.

Een tweede veronderstelling achter de rationele benadering van beleidsevaluatie is dat sprake is van een grondige doordenking van het realiseren van doelstellingen met bepaalde middelen; dat proces zou een weloverwogen en doordachte keuze moeten opleveren. Wordt na evaluatie vastgesteld dat bepaalde veronderstellingen niet uitkomen dan zal de ambtelijke en bestuurlijke neiging bestaan om binnen het bestaande beleid een correctie te onderbouwen, en dus te corrigeren.

Dit is natuurlijk lang niet altijd verstandig, bijvoorbeeld als gevolg van padafhankelijkheid. Wie eenmaal de weg met veel juridische instrumenten belegd heeft, zal niet makkelijk het roer omgooien door beleid veel lichter vorm te geven in termen van communicatieve instrumenten als convenanten. Een jurist als Willem

Konijnenbelt noemde convenantisme eens 'fluiten in het donker'. Dan weet je het wel; hij zag niet zoveel in convenanten. De neiging kan dus bestaan om het bestaande beleid te corrigeren zonder een grote succeskans.

Een derde veronderstelling bij een effectiviteitsoordeel is dat er een goed onderscheid kan en moet worden gemaakt tussen prestaties van beleid ('output') en effecten van beleid ('outcomes'). Dat is echter niet altijd makkelijk. Soms is het gewenste effect zo verheven dat je al als evaluator blij bent dat je de output kunt zien. Zicht op de outcome is dan nog ver weg.

Voorbeeld: Het uitschrijven van verkeersboetes is waarneembaar maar is ook het doel van beter rijgedrag of grotere verkeersveiligheid bereikt?

In de rationele benadering wordt overigens ook wel gekeken naar doelmatigheid van beleid en samenhang van beleid. Die punten licht ik hier niet toe (zie Bekkers, 2007: 302).

2) Nu duidelijk is wat de rationele benadering inhoudt, kunnen de andere benaderingen korter worden behandeld. De tweede benadering van beleid en beleidsevaluatie (en daarmee om te komen tot de keuze van bepaalde beoordelingsmaatstaven) is *de politieke benadering* (Bekkers, 2007; De Kool, 2007).

Deze benadering richt zich *niet* uitsluitend op een doordacht doel-middelenschema, dus op een uitgewerkte beleidstheorie met veronderstellingen over wat beleid 'doet', maar zeker ook op de belangen en voorkeuren die vertegenwoordigers in de politieke en maatschappelijke arena's hebben. Wie wil wat waarom? Wie heeft welke machtspositie in de arena's en kan een opvatting over beleid doordrukken of de bedachte beleidsuitvoering tegenhouden of remmen? Wil een doelgroep van beleid, bij voorbeeld ondernemingen, weinig regelgeving maar wel subsidies en heeft ze de macht om dit te bereiken?

Wie vanuit een politieke benadering kijkt naar beleidsdoelen en middelen zal oog hebben voor symbolisch beleid dat niet of zwak geïnstrumentaliseerd is. Er zal gezocht worden naar de achtergrond hiervan. Willen politici een gebaar maken of ergens lippendienst aan bewijzen maar hebben ze er weinig budget en instrumentalisatie voor over? (Bekkers en Ringeling, 2003: 143 e.v.)

Vanuit de politieke benadering gaat het er bij een beleidsevaluatie niet alleen om de beleidstheorie achter een beleidsprogramma op te sporen maar ook om beoordelingscriteria als positiebehoud en draagvlak voor beleid mee te nemen. De evaluator die hier oog voor heeft moet beginnen met de tot stand koming van beleid te analyseren. Toen de evaluatiecommissie-Oosting waarvan ik deel uitmaakte de Wabm (Wet algemene bepalingen milieuhygiëne) evalueerde begon ze dan ook met een *totstandkomingsanalyse* van de wet.

Wie zo kijkt, komt er (ook) achter dat de kijk op een bepaald beleidsprogramma per actor kan verschillen. Dat is niet merkwaardig want wie politiek kijkt, weet dat beleid niet alleen iets is van 'uitdenken' maar vaak ook een kwestie is van belangen verzoenen en soms van wensen uitruilen en zaken 'uitvechten', dus van cognitie, interactie en positiebehoud (Hoppe, 2012). Politiek is als het ware te zien als een botsing van verschillende 'frames' – een verschillende kijk op een vraagstuk en oplossingsrichting - van politieke partijen en maatschappelijke actoren met machtsaanspraken als werkgevers- en werknemersorganisaties.

Succesvol beleid betekent in deze benadering niet alleen dat sprake is van effectief beleid maar ook van politiek en maatschappelijk geaccepteerd beleid.

Wat van deze politieke benadering te vinden? Verdedigers van de eerder genoemde rationale benadering hebben wel opgemerkt dat de adepten van een politieke benadering bepaalde slechte beleidspraktijken (van slecht onderbouwd beleid) rechtvaardigen. Zij bepleiten dat evaluatoren onafhankelijk zijn en afstand houden tot de politici. De rationalisten menen overigens dat ze zelf meer dan voldoende gebruiksggericht zijn want ze benadrukken 'beleid als argumentatie' (Pröpper, 1993, 2001; Pröpper en De Vries, 1995).

3) Er is een derde manier van kijken naar beleid en beleidsevaluaties: *de culturele benadering*. Ook hier gaat het weer om overeenstemming tussen politici en belanghebbenden.

Succesvol beleid bestaat volgens de culturele benadering uit het scheppen van een gemeenschappelijk beeld over nut en noodzaak van een bepaald beleid door de keuze van een bepaald referentiekader en een bepaalde taal (Bekkers, 2007: 303). Vragen zijn in dit verband onder andere:

- is er communicatie die gericht is op het faciliteren van een collectief leerproces?;
- wordt voor een nieuwe koers draagvlak verworven?

In deze optiek wordt - in een beleidsevaluatie volgens de politieke benadering - niet alleen gelet op de doelstellingen en de verbinding met instrumenten maar meer op de gemeenschappelijke richting die 'partners' in vertrouwen naar en met elkaar inslaan.

Deze benadering is nuttig (zie het werk van Hendriks) en maar staat in de schaduw van de eerder genoemde.

4) Stel dat iemand het overheidsbeleid ten aanzien van de verdroging of vernatting van De Peel wil bestuderen op resultaten en effecten. Heeft het dan zin om de rationele, politieke of culturele benadering te kiezen? Niet echt. Men komt eerder uit bij een netwerkbenadering (Bekkers en Lips, 1997, 1998).

De vierde benadering is daarmee gegeven: *de netwerkbenadering*. Deze benadering is aan de orde als een overheid zelf niet alle doorzettingsmacht heeft maar qua taken, bevoegdheden en middelen afhankelijk is van andere actoren. Dan moeten partijen die elkaar nodig hebben, elkaar ontmoeten en proberen te komen tot goede verhoudingen en uiteindelijk tot een gezamenlijk beeld van probleem- en oplossingsrichting. Dat is bij De Peel het geval omdat hier onder meer diverse departementen bij betrokken waren, alsmede enkele provincies, meerdere gemeenten, minstens een waterschap, tal van grondeigenaren en pachters.

Gemeenschappelijke beeldvorming is in dergelijke gevallen een belangrijk doel. Een dialoog van doven moet worden vermeden (Van Eeten en Termeer, 1996a, b; Van Eeten, 1999; Termeer, 2006; Jansen & Wald, 2007a, b).

Om tot die gemeenschappelijke beeldvorming te komen moet een procesmanager of een van de partners die deze rol op zich neemt ertoe bijdragen dat fixatie op eigen belangen en probleemdefinities in het netwerk overstegen worden in een nieuwe taal. Fixaties belemmeren immers de gemeenschappelijke beeldvorming.

Beleid is zo gezien niet alleen een zaak van cognitie (uitdenken) maar ook een sociaal en interactief proces. In de netwerkoptiek wordt een machtsstrijd voorkomen of de kop ingedrukt maar onderhandelingen en coalitievorming kunnen aan de orde zijn (Abma en In 't Veld, 2001: 29).

Veel analyse vanuit deze optiek is gericht op beschrijving ('hoe komt gemeenschappelijke actie tot stand?) en verklaring van beleidsprocessen. Evaluaties zijn ook mogelijk maar komen niet erg vaak voor.

Als ze aan de orde zijn, zijn ze gericht op onder meer de volgende vragen:

- wordt de afhankelijkheid van elkaar door partijen gezien?;
- wordt erkend dat elk belang heeft bij samenwerking?;
- zijn er convergerende krachten?;
- worden de verschillende partijen aan tafel gebracht en productief gemaakt? ;
- wat zijn de netwerkdoelstellingen?

Beleid wordt beoordeeld op basis van het bereiken van win-winsituaties en procescriteria als openheid, zorgvuldigheid en betrouwbaarheid (Abma en In 't Veld, 2001: 30; Klijn en Koppenjan, 2001; Klijn, 2005).

Uit de weergave van de netwerkbenadering blijkt dat sprake is van een veel meer uitgewerkt beeld van de relatie tussen overheid en samenleving dan in de rationele benadering. Het te vermijden nadeel van evaluaties gebaseerd op de netwerkbenadering houdt in, zoals de Utrechtse hoogleraar milieukunde Piet Glasbergen wel eens verklaard heeft, dat niet alleen naar de procedurele kant gekeken moet worden maar ook naar de kwaliteit van het (netwerk)resultaat.

Een tweede kritiek houdt in: niet alle beleid voltrekt zich in een netwerk. Er zijn nog steeds velden waar een gemeentelijke overheid volledig autonoom en ongestoord door andere overheden kan opereren. Dit kunnen de echte netwerkdenkers zich bijna niet voorstellen.

Nu deze vier benaderingen zijn genoemd, doemt de vraag op of een van deze vier benaderingen de favoriet is van de auteur van dit stuk. Het antwoord is nee. Eigenlijk hebben alle vier de benaderingen betekenis. Ik geef daarvoor de volgende argumenten.

a) Het leerstuk van de rationalistische benadering moet bij elke ambtenaar, dus ook bij een evaluator, bekend zijn. Zoals de Brabantse CdK en bestuurskundige Wim van de Donk begin 2013 nog tegen me zei: *'Ambtenaren horen te weten wat een doel-middel-schema is, wat een doelboom is, en dat je van beleid de achterliggende veronderstellingen kunt proberen te beoordelen. Dat moet tot het basisarsenaal van de provinciale ambtenaar behoren, vind ik'*.

b) Iedereen met politiek-bestuurlijk gevoel moet weten dat je een beleid met doelstellingen en daarmee samenhangende middelen kunt proberen te ontwerpen en onderbouwen met veel informatie. Maar beleid is ook de kunst van het haalbare, een kwestie van macht en steun in een volksvertegenwoordigend orgaan en beleid heeft vaak een relatie met min of meer machtige 'partijen' in de samenleving. De politieke benadering van beleid is dus niet te verwaarlozen. Het is dus niet alleen de kwaliteit van een ontwerp en de kwaliteitsrijke informatiebasis daarvoor die beleid tot goed beleid kan maken.

c) Een overheid heeft voorts lang niet altijd alleen doorzettingsmacht maar opereert in een netwerk van van elkaar afhankelijke overheden, semi-overheden en private organisaties. Dan is het nodig de beginselen van netwerksturing toe te passen en ook bij een beleidsevaluatie daarmee rekening te houden. Dat betekent dat bij de evaluatie andere maatstaven op de proppen komen dan effectiviteit of doelmatigheid van beleid.

De rationalisten hebben overigens geprobeerd om de netwerkbenadering en politieke benadering te verbinden met het denken in termen van beleidstheorie, beleidsdoelstellingen en middelen. Dat vind je terug in het werk van een evaluatiedeskundige als Hans Bressers.

22 SAMENVATTING

Evaluatie is in de bestuurskunde gemeengoed

Er is de afgelopen decennia veel over evalueren geschreven. Beleidsevaluatie is een onderdeel van de beleidscyclus, met als elementen het probleem, het beleidsontwerp, het besluit, de implementatie en evaluatie en tenslotte de beleidsdynamiek. Dat levert een klassiek leerstuk op.

Ook in het dagelijks leven is evaluatie volop aan de orde. Elke burger kijkt wel eens terug en oordeelt over hoe iets gegaan is en of hij of zij iets goed of slecht gedaan heeft om het uiteindelijk beter te (willen) doen.

Beleidsevaluatie is sinds ongeveer 1970 in Nederland geleidelijk aan een opmars bezig geweest. Bij *beleidsevaluatie* kan het gaan om het beoordelen van beleid van een organisatie, van het beleid van een onderwijsorganisatie tot het beleid van een omroepbestuur, van het beleid van de politiekorpsleiding tot dat van een gemeentebestuur. In deze beschouwing ging het om beleidsevaluatie bij *de overheid*.

Beleidsevaluatie heeft doorgaans betrekking op beoordeling van een beleidsprogramma maar hier wordt ook de beoordeling van hele organisaties meegenomen, dus in feite van alle beleid, zoals dat geschiedt in bestuurskrachtmetingen en visitaties van een provincie, een omroep, een politiekorps of een corporatie.

Beleidsevaluatie eenheidsworst?

Beleidsevaluatie, is dat eenheidsworst? Nee, er is meer gelijkenis met sneeuw. De Eskimo's kennen veel sneeuw en kunnen ook vele soort sneeuw van elkaar onderscheiden. Zo is het ook met beleidsevaluatie. Er is niet slechts één soort evaluatieonderzoek. Er zijn typen evaluatie, er zijn keuzen te maken. Maar er is wel een harde kern, zoals ook sneeuw iets essentieels kent om van sneeuw te kunnen spreken: dat essentiële, de *kern van evaluatie*, is beoordelen in het licht van de mogelijkheden tot verbetering.

Die variëteit aan evaluatievormen, concepten en aanpakken zal belangrijk blijken. Als een lezer thuis is in deze wereld kan hij die evaluaties onderscheiden zoals Eskimo's sneeuw kunnen onderscheiden en kan hij het gewenste type toepassen.

Kortom, er is een kern en er bestaan typen.

Beleidsevaluatie impliceert een onderzoek van de vraag *wat van beleidsprocessen terecht kan komen, komt of gekomen is uitgaande van een beoordeling op basis van bepaalde maatstaven*. Dat betekent dat een evaluatie van een beleidsontwerp mogelijk is, maar ook van genomen besluiten of van de organisatie van het beleid of van de beleidsuitvoering en – effecten.

Hier was het vertrekpunt *de evaluatie van bestaand beleid* aan de hand van maatstaven of criteria.

Die maatstaven, dat zijn vaak de *officiële* – dus vastgelegde en democratisch vastgestelde – overheidsdoelstellingen achter het beleid zelf. Beleid is immers een gerichte poging om met behulp van beleidsmiddelen, zoals een wet of voorlichtingscampagne, iets te bereiken in een bepaalde tijdspanne. Achter die gerichtheid gaan doelstellingen schuil en vaak ook verwachtingen, bij voorbeeld dat die doelstellingen ook gerealiseerd kunnen worden. doelstellingen zijn doorgaans te beschouwen als *gewenste effecten*.

Beleidsevaluatie in enge zin

Beleidsevaluatie komt in enge en in brede zin voor.

Beleidsevaluatie in enge zin is vaak slechts een *doelbereikingsanalyse*. In bredere zin speelt ook de beantwoording van de vraag mee of doelstellingen bereikt worden door middel van *de uitvoering* van het beleid zelf en/of door omstandigheden en processen *buiten* het beleid om. Dat betekent dat de beleidsevaluator zal kijken naar de doelstellingen achter het beleid en het realiseren van deze doelstellingen, te zien als beoogde effecten of hoe de uitvoering van het beleid verliep en bijdroeg aan deze realisatie. Deze *totaalanalyse* wordt *effectiviteitonderzoek* genoemd. Een beleidsevaluatie in brede zin, dus een effectiviteitsanalyse, omvat derhalve zowel een doelbereikingsanalyse als een oorzakelijkheidsanalyse.

Dit effectiviteitonderzoek is in feite de *klassieke* beleidsevaluatie, zoals die in de jaren zeventig en daarna werd uitgevoerd. Vingeroefeningen op dit vlak zijn gepleegd door onderzoekers als Hoogerwerf, Bressers en Maarse.

De effectiviteitsanalyse vereist nogal wat aan inzet en data en daarom is dat vooral een academische aanpak geworden. In de werkelijkheid komen tal van andere typen evaluatie voor. Zoals Eskimo's ook vele soorten sneeuw kennen.

Ideale onderzoeksopzet

Achter die *andere* typen beleidsevaluatie komen we door eerst te kijken naar *de ideale opzet* van een evaluatieonderzoek.

Idealiter moet er sprake zijn van een experiment. Dat behelst dat er twee groepen zijn en vier metingen volgens de 'before-after'- formule.

- Een *experimentele groep* die de beleidsinterventie ontvangt, wordt gemeten voor en na de interventie om te zien of de interventie iets doet en zo ja, wat de interventie doet.
- Daarnaast wordt een *controlegroep* gekozen die niet aan een interventie wordt blootgesteld. Ook hier een meting op twee momenten.
- De controlegroep en de experimentele groepen zouden gelijk moeten zijn in relevante kenmerken om de invloed van factoren uit te sluiten die een juist inzicht in de werking van de interventie zouden kunnen beïnvloeden (lees verstoren).

Een evaluatie op basis van een design met een experimentele groep en controlegroep en vier metingen (ook wel eens de Donald Campbell-aanpak genoemd) is natuurlijk te verrichten in de medisch-farmaceutische sfeer.

Een voorbeeld. Een groep werkzame mensen van zeg 20-40 jaar heeft wel eens ernstige hoofdpijn. Stel dat een farmaceutisch bedrijf een hoofdpijnstiller op de markt wil brengen. Dan zal dat bedrijf een certificaat willen van de goede werking van het medicijn en het uitblijven van substantiële ongewenste bijwerking. Een universitaire vakgroep, TNO of andere groep kan dan een onderzoek uitvoeren naar de kwaliteit van het medicijn. Besloten wordt om een experimentele groep te vormen en een controlegroep; die wordt samengesteld uit steekproeven uit de bevolking. Beide groepen zijn zoveel mogelijk gelijk of worden gelijk gemaakt. Dat betekent dat factoren die op vermindering of vermeerdering van hoofdpijn van invloed zouden kunnen zijn zoveel mogelijk gelijk verdeeld zouden moeten zijn verdeeld ('at random'). Denk aan gelijkenschakeling van variabelen als het soort werk of de keuze van het vervoermiddel naar en van het werk en aan het doen aan sport en fitness. Groepen moeten zoveel mogelijk gelijk zijn. Dan zou de ene groep gedurende een bepaalde periode het nieuwe middel (medicijn) krijgen, en de andere groep niets. Vervolgens vindt voor- en nameting plaats om het effect te zien van de interventie, het hoofdpijnremmend

middel dat wordt genomen. De opzet is nog uit te breiden, door meerdere experimentele groepen te maken. bijvoorbeeld de ene groep krijgt een pil met dosering 1 en de tweede groep een pil met een andere dosering 2. En men kan twee controlegroepen ontwerpen. De ene groep krijgt helemaal geen pil en de andere controle groep krijgt een bestaand middel of een 'fake' middel dat voor de hoofdpijn van geen betekenis is (maar burgers weten dat niet).

Enfin, duidelijk is dat om een goede opzet te krijgen van het onderzoek, die ook werkelijk het effect van de interventie naar voren brengt, tal van beslissingen genomen moeten worden. Deze aanpak *van before-aftermeting met experimentele groep en controlegroep* is arbeidsintensief en duur. De farmaceutische industrie heeft er veel belang bij dus die wil graag bijdragen aan de financiering hiervan. Daarmee de vraag oproepend: kan de onderzoeker wel onafhankelijk blijven?

Ook onderwijskundig en onderwijssociologisch onderzoek is wel verricht met een dergelijke before-afteeraanpak met controlegroepen. Denk maar aan de opzet om het effect van een klassenverkleining in het basis- of voortgezet onderwijs te meten, een heet hangijzer in de jaren zeventig van de vorige eeuw. De gedachte was dat klassenverkleining betere mogelijkheden tot goed onderwijs zouden bieden omdat de leerkracht te maken had met een beter overzienbaar en begeleidbaar aantal leerlingen. Ook sociaal-psychologen werken vaak met een experimentele onderzoeksopzet.

Evaluatieve gebreken

Maar bij beleidsinterventies van de overheid is het lang niet altijd mogelijk om een dergelijk gewenst design te kiezen. Om meerdere redenen. Ik noem er enkele.

1. *Geen controlegroep te kiezen.* Het kan zijn dat er geen burgers of anderen zijn uit te sluiten van een beleidsinterventie, bij voorbeeld omdat het door het parlement onrechtvaardig wordt gevonden om iemand iets aan beleid te ontzeggen, ook al is dat misschien maar tijdelijk. Dan kun je dus geen controlegroep vinden. Stel dat je een evaluatie van een belastingmaatregel onderneemt, dan geldt die voor iedereen die onder het betreffende belastingregime valt.
2. *Geen meting vooraf mogelijk.* Ten tweede is het mogelijk dat op het moment dat de evaluatieopdracht wordt geformuleerd de interventie al loopt. Dan is geen before-meting, dat wil zeggen geen zogenaamde nulmeting meer mogelijk.
3. *Geen vrije keuze in de after-meting.* Ten derde is het mogelijk dat op het moment van een opdracht de interventie al is gepleegd en al weer voorbij is. Dat is het geval met crisis. Stel een vliegtuig stort neer, wat niet direct een gewilde interventie is te noemen, maar waarvan een evaluatie wel nodig wordt geacht ('hoe kon dit gebeuren'), dan sta je behoorlijk met lege handen. Dergelijke evaluaties komen volop voor. Denk maar aan de Herculesramp in Eindhoven, jaren terug.
4. *Moment van after-meting nog niet aangebroken.* Een vierde mogelijkheid is dat nog geen prestaties en/of effecten zichtbaar zijn. Het is mogelijk dat op het moment van de evaluatieopdracht de interventie nog maar net gestart is. Het is eerst pas bekend dat beleid in werking is. Onder die omstandigheden is het eigenlijk nog te vroeg voor een evaluatie omdat de beleidsprestatie nog niet geleverd is, laat staan dat al beleidseffecten zichtbaar zijn. Geen evaluatie verrichten?

5. *De interventie is bijgesteld.* Ten vijfde is er een mogelijkheid dat op het tijdstip van een evaluatieopdracht het beleidsmiddel al een paar keer wijziging heeft ondergaan als gevolg van kritiek of onvoorziene omstandigheden. Denk aan een maatregel die op Europees vlak is genomen en in Nederland geïmplementeerd moet worden.
6. *Doelstellingen verschoven.* Ten zesde is het mogelijk dat ten tijde van de evaluatieopdracht het beleidsmiddel wel hetzelfde is gebleven maar de beleidsdoelstellingen na verloop van tijd zijn verschoven. Wellicht hebben debatten tussen een bestuur en een volksvertegenwoordiging ertoe geleid om doelstellingen terug te schroeven, dus om de ambitie bij te stellen. Of de doelstellingen verschoven door andere accenten in nieuw beleid of een nieuwe maatregel van de Europese Unie die nationaal implementatie behoeft. Wat neemt de evaluator dan als maatstaf voor de evaluatie: de officiële vastgelegde doelstellingen achter een beleid op moment 1 of de doelstellingen op moment 2 of moment 3? Deze dynamiek deed zich bij voorbeeld voor in het asielbeleid onder minister Leers ten aanzien van jeugdige ‘verwesterde’ asielzoekers. De mogelijkheid voor asiel aan hen werd versoepeld. Een evaluator moet daar rekening mee houden.
7. *Tenslotte is er nog de netwerkomstandigheid.* Voor tal van beleidsissues geldt dat er niet een overheid is die alle taken, bevoegdheden en middelen bezit om doelstellingen te realiseren en middelen zelfstandig in te zetten. Geen ‘partij’ heeft alle doorzettingsmacht in handen. In een dergelijke context van afhankelijkheden van actoren van elkaar kunnen fixaties optreden en een krachtige gerichtheid op het eigen belang. Dat moet doorbroken worden door gedeelde beeldvorming, en een doelverrijking en middelenballing. Kortom, door netwerkconstitutie (goede partijen aan tafel) en netwerkmanagement (het productief maken van de afhankelijkheden door verbindingen). Denk bij voorbeeld aan de verdroging en vernatting van De Peel. Als in deze context slechts een partij/ actor een opdracht geeft voor de beleidsevaluatie van beleid dat zich in een netwerkcontext voltrekt en zich moet richten op de netwerkdoelstellingen, ontstaat een risico op non-gebruik door andere partijen van de resultaten.

Vanwege het voorkomen van deze omstandigheden die het navolgen van een ideaal before-after-design met experimenteer- en controlegroep bemoeilijken, spreek ik van *de wet van de evaluatieve gebreken*. Komt die wet vaak voor?

Er is bij beleidsevaluatie vrijwel altijd sprake van een niet ideale omstandigheid voor het uitvoeren van een evaluatie. Het is vrijwel nooit mogelijk om het zuivere experiment met before-aftermeting bij een experimentele en controlegroep toe te passen. In onderzoek naar beleidseffecten is het zogenaamde *before-aftermodel zonder controlegroep* overigens wel wat vaker toegepast (Bressers, 2008: 169). Maar in gebruikgerichte contexten (dus bij op besluitvorming gerichte opdrachtresearch) toch niet vaak.

Deze wet van de evaluatieve gebreken heeft enkele consequenties in de vorm van een aanbeveling. Wat te doen?

Als er gebreken zijn dan goed nadenken over de opdracht

Ten eerste is *de opdrachtcontext* van groot belang voor opdrachtverlener(s) en onderzoeker(s). Zeg maar de entree-onderhandelingen tussen opdrachtgever en evaluerende opdrachtnemer. Te stellen vragen zijn: Om welk te evalueren beleid gaat het precies? Is er al tot dit beleid besloten? Wanneer, waarom? Is er sprake van implementatie? Zijn al prestaties of effecten in zicht? Zo nee, wanneer dan wel? Waarop zou de evaluatie zich moeten richten? Welke opzet

hiervoor te kiezen? Waartoe zou de evaluatie moeten leiden, dus welke functie kan die worden toegedicht?

Evaluatoren zouden moeten proberen vanuit de antwoorden op deze vragen creatief methoden voor beleidsevaluatie toe te passen (Swanborn, 2002; Bongers en Den Hertog, 2007).

Als er gebreken zijn welk evaluatietype vermijdt bezwaren

De tweede consequentie is dat er *evaluatietypen bestaan die tegemoet komen aan bepaalde bezwaren*. Ik heb daar zelf ervaring mee opgedaan en/of kom ze tegen in de literatuur.

Een voorbeeld is de *early warning evaluation* (EWA). Als de evaluatieopdracht tamelijk vlug na de beleidsvaststelling wordt verleend omdat er een bestuurlijke toezegging is gedaan die ertoe leidde dat een cruciale partij in een volksvertegenwoordiging toch over de streep is getrokken, dan kan een evaluator proberen een *consistentietoets van de schakels* te doen om te zien of het überhaupt wel mogelijk is in de buurt van gewenste effecten uit te komen.

Stel dat de beleidsprestatie nog niet geleverd is en de effecten nog niet zichtbaar zijn, wat dan te doen? Neem het voorbeeld van verkeersveiligheidsbeleid. Er zijn hot spots in wijken K en L waar veel verkeersincidenten optreden en ook jonge verkeersslachtoffers vallen. Stel het beleid behelst verkeersremmende maatregelen in de vorm van wegversmallingen en bobbels in het wegdek die snelheidsremmend werken in de woonwijken in de gehele gemeente. Stel dat dit beleid door de gemeenteraad is goedgekeurd. Stel ook dat er voorlichting over dit beleid plaatsvond en een tweede concept van het implementatieplan is opgesteld. De aanpak behelst dat eerst de wijken A tot en met F verkeerssluw gemaakt worden en dat dit 2 jaar duurt. Wat kan nu een early warning evaluation (EWA) doen? De evaluator kan de beleidsdoelstellingen bezien, vaststellen dat 'hot spot'-incidenten en slachtoffers aanleiding waren voor het nieuwe beleid maar dat dit langzamerhand uit het oog verloren is. Waardoor de echte urgente wijken K&L veel te laat aan de beurt komen om voorzieningen aan te brengen. Dan is de EWA een soort *consistentietoets* die kan vaststellen of het beleid uiteindelijk *naar verwachting* wel of niet uitkomt bij de gewenste effecten.

Als doelstellingen verschuiven dan dynamische analyse

Wie vindt dat het een probleem is om de oorspronkelijke doelstellingen achter beleid tot maatstaf te nemen voor een zoektocht naar gewenste effecten omdat de doelstellingen in de loop der tijd steeds een ander accent kregen, kan een *dynamische analyse* verrichten: kijk naar de eerste doelstellingen en naar de meer actuele. Opnieuw dus een oplossing voor een van de genoemde gebreken.

Goal free approach als neveneffecten in beeld moeten komen

Wie meent dat er neveneffecten zijn van beleid en die in de kijker wil krijgen, doet er goed aan om een 'goal free approach' te kiezen. Door te kijken naar de oppositie tegen het beleid, in het kader van de besluitvorming oevr beleid, komt de onderzoeker kritische punten op het spoor en daar kunnen neveneffecten bij horen, die verwacht worden. Ook zijn gesprekken mogelijk met belanghebbenden. Voor je het weet is overigens sprake van een combinatie van de goal free-aanpak en een responsieve evaluatietype.

Kiezen

Op die opdrachtcontext ga ik nog even door. De literatuur laat zien dat er allerlei typen evaluatie bestaan. Opdrachtgevers moeten dus *kiezen*. Opdrachtgevers zullen vaak voor een *utilization-focused evaluation* kiezen, een gebruiksgerichte evaluatie. Maar daar ben je er niet mee. Ik noem nog enkele mogelijkheden.

Kiezen: ex post of toch ante?

Een keuze is: *ex post evaluatie* of een *ex ante evaluatie*? *Ex post evaluatie* is evaluatie van een beleidsproces dat is vastgesteld en in uitvoering is genomen. *Ex ante evaluatie* is evaluatie in de ontwerpfase van beleid, voorafgaand aan de beleidsvaststelling.

Omdat *ex post evaluatie* vaak aan het licht brengt dat beleid niet 'doet' wat het moet doen (niet de beoogde effecten geeft, zie Leeuw en Van Gils, 2010), is de aandacht de laatste decennia verschoven naar het vaker uitvoeren van *ex ante evaluatie*. Je kunt al in de fase van het ontwerpen van beleid proberen om beleidsopties te doordenken op *kosten en baten* en op de vraag of men op de reis wel kan aankomen waar men wil uitstappen. Als de kans klein is dat men doelstellingen realiseert, is een rechtvaardiging mogelijk door te stellen dat de ambities nu eenmaal groter zijn dan wat je realiseert. Dat is een rechtvaardiging voor falen, op voorhand. Men kan echter ook zeggen dat uit een oogpunt van zuinigheid en doelmatigheid beleid dat op een fiasco dreigt uit te lopen beter niet vastgesteld moet worden. Dus is uitvoeringsgericht ontwerpen bepleit in de vorm van *forward & backward mapping* (Elmore).

Ex ante evaluatie sluit daar op aan: je doordenkt de doelstellingen en kwetsbare punten in het traject daarheen nog voor beleid is vastgesteld. Een effectrapportage is een doordenking van de vraag of en zo ja welk effect je wel of niet bereikt. Maar effectrapportages zijn meestal niet integraal maar slechts partieel (denk maar aan de milieu-effectrapportage). Zuivere *ex ante evaluatie* is juist wel integraal.

Als een overheidsbestuur wil kiezen voor *ex post evaluatie* zijn er in werkelijkheid nog tal van *andere* keuzen te maken, zoals in:

- Een *doelbereikingsevaluatie* of een *procesevaluatie*;
- Een *retrospectieve* en *prospectieve evaluatie*;
- Een evaluatie volgens een model van *zelfevaluaties* onder auspiciën van een *visitatiecommissie*;
- Een *intern* of *extern* uit te voeren evaluatie.

Evaluatieladder

Ik heb een *evaluatieladder* opgesteld voor *ex post evaluatie* om een potentiële opdrachtgever van dienst te zijn bij het doordenken van een keuze uit evaluatievormen. De klassieke effectiviteitsanalyse met before-after design met een experimentele en controlegroep staat dan aan de top van de ladder, en direct daaronder een evaluatie met before-after meting *zonder* controlegroep.

De klassieke effectiviteitsevaluatie kan zicht geven op doelstellingen en gewenste effecten en ook de bijdrage van beleid aan die effecten in beeld brengen (oorzakelijkheidsanalyse genoemd). Een dergelijke opzet vereist veel onderzoeksexpertise aan de kant van evaluatoren, is uitgebreid en rijk van inhoud maar is duur (prijzig). Dat zijn enkele maatstaven voor de beoordeling van een evaluatievorm.

Procesevaluaties staan lager op de ladder. Ze worden ook minder gebruikt dan effectiviteitsevaluaties.

Visitatiecommissies

Hoe zit het met *visitatiecommissies* die werken met *zelfevaluaties*: nuttig? Zijn die procedures juist wel prospectiever of toch niet? *Visitatiecommissies* kwamen we tegen in het hoger onderwijs, het grotestedenbeleid, de mediawereld, bij politiekorpsen en in gemeenten.

In het hoger onderwijs voldeed de aanpak om te werken met commissies vele jaren, omdat die goed doordacht bleek. Ik heb er de *beginselen van behoorlijke visiteren* aan ontleend. Daartoe behoort: een visitatiecommissie moet onafhankelijk zijn; een visitatiecommissie moet een protocol beschikken waaraan een zelfevaluatie moet voldoen en welke gegevens die moet bevatten; een visitatiecommissie moet ter plaatse gaan kijken, de analyse aanvullen en de zelfevaluatie checken; een visitatiecommissie moet een mogelijkheid geven tot reactie op een concept-rapport over de visitatie door de gevisiteerden opdat de feiten kloppen en volledig zijn weergegeven.

Als aan dergelijke beginselen vastgehouden wordt, is de kans groot dat een te visiteren organisatie zich herkent in het portret dat de visitatiecommissie van haar schetst en de bevindingen ook erkent. De beginselen zijn dus krachtig bij aan een goede visitatie.

Ik heb die beginselen nog eens op geldigheid 'getoetst'. Ook visitaties in het grotestedenbeleid hebben heel veel weg gehad van visitaties in het hoger onderwijs. Dat geldt ook voor de visitatiecommissie die in twee rondes enkele tientallen gemeentebestuur in Limburg heeft beoordeeld op bestuurskracht. Beide visitatieprocessen zijn positief beoordeeld. Daarom is sprake van *evidence-based evaluation policy*. We weten langzamerhand hoe het evalueren doro middel van visitatie moet en wat geen verstandige aanpak is.

Meer toekomstgericht

Veel van de ex post evaluaties hebben echter ook een nadeel. Ze zijn vaak retrospectief en *niet* prospectief (genoeg); ze bieden geen verbeeldingskrachtige blik op de toekomst als handreiking voor bestuur.

Wie kijkt naar '*the politics of evaluation*', zeg maar de politieke context van evaluatie, komt er achter dat een politiek bestuur hier en daar juist een evaluatie wil die zowel retrospectief bevat als een meedenken over de opties die nu opstaan voor een bestuur om zich op de toekomst te richten.

Een visitatie van Vlaamse steden in opdracht van de Vlaamse overheid, die ik leidde als covoorzitter, was juist meer gebaseerd op zekere zelfevaluatie, en op meedenken van de betrokken ambtenaren en bestuurders, en dus retrospectief maar toch ook met een belangrijke component aan prospectie.

Lukte die prospectie? De gedachte was dat meedenken door aanreiking van 'best practices' een mogelijkheid kon zijn. En dat dus overleg tussen gemeenten, waarbij ook voorlopers betrokken waren, de achterblijvers zou kunnen meeslepen. Dat laatste is mogelijk maar het eenvoudig benoemen van 'best practices' is bepaald niet eenvoudig. Wat goed werkt in de ene context werkt nog niet altijd in een andere.

Lager op de ladder

Terugkerend naar de evaluatieladder stel ik voor visitatiecommissies met zelfevaluatie hoog op de ladder te plaatsen. Zijn er dan ook onderzoekstypen die a prima vista lager op de ladder te plaatsen zijn? Ik meen van wel en noem hier *prestatiemeting* (De Bruijn, 2004), *monitoring* (De Kool, 2007a) en *benchmarking* (Hakvoort en Klaassen, 2004). Monitoring wordt door velen niet eens als echte beleidsevaluatie werkend.

Prestatiemeting impliceert de vraag of beleid ook tot resultaat leidt. Het gaat om de output, vaak nog *niet* om de effecten of outcomes.

Prestatiemeting is de laatste jaren een onderdeel geworden van programmabegrotingen. In deze begrotingen zijn de drie W's aan de orde: wat willen

we, wat doen we daarvoor en wat mag het kosten? De vierde W: wat is het resultaat? Wie de begroting van de gemeente Amsterdam bekijkt ziet daarin tal van prestatiegegevens opgenomen. Waarom de prestatie wel of niet bereikt wordt, is daarin meestal niet aan de orde. Er is weinig tegen prestatiemeting maar deze aanpak kent zijn grenzen.

Monitoring komt tegenwoordig nog al eens voor. Denk aan een veiligheidsmonitor. Veelal reiken dergelijke monitors niet ver. Het is niet zelden veredelde statistiek over prestaties en prestatie-indicatoren. De keuze van indicatoren is nogal eens arbitrair of de belangrijkste indicatoren zijn afwezig. Ze geven geen diepgaande analyse van doelbereiking, zeggen weinig over beleidsuitvoering en de bijdrage daarvan aan het realiseren van effecten. De verklarende betekenis is nagenoeg afwezig. Daar staat tegenover dat bestuurders soms wel kijken naar cijfers uit monitors, zeker als de monitor het enige middel is om iets te zeggen over vooruitgang of achteruitgaan. Denk aan de bestuurders van de gemeente Rotterdam.

Benchmarking staat ook betrekkelijk laag op de ladder. Benchmarking impliceert een vergelijking van organisaties of onderdelen daarvan met andere organisaties met het doel om te verbeteren en zelfs best practices te detecteren en tot uitgangspunt te nemen voor verbetermanagement (Askim, 2008). Deze benchmarks hebben in de praktijk vaak volgende kenmerken:

- dat alleen wat kwantificeerbaar is, wordt gemeten;
- dat operationeel beleid belicht wordt (zoals de kwaliteit van rioleringen);
- dat vaak geen rekening gehouden wordt met beoordelingsmaatstaven in overheidsbeleid zoals duurzaamheid, rechtvaardigheid, solidariteit of level-playing fields;
- dat er slechts één rapport voor tal van opdrachtgevers verschijnt met weinig toelichting, wat negatief inwerkt op de gebruiksgerichtheid van deze rapportages;
- dat vaak geen zicht geboden wordt waarom de ene organisatie boven of onder het gemiddelde zit; en
- dat de gebruikers niet goed weten wat ze met het rapport aanmoeten.

Van de inhoud van benchmarking-rapporten worden wel kennis genomen door de meest betrokken ambtenaren maar ze bereiken vaak niet het dagelijks bestuur (of controlerend orgaan) omdat dit de resultaten niet goed kan interpreteren en rapportages niet van aanbevelingen zijn voorzien. Deze bevinding sluit aan bij het eerder vastgestelde matig tot geringe gebruik van begrotingskengetallen in politiek-bestuurlijke contexten.

Maar onder bijzondere omstandigheden kan een bepaald benchmarkrapport er ineens wel toe doen. In tijden van bezuinigingen kan het interessant zijn om uit een benchmark van ict-kosten te halen dat de ict-kosten per inwoner van gemeenten met meer dan 100.000 inwoners 17 procent hoger liggen dan voor gemeenten met een geringer inwonertal. Dan is het schaalvoordeel voor de grotere gemeente blijkbaar verdwenen. Hoe komt dat? Grotere gemeenten hebben meer eigen medewerkers op het vlak van informatiemanagement. En hoe groter de gemeente hoe groter de afstemmingsbureaucratie. Minder grote gemeenten huren tijdelijke krachten in (BB, 150213).

Een evaluatieladder is wel nuttig maar heeft als nadeel het beeld dat een staalkaart wordt gepresenteerd waaruit te kiezen is. Alsof het een boek betreft waaruit behang gekozen wordt. Een waarschuwing voor teveel staalkaartdenken is nodig.

De conclusie moet zijn dat bezinning op de opdracht heel belangrijk is voor de evaluatieopzet en het verdere vervolg. In deze fase zijn tal van keuzen te maken: geen evaluatie verrichten, later verrichten, een uitgebreide of smallere evaluatievorm kiezen, etc.

Wat laat de praktijk van evaluatie nog meer zien? Dat evalueren in een netwerkcontext mogelijk is.

Valt er te leren van evaluatie? Inderdaad. We gaven een aantal suggesties voor 'leren van evalueren'.

Literatuur

- Aardema, H. e.a. (red.), *Meerwaarde van de bestuurskunde*, Lemma/Boom, 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 2009b, pp. 35-39.
- Abma, K. en A.F.A. Korsten, *Gemeenten in rapportcijfers*, Eburon, Delft, 2009a.
- 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., Bestuurskracht en kwaliteit, in: *Openbaar bestuur*, jrg. 19, nr. 10, oktober 2009.
- Abma, K., Bestuurskrachtmeting inclusief visitatie, in: A. Eversdijk en J. Schutgens (red.), *Openbaar bestuur op expeditie - Verkenningen van bestuurlijke vernieuwingspogingen - Liber Amicorum voor prof. dr. Arno F.A. Korsten*, Boom Lemma, Den Haag, 2010, pp. 18-39.
- Abma, T. en R.J. in 't Veld (red.), *Handboek beleidswetenschap*, Boom, Amsterdam, 2001.
- Abma, T., *Responsief evalueren*, Rotterdam, 1996 (diss.).
- Abma, T., Variaties in beleidsevaluaties: meten, beoordelen en onderhandelen, in: Abma, T. en R.J. in 't Veld (red.), *Handboek beleidswetenschap*, Boom, Amsterdam, 2001, pp. 311-321.
- Abma, T.A., Beleidsevaluatie zonder beter weten, in: Gunsteren, H. van, en E. van Ruyven (red.), *Bestuur in de ongekende samenleving*, Sdu, Den Haag, 1995, pp. 163-180.
- Abma, T.A., Onderhandelend evalueren, in: *Bestuurskunde*, 2000, nr. 8, pp. 393-403.
- Abma, T.A., Variaties in beleidsevaluatie, in: *Bestuurskunde*, 1997, nr. 2, pp. 77-87.
- Algemene Rekenkamer, *Bestuur op afstand in beeld- Tien jaar onderzoek van de Algemene rekenkamer naar instellingen op afstand van het Rijk*, Sdu, Den Haag, 2008a.
- Algemene Rekenkamer, *Kwaliteitsbewaking in het hoger onderwijs in Nederland en Vlaanderen*, Verslag aan de Tweede Kamer der Staten-Generaal en het Vlaams Parlement, Sdu Uitgevers, Den Haag, 2008b.
- Algemene Rekenkamer, *Kwaliteitszorg Hoger Onderwijs*, Den Haag, 2000.
- Algemene Rekenkamer, *Onregelmatigheden bekostiging (hoger) onderwijs*, Den Haag, 2003.
- Algemene Rekenkamer, *Organisatie van de beleidsevaluatie*, Sdu, Den Haag, 2000.
- Amaral, A.M.S.C., The US Accreditation System and the CRE's Quality Audits - A Comparative Study, in: *Quality Assurance in Education*, vol. 6, 1998, nr. 4, pp. 184-196.
- Argyris, C. & D. Schön, *Organizational Learning: A Theory of Action Perspective*, Addison-Wesley, Reading, 1978.
- Argyris, C., *Leren in en door organisaties*, Scriptum, Schiedam, 1996.
- Argyris, C., *On Organizational Learning*, Blackwell Publishing, Malden, USA, 1999.
- Askim, J., Determinants of Performance Information Utilization in Political Decision Making, in: Dooren, W. van, en 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.
- B&A Groep, *Bestuurskrachtmeting provincie Zeeland*, Den Haag, 2011 (concept).
- Bagchus, R., *Onderwijs, Cultuur en Wetenschappen - Profiel van een ministerie*, Departementale studies nr. 6, VUGA, Den Haag, 1995.
- Bagchus, R., *Onderwijs, Cultuur en Wetenschappen*, in: Twist, M.J.W. van, R. Bagchus en F.O.M. Verhaak (red.), *Kerndepartementen op afstand?*, Eburon, Delft, 1996, pp. 147-167.
- Bardach, E., Policy Dynamics, in: Moran, M., M. Rein & R.E. Goodin (eds.), *The Oxford Handbook of Public Policy*, Oxford University Press, 2006, pp. 336-367.
- Baumgartner, F.R. & B.D. Jones (eds.), *Policy Dynamics*, Chicago UP, Chicago, 2002.
- Becker, H.A., Ex ante evaluatie: consequentie-analyses, in: *Beleid en Maatschappij*, 1983, nr. 11, pp. 314-325.
- Becker, H.A., Ex ante-evaluatie: consequentie-analyses, in: Lehning, P.B. en J.B.D. Simonis (red.), *Handboek beleidswetenschap*, Boom, Meppel, 1987, pp. 116-130.
- Becker, H.A., *Social impact assessment*, UCL Press, Londen, 1997.
- Beek, D. van, *Sturen naar een intelligente overhead*, Reed Business, Den Haag, 2007.
- 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.
- Bekker, R., *Marathonlopers rond het Binnenhof – Topambtenaren bij het rijk 1970-2010*, Boom/Lemma, 2012.
- Bekkers, V. en A. Ringeling (red.), *Vragen over beleid- Perspectieven op waardering*, Lemma, Utrecht, 2003.
- Bekkers, V.J.J.M. , M. Fenger en V. Homburg, *Doorwerking van strategische beleidsadvisering*, Erasmus Universiteit, Rotterdam, 2004.
- Bekkers, V.J.J.M. en M. Lips, Spierballen, veel stroop en weinig water – Verdrogingsbeleid in de provincie Noord-Brabant, in: Hendriks, F. en Th.A.J. Toonen (red.), *Schikken en plooiën*, Van Gorcum, Assen, 1998, pp. 153-168.
- Bekkers, V.J.J.M., *Beleid in beweging*, Boom/Lemma, Den Haag, 2007.
- Bekkers, V.J.J.M., M. Fenger en V. Homburg, Doorwerkingspatronen van strategische advisering, in: *Bestuurswetenschappen*, 2005, nr. 5, pp. 431-450.
- Bekkers, V.J.J.M., *Nieuwe vormen van sturing en informatisering*, Eburon, Delft, 1994.
- Bemelmans-Videc, M.L. en M. de Vries, Ontwikkelingen in de evaluatie van overheidshandelen, in: Aardema, A. e.a. (red.), *Meerwaarde van de bestuurskunde*, Boom/Lemma, Den Haag, 2010, pp. 127-137.
- Bemelmans-Videc, M.L., B. Perrin & J. Lonsdale (eds.), *Making Accountability Work: Dilemmas for Evaluation and for Audit*, Transaction Publ., Edison (NJ), 2007.
- Benneworth, P., H. de Boer, L. Cremonini, B. Jongbloed, L. Leisyte, H. Vossensteyn & E. de Weert, *Quality related funding, performance agreements and profiling in higher education: An international comparative study*, CHEPS Universiteit Twente, Enschede, 2011.
- Benz, A., S. Lütz, U. Schimank & G. Simonis (Hrsg.), *Handbuch Governance*, VS Verlag, Wiesbaden, 2007.
- Berkel, H. van, De visitatie-illusie, in: *Tijdschrift voor hoger Onderwijs en Management (THEMA)*, 2000, nr. 5, pp. 25-29.
- Berndsen, R., M.Fraanje, A.F.A. Korsten en M. Kort, Beoordeling van provinciale bestuurskracht, in: *Bestuurswetenschappen*, 2007, nr. 6, pp. 11-49.
- Beus, J. de, Te veel transparantie werkt niet, laat politici de ruimte om te besturen, in: *NRC*, 24 juni 2006.
- Bijl, R., Evaluatieonderzoek en het integratiebeleid, in: *Justitiële Verkenningen*, jrg. 31, 2005, nr. 8, pp. 85-100.
- Boekenoogen, M., P. van Hoesel, Y. Prince en C. Verheijen (red.), *Methoden van beleidsonderzoekers: creatief en oplossingsgericht*, Lemma, Den Haag, 2009.
- Bogt, H.J, ter, P. Jansen en S. Tillema (red.), *Met het oog op de publieke zaak*, RUG, Groningen, 2010.
- Bogt, H.J. ter, Bronnen van controlinformatie voor politieke bestuurders, in: *Maandblad voor Accountancy en Bedrijfseconomie*, jrg. 75, 2001a, nr. 12, pp. 525-538.
- Bogt, H.J. ter, De doelmatigheid van outputgerichte managementinformatie, in: *Beleidswetenschap*, jrg. 16, 2002, nr. 2, pp. 114-144.
- Bogt, H.J. ter, en G.J. van Helden, Kwaliteit van prestatiemeting bij gemeentelijke overheden, in: *Bestuurskunde*, jrg. 3, 1994, nr. 1, pp. 2-13.
- Bogt, H.J. ter, en J.G. van Helden, Gebruik van outputinformatie bij gemeentelijke overheden, in: *Bestuurskunde*, 1994, nr. 1, pp. 2-13.
- Bogt, H.J. ter, Gebruik van outputinformatie door gemeentebestuurders, in: *Overheidsmanagement*, 2000, nr. 7/8, pp. 183-187.
- Bogt, H.J. ter, *Neo-institutionele economie, management control en verzelfstandiging*, Labyrint, Cappelle a.d. IJssel, 1997.
- Bogt, H.J. ter, Performance evaluation styles in governmental organizations: how do professional managers facilitate politicians' work?, in: *Management Accounting Research*, vol. 14, 2003, nr. 4, pp. 311-332.
- Bogt, H.J. ter, Politicians and output-oriented performance evaluation in municipalities, in: *The European Accounting Review*, vol. 10, 2001b, pp. 621-643.
- Bogt, H.J. ter, Politieke bestuurders, productbegrotingen en prestatiebeoordelingen bij gemeenten, in: *Maandblad voor Accountancy en Bedrijfseconomie*, jrg. 73, 1999, nr. 12, pp. 671-683.
- Bogt, H.J., ter, *Managementvernieuwing bij de overheid - Mooie woorden of echte daden?*, Groningen, 2005 (oratie).
- Bogt, H.J., ter, Prestatiegegevens: nuttig voor sturing van gemeentelijke taken, in: *Public Controlling*, juni 2004, pp. 52-58.

- Bongers, F.J. en D. Verlet, Indicatoren en beleid: doe meer met ongeveer, in: *Bestuurswetenschappen*, jrg. 67, 2013, nr. 1, pp. 36-55.
- Bongers, F.J. en P. den Hertog, Vernieuwing van methoden voor beleidsevaluatie, in: *Bestuurswetenschappen*, 2007, nr. 3, pp. 29-44.
- Bonney, M., Evaluatieonderzoek in de praktijk: klimaatbeleid, in: Knaap, P. van der, en A. Schilder (red.), *Resultaatgericht sturen en evalueren*, Sdu, Den Haag, 2004, pp. 145-155.
- Boomkens, R., *Topkisch en slow science – Kritiek van de academische rede*, Uitgeverij Van Gennep, Amsterdam, 2008.
- Borst, R., C. Lako en M.S. de Vries, Is prestatie management in de publieke sector toepasbaar? – Een vergelijkend onderzoek naar de opvattingen onder ambtenaren, in: *Bestuurswetenschappen*, jrg. 66, 2012, nr. 4, pp. 14-33.
- 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. van Dooren, Performance Measurement and Management in Public Sector Organizations, in: Boivard, T. & E. Löffler (eds.), *Public Management and governance*, Routledge, Londen, 2004, pp. 127-137.
- Bouckaert, G., B. De Peuter en W. van Dooren, *Meten en vergelijken van lokale bestuurlijke ontwikkeling: een monitoringsysteem voor het lokaal bestuur in Vlaanderen*, Die Keure, Brugge, 2003.
- Bouckaert, G., Prestaties en prestatie management in de publieke sector, in: *Tijdschrift voor Economie en Management*, 2006, nr. 3, pp. 237-265.
- 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.
- Bovens, M. & P. 't Hart, *Understandig Policy Fiascoes*, Transaction Books, New Brunswick, 1996.
- Bovens, M. en P. 't Hart, Beleidsfiasco's tussen bestuurlijk falen en bestuurlijke pech, in: Abma, T. en R.J. in 't Veld (red.), *Handboek beleidswetenschap*, Boom, Amsterdam, 2001, pp. 127-145.
- Bovens, M. en Th. Schillemans (red.), *Handboek publieke verantwoording*, Boom Lemma, Den Haag, 2009.
- Bovens, M., & P. 't Hart, *Understanding policy fiascoes*, Transaction Publ., New Brunswick, 1996.
- Bovens, M., P. 't Hart & S. Kuipers, The Politics of Policy Evaluation, in: Moran, M., M. Rein & R.E. Goodin (eds.), *The Oxford Handbook of Public Policy*, Oxford University Press, 2006, pp. 319-335.
- Bowerman, M., A. Ball & G. Francis, Benchmarking as a Tool for Modernization of Local Government, in: *Financial Accountability and Management*, jrg. 17, 2001, nr. 4, pp. 321-329.
- Bressers, J. & A. Hoogerwerf (red.), *Beleidsvaluatie*, Samsom, Alphen, 1995 (3^e druk).
- Bressers, J.Th.A., *Beleidsdoelmatigheid en waterkwaliteitsbeleid*, THTWente, Enschede, 1983.
- Bressers, J.Th.A., Beleidsvaluatie en beleidseffecten, in: Hoogerwerf, A. (red.), *Overheidsbeleid*, Samsom, Alphen, 1993, pp. 161-179.
- Bressers, J.Th.A., Evaluatie van beleid, in: Hoogerwerf, A. en M. Herweijer (red.), *Overheidsbeleid*, Kluwer, Alphen, 2008 (8^e druk), pp. 155-179.
- Bressers, J.Th.A., P. de Jong, P.J. Klok en A.F.A. Korsten (red.), *Beleidsinstrumenten bestuurskundig beschouwd*, Van Gorcum, Assen, 1993.
- Bressers, J.Th.A., Triangulatie in evaluatie-onderzoek; beleidseffectiviteit en waterkwaliteitsbeleid, in: Derksen, W., A.F.A. Korsten en A.F.M. Bertrand (red.), *De praktijk van onderzoek*, Wolters-Noordhoff, Groningen, 1988, pp. 263-287.
- Brewer, G. & P. deLeon, *The Foundations of Policy Analysis*, The Dorsey Press, Homewood, 1983.
- Brink, G. van den, Th. Jansen en D. Pessers (red.), *Beroepszeker- Waarom Nederland niet goed werkt*, Boom, Amsterdam, 2005.
- Bruijn, H. de, & G.J. van Helden, A Pea for Dialogue Driven Performance-based Management Systems: Balance from the Dutch Public Sector, in: *Financial Accountability & Management*, vol. 22, 2006, nr. 4, pp. 405-423.
- Bruijn, H. de, en E. ten Heuvelhof, *Handhaving- Het spel tussen inspecteur en inspectee*, Lemma, Utrecht, 2005.
- Bruijn, H. de, en E. ten Heuvelhof, Procesmanagement, in: Abma, T. en R.J. in 't Veld (red.), *Handboek beleidswetenschap*, Boom, Amsterdam, 2001, pp. 193-208.
- Bruijn, H. de, en J. van Helden, Effectief prestatie management bij professionele organisaties in de publieke sector, in: *Overheidsmanagement*, febr. 2007, nr. 1, pp. 20-26.

- Bruijn, H. de, *Framing- Over de macht van taal in de politiek*, Atlas, Amsterdam, 2011.
- Bruijn, H. de, G.R. Teisman, J. Edelenbos en W. Veeneman (red.), *Meervoudig ruimtegebruik en het management van meerstemmige processen*, Lemma, Utrecht, 2004.
- Bruijn, H. de, *Managers & professionals*, Academic Service/Sdu, Den Haag, 2008.
- Bruijn, H. de, *Managing Performance in the Public Sector*, Routledge, Londen, 2007.
- Bruijn, H. de, Prestatiemeting in de publieke sector, in: Knaap, P. van der, en A. Schilder (red.), *Resultaatgericht sturen en evalueren*, Sdu, Den Haag, 2004, pp. 195-205.
- Bruijn, H. de, *Prestatiemeting in de publieke sector*, Lemma, Utrecht, 2001.
- Bruijn, J. de, E. ten Heuvelhof en R.J. in 't Veld, *Procesmanagement*, Academic Service, Schoonhoven, 2002.
- Bruinsma, F., De ondraaglijke lichtheid van de rechtenstudie, in: *Nederlands Juristenblad*, jrg. 28, 2000, pp. 1371-1374.
- Burg, W. van der Burg, De regulering van professionals: twee botsende perspectieven, in: Jansen, Th., G. van den Brink en J. Kole (red.), *Beroepstrots- Een ongekende kracht*, Boom, Amsterdam, 2009, pp. 168-182.
- Camps, Th.W.A., *Het besloten concilie der deskundigen*, Lemma, Utrecht, 1995 (diss.).
- Caniëls, M.C.J., *Lerende organisaties in de kenniseconomie*, Open Universiteit, Heerlen, 2011.
- Caplan, N., The Two-Communities Theory and Knowledge Utilization, in: *American Behavioral Scientist*, vol. 22, 1979, nr. 3.
- Carter, R., Maximizing the Use of Evaluation Results, in: Wholey, J.S., H.P. Hatry & K.E. Newcomer (eds.), *Handbook of Program Evaluation*, Jossey-Bass Publ., San Francisco, 1994, pp. 576-590.
- Casey, D., *Managing Learning in Organizations*, Open University Press, Buckingham, 1993.
- 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.
- Coleman, J.S., *Policy research in the social sciences*, General learning Press, Morristown, 1972.
- Commissie Bestuurlijke Evaluatie Algemene Inlichtingen- en Veiligheidsdienst (commissie-Havermans), *De AIVD in verandering*, Den Haag, november 2004.
- Commissie 'Project x' Haren (commissie-Cohen), *De weg naar Haren- Deelrapport 2*, Haren, 2013 (c).
- Commissie 'Project x' Haren (commissie-Cohen), *Er is geen feest - Deelrapport 1*, Haren, 2013 (b).
- Commissie 'Project x' Haren (commissie-Cohen), *Hoe Dionysos in Haren verscheen - Deelrapport 3*, Haren, 2013 (d).
- Commissie 'Project x' Haren (commissie-Cohen), *Twee werelden - Hoofdrapport*, Haren, 2013 (a)
- Commissie Accreditatie Hoger Onderwijs, *Prikkelen, presteren, profileren*, Eindrapport, Den Haag, 2001.
- Commissie Diplomering Langstudeerders Hogeschool INHolland (commissie-Leers), *Veel ruimte, weinig rekenschap*, Haarlem, september 2010.
- Commissie Toekomst Lokaal Bestuur (Cie-Bovens), *Wil tot Verschil - Gemeenten in 2015*, Vereniging van Nederlandse Gemeenten, Den Haag, 2006.
- Commissie voor Beleidsevaluatie, *Terugblikken met het oog op de toekomst - Een zelfevaluatie van de commissie voor Beleidsevaluatie over de bestuursperiode 2007-2011*, Den Bosch, 2011.
- Commissie-Korsten, *De kunst van het balanceren - Visitatierapport hbo-opleidingen bestuurskunde en overheidsmanagement*, Den Haag, 2003.
- Commissie-Korsten, *Ontwikkeling door visie - Midterm review-commissie opleidingen bestuurskunde aan de Erasmus Universiteit*, Rotterdam, 2007.
- Cörvers, R.J.M., *Netwerksturing bij natuurontwikkeling - Bestuurskundige consequentieanalyse bij gebiedsgerichte projecten*, Maastricht, 2001.
- Cörvers, R.J.M., P. Glasbergen en A.F.A. Korsten, Netwerksturing bij natuurontwikkeling, in: *Bestuurswetenschappen*, 2003, nr. 3, pp. 205-228.
- Cousins, J.B., Commentary: Minimizing evaluation misuse as principled practice, in: *American Journal of Evaluation*, vol. 25, 2004, nr. 3, pp. 391-397.
- Daemen, H. en Th. Van der Krogt, Vier functies van internationale accreditatie, in: Kickert, W.J.M., L. Cachet, F.B. van der Meer en L. Schaap, *Liefde voor het openbaar bestuur en liefde voor de bestuurskunde*, Eburon, Delft, 2007, pp. 105-119.
- Dal Santo, T., S. Goldberg, P. Choice & M.J. Austin, Exploratory Research in Public Social Service Agencies - An Assessment of Dissemination and Utilization, in: *Journal of Sociology and Social Welfare*, vol. 29, 2002, nr. 4, pp. 59-81.
- Daniels, M.R., *Terminating Public Programs*, M.E. Sharpe, Armonk, 1997.

Denhardt, K.D. & M.P. Aristigueta, Performance Management Systems: Providing Accountability and Challenging Collaboration, in: Dooren, W. van, en S. van de Walle (eds.), *Performance Information in the Public Sector- How it is used*, Palgrave MacMillan, Londen, 2008, pp. 106-125.

Derksen, W., A.F.A. Korsten en A.F.M. Bertrand (red.), *De praktijk van onderzoek*, Wolters-Noordhoff, Groningen, 1988.

Derksen, W., *Kennis en beleid verbinden*, Boom/Lemma, Den Haag, 2011.

Dijk, A. van, en H. Woldendorp, Sturen in complexe organisaties, in: A. van Dijk, G. Smid en H. Woldendorp (red.), *Sturen in complexe organisaties- Theorie en praktijk van het hoger onderwijs*, Assen, 1993, pp. 11-21.

Dill, D.D., Accreditation, Assessment, Anarchy? The Evolution of Academic Quality Assurance in the US, in: Brennan, J. e.a., *Standards and Qualities in Higher Education*, Jessica Kingsley Publ., Londen, 1997, pp. 15-44.

Dittrich, K., *Accreditatie en de mogelijke gevolgen ervan*, speech HvA/UvA, 21 okt. 2003.

Dittrich, K., Het nieuwe Nederlandse accreditatiestelsel, in: TH&MA- *Tijdschrift voor Hoger onderwijs en Management*, 2010, nr. 4, pp. 23-26.

Doelen, F.C.J. van der, Cultuur en structuur van beleidsevaluaties, in: *Beleidswetenschap*, jrg. 3, 1989, nr. 4, pp. 342-358.

Doelen, F.C.J. van der, Legitiem en effectief overheidsbeleid: geven en nemen, in: *Beleid en Maatschappij*, jrg. 18, 1991, nr. 2, pp. 90-100.

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

Dooren, W. van, en S. van de Walle, Zelfevaluatie of zelfpromotie? – Een onderzoek naar de motieven voor het gebruik van een kwaliteitsmodel: het Common Assessment Framework, in: *Bestuurskunde*, jrg. 13, maart 2004, nr. 2, pp. 87-98.

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

Dooren, W. van, *Performance Measurement in the Flemish Public Sector*, Instituut voor de Overheid, Leuven, 2006.

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 (zelf)diagnose - Beschouwing over de werking van het stelsel*, Sdu, Den Haag, 2004.

Driessen, P.P.J., *Sturen op kwaliteit*, Utrecht, 2005 (oratie).

Drooglever Fortuijn, E., *Onderwijsbeleid – Maatschappelijke functies en strategische keuzen*, Aksant, Amsterdam, 2003.

Dubnick, M., Accountability and the Promise of Performance, in Search of the Mechanisms, in: *Public Performance and Management Review*, vol. 28, 2005, nr. 3, pp. 376-417.

Dunn, W.N., *Public Policy Analysis*, Pearson, Boston, 2012 (5e editie).

Edelenbos, J. en A. van Buuren, Evalueren als leerproces, in: *Bestuurskunde*, jrg. 14, sept. 2005, nr. 6, pp. 2-13.

Edelenbos, J. en M. van Eeten, the Missing Link- Processing Evaluation in Dialogical Evaluation, in: *Evaluation*, vol. 7, 2001, nr. 2, pp. 211-217.

Edelenbos, J., M. van Buuren, M. van As e.a., *Milieubeleid meer van, voor en door de burger – Resultaten van de lerende evaluatie van het Stimuleringsprogramma Burger en Milieubeleid*, Erasmus Universiteit, Rotterdam, 2003.

Eeten, M. van, *Dialogues of the Deaf – Defining New Agenda's for Environmental Deadlocks*, Eburon, Delft, 1999.

Eeten, M. van, en K. Termeer, Het faciliteren van interactieprocessen in beleidsnetwerken, in: *Werken, leren en leven met groepen*, okt. 1996a, C3100, pp. 1-30.

Eeten, M. van, en K. Termeer, Routebeschrijvingen door het gebied van netwerkmanagement, in: *Werken, leren en leven met groepen*, Samsom, Alphen, okt. 1996b, E2200, pp. 1-9.

Ehren, M., Het spel der neveneffecten. Strategisch gedrag van toezichthouders en scholen, in: Leeuw, F.L., J.S. Kerseboom en R. Elte (red.), *Turven, tellen toetsen*, Boom Juridische uitgevers, Den Haag, 2007, pp. 195-203.

Ellemers, J.E., Veel kunnen verklaren of iets kunnen veranderen: krachtige versus manipuleerbare variabelen, in: *Beleid en Maatschappij*, nov. 1976, pp. 281-291.

- Ellemers, J.E., Veel kunnen verklaren of iets kunnen veranderen: krachtige versus manipuleerbare variabelen, in: Lehning, P.B. en J.B.D. Simonis (red.), *Handboek beleidswetenschap*, Boom, Meppel, 1987, pp. 206-226.
- Elsacker, W.J., *Roles of performance management in local government*, EPM Sneek, Groningen, 2007.
- Elte, R., Het ministerie van Onderwijs en Wetenschappen, in: Breunese, J. en L.J. Roborgh (red.), *Ministeries van algemeen bestuur*, Spruyt, Van Mantgem en De Does, Leiden, 1989, pp. 123-139.
- Elte, R., Winst- en verliesrekening van bureaucratisering van het toezicht op onderwijs, in: Leeuw, F.L., J.S. Kerseboom en R. Elte (red.), *Turven, tellen toetsen*, Boom Juridische uitgevers, Den Haag, 2007, pp. 179-193.
- Engbersen, G., *Fatale remedies – Over onbedoelde gevolgen van beleid en kennis*, Pallas Publications, Amsterdam University Press, Amsterdam, 2009.
- Engbersen, G., Ontmantel de visitatie-industrie, in: *Beleid en Maatschappij*, 2010, nr. 2, pp. 161-164.
- Engelen, E., De manager heeft gewonnen, in: *NRC*, 18 juli 2012, p. 15.
- Erasmus Universiteit Rotterdam– Rijksuniversiteit Leiden, *Zelfstudie Bestuurskunde ten behoeve van Visitatie onderwijs 1997*, Rotterdam/Leiden, 1997.
- Evaluatiecommissie Wet Milieubeheer, *Achterafbekeken*, Den Haag, dec. 2007.
- Farnworth, W., *The Legal Analyst- A Toolkit for Thinking about the Law*, University of Chicago Press, Chicago, 2007.
- Farrington, D.P. en B.C. Welsh, Het belang van experimentele evaluaties in de criminologie, in: *Justitiële Verkenningen*, jrg. 31, 2005, nr. 8, pp. 11-42.
- Fijnaut, C., H. Boutellier e.a., *Veel te winnen- Visitatiecommissie Veiligheid Eindhoven*, Gemeente Eindhoven, 2008.
- Foppen J.W., *Gistend beleid – Veertig jaar universitaire onderwijspolitiek*, Vuga, Den Haag, 1989.
- Francis, G. & J. Holloway, What have we learned? Themes from the literature on best-practice benchmarking, in: *International Journal of Management Reviews*, vol., 2007, nr. 3, pp. 171-189.
- Frederiks, M.M.H., *Beslissen over kwaliteit: verklaring van besluitvorming over het gebruik van onderwijsvisitatierapporten in Nederlandse universiteiten*, CHEPS/De Tijdstroom, Enschede, 1996.
- Frey, B.S. & M. Osterloh, Gaat de wetenschap ten onder aan de beoordeelziekte evaluïtis?, in: *NRC*, 29 november 2008.
- Frey, B.S. & M. Osterloh, Onderzoeksevaluaties: verborgen kosten, twijfelachtige voordelen en betere alternatieven, in: Jansen, Th., G. van den Brink en J. Kole (red.), *Beroepstrots- Een ongekende kracht*, Boom, Amsterdam, 2009, pp. 194-219.
- Friedrich, H., *Staatliche Verwaltung und Wissenschaft*, Europäische Verlagsanstalt, Frankfurt am Main, 1970.
- Gardner, H., M. Csikszentmihalyi & W. Damon, Het GoodWork Project®: enkele resultaten van vijftien jaar onderzoek, in: Jansen, Th., G. van den Brink en J. Kole (red.), *Beroepstrots- Een ongekende kracht*, Boom, Amsterdam, 2009, pp. 48-77.
- Garré, P., Externe verantwoording, kwaliteitsverbetering en –verankering in het hoger onderwijs. Accreditering: hefboom of keurslijf, droom of illusie?, in: *Tijdschrift voor Onderwijsrecht en Onderwijsbeleid*, sept.-okt. 2002, nr. 1, pp. 53-60.
- Gerrichhauzen, J.T.G., A.F.A. Korsten en H.J.L. Fijen, De lerende organisatie in een ambtelijke cultuur, in: *Bestuurswetenschappen*, juni 2002, nr. 3, pp. 221-236.
- Gerritsen, E., *De slimme gemeente nader beschouwd*, Nicis, Amsterdam UP, Amsterdam, 2011.
- Geul, A., *Pragmatisme en pragmatiek- Perspectieven op beleidsconstructie*, Lemma, Utrecht, 2006.
- Glasbergen, P., Beleidswetenschappelijk milieuonderzoek: profiel, kernvragen en betekenis, in: Boersema, J.J., T. Pulles, J. van der Straaten, J. Bertels (red.), *De oogst van milieu*, Boom, Amsterdam, 2003, pp. 250-267.
- Glasbergen, P., Voortgang en vooruitgang in de bestuurskunde, in: *Bestuurswetenschappen*, jrg. 42, 1988, nr. 2, pp. 86-97.
- Glasbergen, P.G. & A. Blowers (ed.), *Environmental policy in an international context*, Arnold, Londen, 1995 (drie delen: perspectives, conflicts, prospects).
- Glasbergen, P.G. (ed.), *Co-operative environmental governance - Public-private agreements as a policy strategy*, Kluwer Academic Publishers, Dordrecht, 1998.
- Glasbergen, P.G. (ed.), *Managing environmental disputes - Network management as an alternative*, Kluwer Academic Publ., Dordrecht, 1995.
- Glasbergen, P.G. (red.), *Milieubeleid*, Vuga, Den Haag, 1994.
- Glasbergen, P.G. (red.), *Milieubeleid*, Vuga, Den Haag, 1994.

Glasbergen, P.G. , Beleidsuitvoering als probleem: oorzaken en perspectieven, in: Lehning, P.B. en J.B.D. Simonis (red.), *Handboek beleidswetenschap*, Boom, Meppel, 1987, pp. 80-94.

Glasbergen, P.G. e.a., *Evaluatie meerjarenafspraken over energie-efficiency*, Ministerie van EZ, Den Haag, 1997.

Glasbergen, P.G. en C. Kuijpers, Beleidsontwikking: de case van het grondwaterbeleid, in: *Beleid en Maatschappij*, 1985, nr. 12, pp. 324-332.

Glasbergen, P.G. en P. Driessen, Integrale omgevingsplanning op regionale schaal, in: *Milieu*, jrg. 6, 1991, pp. 76-80.

Glasbergen, P.G. en P.P.J. Driessen (red.), *Innovatie in het gebiedsgericht beleid*, Den Haag, 1993.

Glasbergen, P.G. en R. Groenberg, Partners in duurzame energie – Samenwerkingsrelaties tussen milieuorganisaties en het bedrijfsleven, in: *Bestuurswetenschappen*, 2000, nr. 6, pp. 462-482.

Glasbergen, P.G., *Beleidsnetwerken rondom milieuproblemen*, Vuga, Den Haag, 1989.

Glasbergen, P.G., *Beleidsnetwerken rondom milieuproblemen*, Vuga, Den Haag, 1989.

Glasbergen, P.G., Beleidsveranderingen en leren in het milieubeleid, in: Eberg, J. en H. van de Graaf (red.), *Leren met beleid*, Amsterdam, Het Spinhuis, 1996, pp. 67-83.

Glasbergen, P.G., Besturingsproblemen rond watersystemen, in: *Bestuurswetenschappen*, 1990, nr. 1, pp. 41-52.

Glasbergen, P.G., Geïntegreerd milieubeleid; fictie of werkelijkheid?, in: Leroy, P. (red.), *Milieubeleid in de jaren negentig; integratie van grijs en groen*, Nijmegen, 1990, pp. 22-34.

Glasbergen, P.G., Milieuvernieuwing: een nieuwe sturingsstrategie, in: Nelissen, N.J.M. e.a. (red.), *In staat van vernieuwing*, Coutinho, Muiderberg, 1996, pp. 177-199.

Glasbergen, P.G., P. Driessen en J. van der Veen, Gebiedsgericht milieubeleid; reflectie op een sturingsconcept, in: *Milieu en Recht*, 1992, nr. 7/8, 393-402.

Glasbergen, P.G., Ruimtelijk dwarsliggen- Revisie op een planningstelsel in revisie, in: *Bestuurskunde*, 1996, nr. 6, pp. 286-296.

Glasbergen, P.G., Seven steps towards an instrumentation theory for environmental policy, in: *Policy and Politics*, 1992, nr. 3, pp. 191-200.

Glasbergen, P.G., *Visies op beleid - Sociaal-wetenschappelijke analyse van overheidsbeleid*, Kobra, Amsterdam, 1984.

Goedegebuure, L., O.A.M. Maassen en D.F. Westerheijden (eds.), *Peer Review and Performance Indicators – Quality Assessment in British and Dutch Higher Education*, Cheps, Enschede, 1990.

Gooren, W.A.J. , m.m.v. I. Snellen, A.F.A. Korsten en R. Verstegen, *Onderzoek zaaien, innovatie oogsten - Eindverslag van een evaluatie-onderzoek van de Stichting Technische Wetenschappen*, IVA/KUN, Tilburg, 1984.

Gooren, W.A.J. en A.F.A. Korsten, *Kennisoverdracht aan kleine en middelgrote ondernemingen via transferpunten in Nederland*, Staatsuitgeverij, Den Haag, 1983.

Gooren, W.A.J., A.F.A. Korsten en F. Prakke, *Evaluatie van innovatiebeleid- Een overzicht over uitvoering en wenselijkheid van evaluatie van beleidsvoornemens uit de Innovatienota*, Minister van O&W/Economische Zaken, Den Haag, 1983.

Gooren, W.A.J., Early Warning Evaluation- A Beacon in Pursuing Policy, in: Snellen, I.Th.M. (red.), *Limits of Government – Dutch Experiences*, Kobra, Amsterdam, 1985, pp. 187-212.

Goossens, J., P. Fröhling en M. Laman, *De taakuitvoering van visitatiecommissies in het hoger onderwijs 1999-2001*, Inspectie voor het Onderwijs, mei 2001 (concept).

Graaf, H. van de, en R. Hoppe, *Beleid en politiek*, Coutinho, Muiderberg, 1989.

Gray, P. & P. 't Hart (eds.), *Public Policy Disasters in Europe*, Routledge, Londen, 1998.

Greener, I., *Public Management – A Critical Text*, Palgrave MacMillan, Londen, 2009.

Greve, H., *Organizational learning from performance feedback: A behavioral perspective on innovation and change*, Cambridge University Press, Cambridge, 2003.

Grinten, T. van der, en P.L. Meurs, Publieke verantwoording in de gezondheidszorg, in: Bakker, W. en K. Yesilkagit (red.), *Publieke verantwoording*, Regimes van inzicht en rekenschap bij de uitvoering van publieke taken, Boom, Amsterdam, 2005.

Grit, K., *Economisering als probleem – Een studie naar de bedrijfsmatige stad en de ondernemende universiteit*, Van Gorcum, Assen, 2000.

Groot, H. de, e.a., *Benchmarking in de publieke sector*, Sdu, Den Haag, 2004.

Grooteman, E., Interdepartementale beleidsvisitaties – Instrument voor het evalueren van beleidsontwikkeling, in: Knaap, P. van der, en A. Schilder (red.), *Resultaatgericht sturen en evalueren*, Sdu, Den Haag, 2004, pp. 215-222.

Grünwald-Schindl, C.A. en A. Kraan-Jetten, Ex post-evaluatieonderzoek, in: Lehning, P.B. en J.B.D. Simonis (red.), *Handboek beleidswetenschap*, Boom, Meppel, 1987, pp. 131-154.

Guba, E.G. & Y.S. Lincoln, *Fourth Generation Evaluation*, Sage, Londen, 1989.

Habermas, J., *Theorie des kommunikativen Handelns*, Suhrkamp, Frankfurt, 1987.

Hajer, M.A., J.P.M. van Tatenhove en C. Laurent, *Nieuwe vormen van Governance*, RIVM, Bilthoven, 2004.

Hakvoort, J. en H. Klaassen, Benchmarking: resultaten vergelijken, in: Knaap, P. van der, en A. Schilder (red.), *Resultaatgericht sturen en evalueren*, Sdu, Den Haag, 2004, pp. 205-215.

Hakvoort, J. en H. Klaassen, Doelstellingen in de publieke sector- Nut of noodzaak?, in: *Management en Organisatie*, jrg. 56, nov./dec. 2002, pp. 43-55.

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

Halachmi, A., Performance Measurement, Accountability, and Improved Performance, in: *Public Performance and Management Review*, vol. 25, 2002b, nr. 4, pp. 370-374.

Halachmi, A., Who Gets What, When and How: Performance Measures for Accountability?; for Improved Performance?, in: *International Review of Public Administration*, vol. 7, 2002a, nr. 1, pp. 85-95.

Hanemaayer, D., *Ex ante evaluatie in Nederland- De stand van zaken*, BoomLemma, Den Haag, 2012.

Hanney, S.R., M.A. Gonzalez-Blok, M.J. Buxton & M. Kogan, The utilization of health research in policy-making: concepts, examples and methods of assessment, in: *Health Research and Policy Systems*, vol.1, 2003, nr. 3, pp. 1-28.

Hart, P. 't, J. Hufen en M.J. van Duin, De lerende overheid: mogelijkheden en grenzen van een modieuze metafoor, in: *Beleid en Maatschappij*, jrg. 15, 1988, nr. 2, pp. 83-103.

Hart, P. 't, P. de Jong en A.F.A. Korsten (red.), *Groepsdenken in het openbaar bestuur*, Samsom, Alphen, 1991.

Haselbekke, A.G.J., H.L. Klaassen, A.P. Ros en R.J. in 't Veld, *Prestaties tellen – Kengetallen als instrument voor een bedrijfsmatig(er) bestuur en beheer van decentrale overheden*, VNG Uitgeverij, Den Haag, 1990.

Hatry, H., Epilogue: The Many Faces of Use, in: Dooren, W. van, en S. van de Walle (eds.), *Performance Information in the Public Sector- How it is used*, Palgrave MacMillan, Londen, 2008, pp. 227-241.

Hatry, H., Performance Measurement Principles and Techniques: An Overview for Local Government, in: Kearny, R. & E.V. Berman (ed.), *Public Sector Performance – Management, Motivation and Measurement*, ASPA, Westview Press, Boulder, 1999, pp. 304-329.

Haverland, M. en D. Yanow, Hoe overleef ik discussies over bestuurskundig onderzoek?, in: *Bestuurskunde*, 2010, nr. 4, jrg. 19, pp. 87-95.

Heffen, O. van, Veronderstellingen achter beleid en hun betekenis, in: *Strategie en beleid in de publieke sector*, Samsom, Alphen, mei 1996, D2020.

Heiber, H., *Messung von Forschungsleistungen der Hochschulen*, Nomos Verlagsgesellschaft, Baden-Baden, 1983.

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

Helden, G.J. van, & A. Johnson, A comparative analysis in the development of performance-based management systems in Dutch and Norwegian local government, in: *International Public Management Journal*, vol. 5, 2002, nr. 1, pp. 75-95.

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

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

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

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

Helden, G.J. van, en S. Tillema, Naar een benchmarkingtheorie voor de publieke sector, in: *Tijdschrift voor Openbare Financien*, jrg. 36, 2004, nr. 3, pp. 99-113.

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

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

Helden, J. van, S. Tillema en R. van Esch, Benchmarking in de publieke sector – Ervaringen met de bedrijfsvergelijking zuiveringsbeheer, in: *Economienblad*, juni 2005, pp. 13-16.

Hellendoorn, J.C. (red.), *Evaluatiemethoden ex ante*, Ministerie van Financiën, Sdu, Den Haag, 2001 (5^e druk).

Hemerijck, A., Vier beleidsvragen, in: Bekkers, V. en A. Ringeling (red.), *Vragen over beleid-Perspectieven op waardering*, Lemma, Utrecht, 2003, pp. 33-49.

Henry, G. & M. Mark, Beyond use: Understanding Evaluation's Influence on Attitudes and actions, in: *American Journal of Evaluation*, vol. 24, 2003, nr. 3.

Herweijer, M., De dynamiek van doelstellend gedrag: een struikelblok bij evaluatie, in: *Bestuurswetenschappen*, jrg. 35, 1981, nr. 5, pp. 348-366.

Heuvel, H. van den, Beleidsinstrumenten: naar strategisch beleid in netwerken, in: Abma, T. en R.J. in 't Veld (red.), *Handboek beleidswetenschap*, Boom, Amsterdam, 2001, pp. 281-292.

Heuvelhof, E.F. ten, en C.M. Nauta, Milieueffectrapportage als leerproces, in: *Beleidswetenschap*, 1997, nr. 1, pp. 25-41.

Hill, M., *The Public Policy Process*, Pearson, Harlow, 2009 (5^e editie).

Hoesel, P. H.M. van, F.L. Leeuw en J.W.M. Mevissen (red.), *Beleidsonderzoek in Nederland*, Van Gorcum, Assen, 2005.

Hoesel, P. van, *Beleidsonderzoek als professie- Een lang gekoesterd ideaal*, Stichting Economisch Instituut voor het Midden- en Kleinbedrijf, Zoetermeer, 2003 (oratie)

Hoesel, P. van, Methoden van beleidsonderzoek in de toekomst, in: Boekenoogen, M., P. van Hoesel, Y. Prince en C. Verheijen (red.), *Methoden van beleidsonderzoekers: creatief en oplossingsgericht*, Lemma, Den Haag, 2009, pp. 417-429.

Hogwood, B.W. & B.G. Peters, *Policy Dynamics*, Wheatsheaf Books Ltd., Brighton, 1983.

Hollander, A.E.M. en A.H. Hanemaayer, Management van gezondheidsrisico's: mogelijkheden en beperkingen van kosten-batenanalyses, in: *Justitiële Verkenningen*, jrg. 29, 2003, nr. 9, pp. 70-83.

Hood, C., Public Service Management by Numbers: Why Does it Vary? Where has it Come From? What are the Gaps and the Puzzles?, in: *Public Money and Management*, vol. 27, 2007a, nr. 2, pp. 95-102.

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 Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*, Princeton University Press, Princeton, 2011.

Hood, C., What Happens when Transparency meets Blame-Avoidance?, in: *Public Management Review*, vol. 9, 2007b, nr. 2, pp. 191-210.

Hoogerwerf, A. (red.), *Succes en falen van overheidsbeleid*, Samsom, Alphen, 1983.

Hoogerwerf, A. en M. Herweijer (red.), *Overheidsbeleid*, Samsom, Alphen, 1998 (6^e druk)(2003, 7^e druk).

Hoogerwerf, A., Beleid berust op veronderstellingen: de beleidstheorie, in: *Acta Politica*, jrg. 19, 1984, nr. 4, pp. 493-531.

Hoogerwerf, A., De beleidstheorie uit de beleidspraktijk: een tussenbalans, in: *Beleidswetenschap*, 1989, pp. 320-341.

Hoppe, R. e.a., *Beleidsnota's die (door)werken*, Coutinho, Bussum, 1998.

Hoppe, R. en A. Peterse (red.), *Bouwstenen voor argumentatieve beleidsanalyse*, Elsevier, Den Haag, 1998.

Hoppe, R. en S. Huijs, *Werk op de grens tussen wetenschap en beleid: paradoxen en dilemma's*, RMNO, Den Haag, 2003.

Hoppe, R. en W. Halfman, Wetenschappelijke beleidsadvisering in Nederland. Trends en ontwikkelingen, in: *Beleidswetenschap*, jrg. 18, 2004, nr. 1, pp. 31-61.

Hoppe, R., 'Anorexia consulta': afslanking infrastructuur rijksdienst, deel 2, in: *Beleid en Maatschappij*, jrg. 34, 2007, nr.4, pp. 238-250.

Hoppe, R., From 'knowledge use' to 'boundary work'; sketch of an emerging new agenda for inquiry into science-policy interaction, in: Veld, R.J. in 't (ed.), *Knowledge Democracy: Consequences for Science, Politics and Media*, Heidelberg, 2010, pp. 169-186.

Hoppe, R., Na 'doorwerking' naar 'grenzenwerk', in: *Bestuurskunde*, 2008, nr. 2, pp. 15-26.

Hoppe, R., *The Governance of Problems – Puzzling, Powering, Participation*, Policy Press, Bristol, 2011.

Hoppe, R., *Van flipperkast naar grensverkeer- Veranderende visies op de relatie tussen wetenschap en beleid*, AWT, Den Haag, 2002.

Hoppe, R., Waarom uitvechten uitdenken overtroeft, in: *Bestuurskunde*, jrg. 21, 2012, nr. 2, pp. 7-18.

Hoppe, R., Werken op de grens tussen wetenschap en politiek: naar een typologie van grensarrangementen, in: *Beleidswetenschap*, jrg. 17, 2003, nr. 2, pp. 144-170.

- House, E. & K. Howe, *Deliberative Democratic Evaluation*, in: *New Directions for Evaluation*, Sage, Londen, 2000, pp. 3-12.
- Huisman, H., Hoe verander je de stijl van management bij resultaatgerichte organisaties?, in: *Bestuurskunde*, jrg. 13, nov. 2004, nr. 7, pp. 301-305.
- Huisman, J. & H. Theisens, Marktwerking in het hoger onderwijs, in: *Beleid & Maatschappij*, jrg. 28, 2001, nr. 3, pp. 165-177.
- Huisman, J. en F. de Vijlder, Hoger onderwijs en beleid in Nederland, in: *Bestuurskunde*, jrg. 8, 1999, nr. 8, pp. 357-367.
- Huisman, J., De regulering van het opleidingsaanbod. Een slingerbeweging tussen overheidsplanning en zelfregulering, in: *Beleidswetenschap*, jrg. 11, 1997, nr. 2, pp. 122-142.
- Huisman, J., *Differentiation, diversity and dependency in higher education*, CHEPS, Utrecht, 1995.
- Huisman, P. en F. de Vijlder, *Sectorstudie toezicht hoger onderwijs*, WRR, Den Haag, 2012 (Webpublicatie 64).
- Huitema, D., M. Jeliaskova en D.F.Westerheijden, Fasen, niveaus en cirkels bij beleidsontwikkeling-Hogeronderwijsbeleid en kwaliteitsinterne milieuzorg, in: *Beleidswetenschap*, jrg. 17, 2003, nr. 1, 29-51.
- Hulshof, M., J. Goossens en J. Meulesteen, Een archeologisch punt ... De kwaliteit van visitatierapporten, in: Hulshof, M., B. Bulder e.a. (red.), *Toezichthouder: lam of leeuw?*, Inspectie van het Onderwijs, Den Haag, 2003, pp. 21-37.
- Hupe, P.L. en F.B. van der Meer, Cruciale koppelingen. Over doorwerking van emancipatie-effectrapportages in overheidsprocessen, in: *Beleidswetenschap*, jrg. 20, 2006, nr. 2, pp. 3-21.
- Hupe, P.L., De mythe van de beleidsrelevantie, in: *Beleid en Maatschappij*, oktober 1984, pp. 278-294.
- Hupe, P.L., *Overheidsbeleid als politiek*, Van Gorcum, Assen, 2007.
- Huysman, M. en R. van der Vlist, Naar een 'organizational learning'-benadering van de lerende organisatie, in: *Gedrag & Organisatie*, jrg. 11, okt. 1998, nr. 5, pp. 219-232.
- in: *Justitiële Verkenningen*, jrg. 31, 2005, nr. 8, pp. 72-85.
- Inspectie van het Onderwijs, *Onderwijsverslag 2000*, Utrecht, 2001.
- Jansen, D. & A. Wald, Netzwerke, in: Benz, A., S. Lütz, U. Schimank & G. Simonis (Hrsg.), *Handbuch Governance*, VS Verlag, Wiesbaden, 2007a, pp. 93-106.
- Jansen, D. & A. Wald, Netzwerktheorien, in: Benz, A., S. Lütz, U. Schimank & G. Simonis (Hrsg.), *Handbuch Governance*, VS Verlag, Wiesbaden, 2007b, pp. 188-200.
- Jansen, Th., G. van den Brink en J. Kole (red.), *Beroepstrots- Een ongekende kracht*, Boom, Amsterdam, 2009.
- Jansen, Th., G. van den Brink en R. Kneyber (red.), *Gezagsdragers - De publieke zaak op zoek naar haar verdedigers*, Boom, Amsterdam, 2012.
- Jeliaskova, M. & D. Westerheijden, The Netherlands: A Leader in Quality Assurance Follows the Accreditation Trend, in: Schwarz, S. en D. Westerheijden (eds.), *Accreditation and Evaluation in the European Higher Education Area*, Kluwer Academic Publ., Dordrecht, 2004, pp. 323-346.
- Jeliaskova, M. & D.F. Westerheijden (red.), *Het zichtbare eindresultaat. Doorwerking van onderwijsvisitaties in hogeronderwijsinstellingen*, Algemene Rekenkamer, Den Haag, 17 februari 2000.
- Jeliaskova, M., & D.F. Westerheijden, Systemic adaptation to a changing environment- Towards a next generation of quality assurance models, in: *Higher Education*, vol. 44, 2002, nr. 3-4, pp. 433-448.
- Jeliaskova, M.I., Running the maze: interpreting external review recommendations, in: *Quality of Higher Education*, vol. 8, 2002, nr. 1, pp. 89-96.
- Jenniskens, I.G.M., *Governmental Steering and Curriculum Innovation - A Comparative Study of the Relationship between Governmental Steering Instruments and Innovation in Higher Education Curricula*, Elsevier/De Tijdstroom, Maarssen, 1997.
- Jensma, F., Interview met Grahame Lock 'De meeste mensen kunnen ook toe met minder weten', in: *NRC*, 23 febr. 2008.
- Johnson, K., L.O. Greenesid, S.A. Toial, J.A.King, F. Lawrenz & B. Volkov, Research on Evaluation Use - A Review of the Empirical Literature from 1986 to 2005, in: *American Journal of Evaluation*, 2009, pp. 307-410.
- Jong, A.H.M. de, en D.L. Kabel, De rol van evaluaties in het begrotingsproces, in: *Justitiële Verkenningen*, jrg. 31, 2005, nr. 8, pp. 100-110.
- Jongbloed, B.W.A., Prestatiesubsidiëring in het hoger onderwijs: een internationale vergelijking, in: *Openbare Uitgaven*, 1999, nr. 4, pp. 212-221.

Jongenelen, J. en F. Winterterp, Methodentriangulatie in beleidsevaluatie, in: Boekenoogen, M., P. van Hoesel, Y. Prince en C. Verheijen (red.), *Methoden van beleidsonderzoekers: creatief en oplossingsgericht*, Lemma, Den Haag, 2009, pp. 253-261.

Karsten, S., A.B. Dijkstra en R.J. Bosker, Openbaarmaking van prestatiegegevens in het basis- en voortgezet onderwijs- Een balans na zeven jaar, in: *Pedagogische Studiën*, jrg. 81, 2004, pp. 397-403.

Katholieke Universiteit Brabant, Faculteit der Sociale Wetenschappen, *Zelfstudie opleiding Beleids- en Organisationswetenschappen ten behoeve van de Onderwijsvisitatie Bestuurs- en Beleidswetenschappen in 1997*, Tilburg, 1997.

Kearny, R. & E.V. Berman (ed.), *Public Sector Performance – Management, Motivation and Measurement*, ASPA, Westview Press, Boulder, 1999.

Kenis, P. & K.G. Provan, Towards an Exogeneous Theory of Public Network Performance, in: *Public Administration*, vol. 87, 2009, nr. 3, pp. 440-457.

Kerseboom, J., F. Leeuw en R. Elte, Een mozaïek van toezicht, inspectie, handhaving en evaluatie in tien thema's, in: Leeuw, F.L., J.S. Kerseboom en R. Elte (red.), *Turven, tellen, toetsen*, Boom, Den Haag, 2007, pp. 9-29.

Kerseboom, J., *Leren van evalueren*, in: *Bestuurskunde*, jrg. 17, 2008, nr. 2, pp.41-47.

Kickert, W.J.M., E.H. Klijn en J.F.M. Koppenjan (eds.), *Managing Complex Networks: Strategies for the Public Sector*, Sage, Londen, 1997.

Kijzers, A., P. van der Knaap en B. van de Putte, Evaluatieonderzoek ex post, in: Knaap, P. van der, en A. Schilder (red.), *Resultaatgericht sturen en evalueren*, Sdu, Den Haag, 2004, pp. 119-139.

Kingdon, J.W., *Agendas, Alternatives and Public Policies*, Longman, New York, 2003

Kirkhart, K., Reconceptualizing evaluation use: an integrated theory of influence, in: *New Directions for Evaluation*, Jossey-Bass, San Francisco, 2000, pp. 5-23.

Klein Haarhuis, C., M. van Ooyen-Houben, E. Kleemans en F. Leeuw, Rechtshandhaving geëvalueerd; een synthese van 31 onderzoeken, *Justitiële Verkenningen*, jrg. 31, 2005, nr. 8, pp.54-72.

Klein Haarhuis, C.M. & E. Niemeijer, *Wet en werkelijkheid - Bevindingen uit evaluaties van wetten*, WODC, Boom, Den Haag, 2008.

Klijn, E.H. en G. Teisman, Effective policy making in a multi-actor setting: networks and steering, in: R. in 't Veld e.a. (eds.), *Autopoiesis and configuration theory: new approaches to societal steering*, Kluwer, Dordrecht, 1991, 99-113.

Klijn, E.H. en G.R. Teisman, Besluitvorming in beleidsnetwerken, in: *Beleidswetenschap*, 1992, nr. 1, pp. 32-52.

Klijn, E.H. en J. Koppenjan, Besluitvormingen management in netwerken: een multi-actor perspectief op sturing, in: Abma, T. en R.J. in 't Veld (red.), *Handboek beleidswetenschap*, Boom, Amsterdam, 2001, pp. 179-193.

Klijn, E.H. en J.F.M. Koppenjan, Beleidsnetwerken als theoretische benadering: een tussenbalans, in: *Beleidswetenschap*, 1997, nr. 2, pp. 143-167.

Klijn, E.H., J.F.M. Koppenjan en C.A.J.M. Termeer, Managing Networks in the Public Sector, in: *Public Administration*, vol. 73, 1995, nr. 3, pp. 437-454.

Klijn, E.H., Netwerken als perspectief op beleid en uitvoering, in: *Beleidswetenschap*, jrg. 19, 2005, nr. 4, pp. 32-55.

Klijn, E.H., Vertrouwen en samenwerking in netwerken, in: *Beleidswetenschap*, jrg. 16, 2002, nr. 3, pp. 259-280.

Kluijtmans, F. (red.), *Leerboek HRM*, Noordhoff Uitgevers, Groningen, 2010.

Knaap, L.M. van der, F. Leeuw, S. Bogaerts & L. Nijssen, Combining Campbell Standards and the Realist Evaluation Approach – The Best of Two Worlds, in: *American Journal of Evaluation*, vol. 29, 2008, pp. 48-57.

Knaap, L.M. van der, L.T.J. Nijssen en S. Bogaerts, *Geweld verslagen?*, Boom, Den Haag, 2006.

Knaap, P. van der, *'Lerende' overheid, intelligent beleid*, Phaedrus, Den Haag, 1997.

Knaap, P. van der, Beleidsevaluatie en het leervermogen van de overheid: lessen uit cybernetica en psychologie, in: Knaap, P. van der, en A. Schilder (red.), *Resultaatgericht sturen en evalueren*, Sdu, Den Haag, 2004, pp. 223-236.

Knaap, P. van der, en A. Schilder (red.), *Resultaatgericht sturen en evalueren*, Sdu, Den Haag, 2004.

Knaap, P. van der, en A. Schilder, Resultaatgericht sturen en evalueren, , in: Knaap, P. van der, en A. Schilder (red.), *Resultaatgericht sturen en evalueren*, Sdu, Den Haag, 2004, pp. 9-17.

Knaap, P. van der, A. Korsten, K. Termeer en M. van Twist (red.), *Trajectmanagement*, Lemma, Utrecht, 2004.

Knippenberg en W. van der Ham, *Een bron van aanhoudende zorg- 75 Jaar ministerie van Onderwijs (Kunsten) en Wetenschappen: 1918-1993*, Van Gorcum, Assen, 1994.

Köbben, A.J.F. en H. Tromp, *De onwelkome boodschap- Of hoe de vrijheid van wetenschap bedreigd wordt*, Uitgeverij Jan Mets, Amsterdam, 1999.

Köbben, A.J.F., *Het gevecht met de engel - Over verheffende en minder verheffende aspecten van het wetenschapsbedrijf*, Mets en Schilt, Amsterdam, 2003.

Koelman, J., *Kosten van de verzuiling*, Erasmus Universiteit, Rotterdam, 1987.

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

Kool, D. de, Monitoring in de publieke sector: De rapportage Water in beeld, in: *Bestuurswetenschappen*, 2011, nr. 1, pp. 21-39.

Kool, D. de, Monitoring in het openbaar bestuur, in: *Overheidsmanagement*, nr. 10, 2007b, pp. 6-10.

Kool, D. de, Monitoring in beeld, in: *Bestuurskunde*, 2007c, nr. 2, pp. 92-103.

Kool, D. de, Rational, Political and Cultural Uses of Performance Monitors: the Case of Dutch Urban policy, in: Dooren, W. van, en S. van de Walle (eds.), *Performance Information in the Public Sector- How it is used*, Palgrave MacMillan, Londen, 2008, pp. 174-192.

Koppenjan, J.F.M., Beleidsonderzoek voor het integraal veiligheidsbeleid: lessen over effectrapportages, in: *Beleidsanalyse*, jrg. 24, 1995, nr. 2, pp. 27-37.

Korsten, A. en F. Leeuw, Handleidingen in het openbaar bestuur, in: *Bestuurskunde*, jrg. 8, 1998, pp. 332-350.

Korsten, A., Effectrapportage, in: *Beleidsanalyse*, jrg. 26, 1997, nr. 1 / 2, pp. 4-7.

Korsten, A.F.A. en I.M.A.M. Pröpper, Beleidsevaluatie, in: Bekke, A. e.a. (red.), *Management in overheidsorganisaties*, Samsom, Alphen, 1993, B1160, pp. 1-14.

Korsten, A.F.A. en I.Th.M. Snellen, Openbare aanbesteding en uitbesteding: evaluatie van de wet ketenaansprakelijkheid, in: Derksen, W., A.F.A. Korsten en A.F.M. Bertrand (red.), *De praktijk van onderzoek*, Wolters-Noordhoff, Groningen, 1988, pp. 473-491.

Korsten, A.F.A. en M. Schoemaker, Gezond bestuur en bestuurlijke probleemgemeenten: Tijd voor de bestuurschout?, in: *Bestuurswetenschappen*, jrg. 65, 2011, nr. 3, pp. 28-49.

Korsten, A.F.A. en P. de Goede (red.), *Bouwen aan vertrouwen in het openbaar bestuur*, Elsevier, Den Haag, 2006d.

Korsten, A.F.A. en R. Hoppe, Van beleidswetenschap naar kennissamenleving, in: *Beleidswetenschap*, jrg. 20, 2006e, nr. 4, pp. 34-73.

Korsten, A.F.A. en Th.A.J. Toonen (red.), *Bestuurskunde - Hoofdfiguren en kernthema's*, Stenfert Kroese, Leiden, 1993.

Korsten, A.F.A., Beoordeling - De betekenis van beleidsnetwerken voor beoordelen van provinciale visitatiecommissies, in: Cörvers, R. en C. Dieperink (red.), *Milieubeleid als sturingsopgave*, Open Universiteit/Universiteit Utrecht, 2011, pp. 201-211.

Korsten, A.F.A., *Bestuurskunde als avontuur*, Kluwer, Deventer, 1988.

Korsten, A.F.A., *Deugdelijk bestuur*, Open Universiteit Nederland, Heerlen, 2010 (afscheidsrede).

Korsten, A.F.A., Effectrapportage, in: *Beleidsanalyse*, 1997, nr. 1/2, pp. 4-8.

Korsten, A.F.A., Evalueren van wetten, in: *Bestuurswetenschappen*, 1981, pp. 462-484.

Korsten, A.F.A., *Grote klasse - Op zoek naar excellente ambtenaren en leiderschap*, Beljon + Westerterp, Heerlen, 2005a.

Korsten, A.F.A., *Maatstaven voor de beoordeling van overheidshandelen*, Open Universiteit, Heerlen, 14 maart 2005b (zie www.arnokorsten.nl)

Korsten, A.F.A., Visitatie goed instrument voor continue verbetering, in: *Provincie en Gemeente*, jrg.55, mei 2004c, nr. 4, pp. 3-5.

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

Korsten, A.F.A., *Visiteren van gemeenten - Evaluatie van de toepassing van de bestuurskrachtmonitor in Limburgse gemeenten*, Open Universiteit Nederland, Heerlen/Maastricht, 2004a.

Korsten, A.F.A., Visiteren, in: *Binnenlands Bestuur*, 2003a, nr. 29-30, p.21.

Korsten, A.F.A., *Wat is goed genoeg?*, Kobra, Amsterdam, 1983.

Kraan-Jetten, A., F.H. Brand en Th.W. de Valk, Ex ante evaluaties: naar een checklist van kwaliteitsnormen, in: *Beleidsanalyse*, 1998, nr. 4, pp. 14-21.

Kraan-Jetten, A., Reconstructie en evaluatie van beleidstheorieën, in: *Beleid en Maatschappij*, 1986, pp. 186-191.

- Kuhlmann, S., J. Bögumil & H. Wollmann (Hrsg.), *Leistungsmessung und – vergleich in Politik und Verwaltung*, VS Verlag, Wiesbaden, 2004.
- Kwakkelstein, T., A. van Dam en A. van Ravenzwaaij (red.), *Van verzorgingsstaat naar waarborgstaat*, Boom Lemma, Den Haag, 2012.
- Kwikkers, P.C., D.F. Westerheijden, N. Vercruyssen, E.V. Schalkwijk, C.J. Peels en J.A. Frederik, *Accreditatie en kwaliteitszorg, De accreditatie van hoger onderwijs in Nederland en Vlaanderen*, Sdu Uitgevers, Den Haag, 2011.
- Laegreid, P. P.G. Roness & K. Rubecksen, Performance Information and Performance Steering: Integrated System or Loose Coupling?, in: Dooren, W. van, en S. van de Walle (eds.), *Performance Information in the Public Sector- How it is used*, Palgrave MacMillan, Londen, 2008, pp. 42-58.
- Laman, M., Façades van Potemkin? – Kwaliteitsverbetering door visitaties, in: Hulshof, M., B. Bulder e.a. (red.), *Toezichthouder: lam of leeuw?*, Inspectie van het Onderwijs, Den Haag, 2003, pp. 37-53.
- Lancer Julnes, P. de, 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 MacMillan, Londen, 2008, pp. 58-72.
- Lancer Julness, P. de & M. Holzer, Promoting the Utilization of Performance Measures in Public Organizations: an Empirical Study of Factor Affecting Adoption and Implementation, in: *Public Administration Review*, vol. 61, 2001, nr. 6, pp. 693-705.
- Landry, R., M. Lamary & N. Amara, The Extent and Determinants of the Utilization of University Research in government Agencies, in: *Public Administration Review*, vol. 63, 2003, nr. 2, pp. 192-205.
- Leeuw, F., Evaluatie-onderzoek, auditing en controle, in: Abma, T. en R.J. in 't Veld (red.), *Handboek beleidswetenschap*, Boom, Amsterdam, 2001, pp. 70-83.
- Leeuw, F.L., G. van Gils, M. Gildemyn en L. Inberg, *Leren van evalueren*, Ministerie van Buitenlandse Zaken, Den Haag, 2009.
- Leeuw, F.L. en G. van Gils, Over benutting van beleidsevaluatie bij het ministerie van Buitenlandse Zaken, in: Aardema, A. e.a. (red.), *Meerwaarde van de bestuurskunde*, Boom/Lemma, Den Haag, 2010, pp. 165-174.
- Leeuw, F.L. en H. van de Graaf, Onderzoek naar beleidstheorieën: uitgangspunten, methodische aspecten en praktische relevantie, in: *Beleid en Maatschappij*, jrg. 15, 1988, nr. 1, pp. 1-7.
- Leeuw, F.L., Beleidstheorieën: veronderstellingen achter beleid, in: Hoogerwerf, A. (red.), *Overheidsbeleid*, Samsom, Alphen, 1993, pp. 89-104.
- Leeuw, F.L., Diminishing returns in evaluation? - Challenges for educational evaluators, in: *Lifelong Learning in Europe*, vol. 9, 2004, nr. 2, pp. 106-114.
- Leeuw, F.L., *Gedragmechanismen achter overheidsinterventies en rechtsregels*, UM, Maastricht, 2008.
- Leeuw, F.L., J.S. Kerseboom en R. Elte (red.), *Turven, tellen, toetsen*, Boom, Den Haag, 2007.
- Leeuw, F.L., R. Rist & R.C. Sonnichsen (eds.), *Can Governments Learn?*, Transaction books, Lexington, 1994.
- Leeuw, F.L., Reconstructing Program Theories: Methods available and Problems to be Solved, in: *American Journal of Evaluation*, vol. 24, 2003, nr.1, pp. 5-20.
- Leeuw, F.L., Toezicht in het hoger onderwijs- Denkkaders voor een woelige tijd, in: *Th&Ma - Tijdschrift voor Hoger Onderwijs en Management*, 2003, nr. 3, pp. 18-23.
- Leeuw, F.L., Visitatie, inspectie en accreditering in het hoger onderwijs, in: *TH&MA- Tijdschrift voor Hoger Onderwijs en Management*, 2001, nr. 5, pp. 44-49.
- Lips, A.M.B., *Autonomie in kwaliteit, Ambigüiteit in bestuurlijke communicatie over de ontwikkeling van kwaliteitszorg in het Hoger Onderwijs*, Eburon, Delft, 1996.
- Lock, G., De onmenselijke maat en de 'crisis in het onderwijs', Inleiding op het symposium van het BON, 4 april 2009.
- Lock, G., The Collectivisation of Dutch Universities, in: *Minerva*, vol. 27, 1989, nr. 2/3, pp. 157-176.
- Luts, M., W. Van Dooren en G. Bouckaert, *Internationale rangschikkingen gerangschikt – Een meta-analyse van rangschikkingen van publieke sectoren*, Steunpunt beleidsrelevant onderzoek- bestuurlijke organisatie Vlaanderen, Leuven, 2008.
- Lütz, S., Policy-Transfer und Policy-Diffusion, in: Benz, A., S. Lütz, U. Schimank & G. Simonis (Hrsg.), *Handbuch Governance*, VS Verlag, Wiesbaden, 2007, pp. 132-144.
- Maarse, J.A.M., *Marktvorming in de zorg*, Maastricht University Press, Maastricht, 2011.
- Mark, M.M. & H.T. Henry, The Mechanisms and Outcomes of Evaluation Influence, in: *Evaluation*, vol. 10, 2004, nr. 1, pp. 35-57.
- Mayne, J., e.a., *Advancing public policy evaluation*, North-Holland, Amsterdam, 1992.

- McDaniel, O.C., *The effects of government policies on higher education- In search of alternative steering methods*, Vuga, Den Haag, 1997.
- Meer, F.B. van der, en J. Edelenbos, Evaluation in Multi-actor Policy Processes - Accountability, Learning and Co-operation, in: *Evaluation*, vol. 12, 2006, nr. 2, pp. 201-218.
- Meer, F.B. van der, G.J. de Vries en G.A.N. Vissers, Evaluatie en leerprocessen bij de overheid: de rol van institutionele condities, in: *Beleidswetenschap*, jrg. 14, 2000, nr. 3, pp. 253-277.
- Meer, F.B. van der, J.F.B. van Gaalen en L. Schaap, Leren van controleren – De rol van de Algemene Rekenkamer, in: *Bestuurskunde*, jrg. 3, 1994, nr. 3, pp. 113-125.
- Meijerink, R., Een gulden eeuw voor universitair onderwijs?, in: *Bestuurskunde*, jrg. 8, 1999, nr. 8, pp. 377-385.
- Mertens, F., *Inspecteren - Toezicht door inspecties*, Sdu Uitgevers, Den Haag, 2011.
- Mertens, F., K. van Schaardenburgh-Verhoeve en S. Sillem (red.), *Methodische aspecten van het onderzoek naar ongevallen*, Eburon, Delft, 2012.
- Mertens, F.J.H., Kwaliteitsbewaking in het Nederlands hoger onderwijs: enkele persistente problemen van het Nederlands hoger onderwijs, in: *Tijdschrift voor Hoger Onderwijs*, jrg. 17, april 1999, nr. 1.
- Mertens, F.J.H., Vertrouwen en controle in het hoger onderwijs, in: Gansewinkel, H., L. Römken en W. Spee (red.), *Complexiteit. Hoera! Kansen en krachten in het hedendaagse onderwijs*, Interstudie NDO, Arnhem, 2008, pp. 103-110.
- Meurs, P.L., E.K. Schrijvers en G.H. de Vries (red.), *Leren van de praktijk – Gebruik van lokale kennis en ervaring voor beleid*, Amsterdam University Press, Amsterdam, 2006.
- Ministerie van O&W, *Hoger onderwijs: autonomie en kwaliteit*, TK 1985-1986, nr. 19253, nr. 1-2.
- Ministerie van OC &W, *Bestuurlijke hantering kwaliteitszorg*. Brief aan het bestuur van de HBO-Raad en de VSNU, Zoetermeer, 3 dec. 1998.
- Ministerie van OC&W, *Keur aan kwaliteit – HOOP 2000*, TK 1999-2000, Den Haag, 2000.
- Ministerie van OC&W, *Kwaliteitszorg hoger onderwijs*, vergaderjaar 1999-2000, 27 010, nrs. 1-2, Sdu Uitgevers, Den Haag, 2000.
- Montfort, C. van, Verantwoording aan departementen, in: Bovens, M. en Th. Schillemans (red.), *Handboek publieke verantwoording*, Lemma, Den Haag, 2009, pp. 81-101.
- Moynihan, D.P., S.K. Pandey & B.E. Wright, Setting the Tabler: How Transformational Leadership Fosters Performance Information Use, in: *Public Administration*, vol. 22, jan. 2012, nr. 1, pp. 143-165.
- Moynihan, D.P., *The Dynamics of Performance Management*, Georgetown University Press, Washington D.C., 2008.
- Mulder, H.P., G. Walraven e.a., Het gebruik van beleidsevaluatie-onderzoek bij de rijksoverheid, in: *Beleidswetenschap*, 1991, nr. 3, pp. 203-277.
- Nationale Accreditatie Organisatie, *Uitgangspunten accreditatie en accreditatiekaders*, Den Haag, maart 2003.
- Nelen, H., *Evidence maze; het doolhof van het evaluatieonderzoek*, UM, 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., P. de Goede en M.J.W. van Twist, *Oog voor openbaar bestuur – Een beknopte geschiedenis van de bestuurskunde*, Elsevier, Den Haag, 2006.
- Newman, D.L., R.D. Brown & L.A. Braskamp, Communication Theory and the Utilization of Evaluation, in: Braskamp, L. & R.D. Brown (eds.), *Utilization of Evaluative Information*, Jossey-Bass Publ., San Francisco, 1980, pp. 29-37.
- 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- Een ongekende kracht*, Boom, Amsterdam, 2009, pp. 115-129.
- Noordegraaf, M. en J. Sterrenburg, Publieke professionals en verantwoordingsdruk, in: Bovens, M. en Th. Schillemans, *Handboek publieke verantwoording*, Lemma, Den Haag, 2009, pp. 231-255.
- Noordegraaf, M. en W.J.M. Kickert, Sturen op afstand in het hoger onderwijs, in: Kickert, W.J.M. (red.), *Veranderingen in management en organisatie bij de rijksoverheid*, Samsom, Alphen, 1993, pp. 77-90.
- Noordegraaf, M., *Management in het publieke domein*, Uitgeverij Coutinho, Bussum, 2004.
- Noordegraaf, M., Prestatiemanagement, in: Noordegraaf, M., K. Geuijen en A. Meijer (red.), *Handboek Publiek Management*, Boom/Lemma, Den Haag, 2011, pp. 365-397.
- Noordegraaf, M., *Professioneel bestuur*, Lemma, Utrecht, 2008.
- NVAO, *Accreditatiekaders, website NVAO*; www. NVAO.nl.
- Oijen, P. van, I.Th.M. Snellen en J.M. van Westerlaak, *Ambtenaren en onderzoekers*, Leiden, 1984.

- Onderwijsraad, *Bureaucratisering in het onderwijs*, Den Haag, 2004.
- Onderwijsraad, *Hoger onderwijs voor de toekomst*, Den Haag, 2010.
- Otten, H.M.P., *Verstrikt in grote projecten- Hoe de stadhuizen in Amsterdam en Apeldoorn tot stand kwamen*, VNG Uitgeverij, Den Haag, 2000.
- Otten, M.H.P., De Stopera – Een prototype van verstriking?, in: *Bestuurskunde*, jrg. 3, 1994, nr. 3, pp. 86-95.
- Parlementaire commissie Privatisering/Verzelfstandiging overheidsdiensten (commissie-Kuiper), *Verbinding verbroken?*, Den Haag, 2012.
- Parsons, W., *Public Policy*, Edward Elgar, Aldershot, 1995.
- Patton, M.Q., *Practical evaluation*, Sage, Londen, 1982.
- Patton, M.Q., *Utilization-focused Evaluation*, Sage, Thousand Oaks, 1997.
- Pawson, R. & N. Tilly, *Realistic Evaluation*, Sage, Londen, 2003.
- Pawson, R. en C. Klein Haarhuis, Evaluatie van complexe programma's: een theoriegestuurde aanpak, in: *Justitiële Verkenningen*, jrg. 31, 2005, nr. 8, pp. 42-54.
- Pawson, R., *Evidence-Based Policy*, Sage, Londen, 2006.
- Peters, K., Impuls voor de lokale democratie?- De casus van de wmo, in: Meurs, P.L., E.K. Schrijvers en G.H. de Vries (red.), *Leren van de praktijk – Gebruik van lokale kennis en ervaring voor beleid*, Amsterdam University Press, Amsterdam, 2006, pp. 41-47.
- Pollitt, C., X. Girre e.a., *Performance or Compliance? – Performance Audit and Public Management in Five Countries*, Oxford University Press, Oxford, 1990.
- Potman, H., *Acceptatie van beleid*, Kerckebosch, Zeist, 1989.
- Power, M., *The Audit Explosion*, Demos, Oxford, 1994.
- Power, M., *The Audit Society - Rituals of verification*, Oxford University Press, Oxford, 1999.
- Power, M., The Theory about the Audit Explosion, in: Ferlie, E., L.E. Lynn jr & C. Pollitt (eds.), *The Oxford Handbook of Public Management*, Oxford University Press, Oxford, 2005, pp. 326-346. pp. 405-423.
- Prick, L., *Drammen, dreigen, draaien – Hoe het onderwijs twintig jaar lang vernieuwd werd*, Mets en Schilt, Amsterdam, 2006.
- Prince, Y., De kunst van beleidsevaluaties, in: Boekenoogen, M., P. van Hoesel, Y. Prince en C. Verheijen (red.), *Methoden van beleidsonderzoekers: creatief en oplossingsgericht*, Lemma, Den Haag, 2009, pp. 169-177.
- Prinsen, H.M. en R.M.M. Vossen, Naleving en handhaving van regelgeving; loont het de moeite?, in: *Justitiële Verkenningen*, jrg. 29, 2003, nr. 9, pp. 54-70.
- Pröpper, I.M.A.M. en D.-D. Reneman, De reconstructie van de beleidstheorie als argumentatie: Een toepassing op het verkeersveiligheidsbeleid, in: *Beleidswetenschap*, 1993, pp. 238-257.
- Pröpper, I.M.A.M. en M.S. de Vries, Presentatie en gebruik van evaluatie-onderzoek, in: Bressers, J.T.A. en A. Hoogerwerf (red.), *Beleidsevaluatie*, Samsom, Alphen a.d. Rijn, 1995, pp. 158-177.
- Pröpper, I.M.A.M., Beleidsevaluatie als argumentatie, in: Abma, T. en R.J. in 't Veld (red.), *Handboek beleidswetenschap*, Boom, Amsterdam, 2001, pp. 145-165.
- Pröpper, I.M.A.M., Beleidsevaluatie als argumentatie, in: *Beleidswetenschap*, jrg. 1, 1987, nr. 2, pp. 113-136.
- Pröpper, I.M.A.M., Het gebruik van beleidsevaluatie: terugkoppeling, in: Hoogerwerf, A. (red.), *Overheidsbeleid*, Samsom, Alphen a.d. Rijn, 1993, pp. 180-198.
- Pröpper, I.M.A.M., Succes en falen van sturing in beleidsnetwerken, in: *Beleidswetenschap*, jrg. 10, 1996, pp. 345-365.
- Putters, K. en T.E.D. van der Grinten, Facetbeleid en gezondheidseffectscreening, in: *Tijdschrift Sociale Gezondheidszorg*, 1998, 76, pp. 258-262.
- Quality Assurance Netherlands Universities (QANU), *Assessment of degree programmes, Public Administration*, Utrecht, September 2005.
- Quality Assurance Netherlands Universities (QANU), *Onderwijsvisitatie Bedrijfswetenschappen*, Utrecht, 2004.
- Quality Assurance Netherlands Universities (QANU), *QANU-KADER*, februari 2004, www.qanu.nl.
- Quality Assurance Netherlands Universities (QANU), *Onderwijsvisitatie Politicologie*, Utrecht, 2004.
- Raad voor het openbaar bestuur, *Presteren door leren – Benchmarken in het binnenlands bestuur*, Rob, Den Haag, 2002.
- Rein, M. & D.A.Schön, Frame-reflective policy discourse, in: *Beleidsanalyse*, 1986, nr. 4, pp. 4-18.
- Rein, M., Reframing Problematic Policies, in: Moran, M., M. Rein & R.E. Goodin (eds.), *The Oxford Handbook of Public Policy*, Oxford University Press, 2006, pp. 389-409.

Rein, M., Reframing Problematic Policies, in: Moran, M., M. Rein & R.E. Goodin (eds.), *The Oxford Handbook of Public Policy*, Oxford UP, Oxford, 2006, pp. 389-409.

Reitsma, A.B., *Kritische succesfactoren bij visitatie en kwaliteitsverbetering van P&A opleidingen*, Open Universiteit, Heerlen, 2003. (diss.).

Rhodes, R.A.W., Policy Network Analysis, in: Moran, M., M. Rein & R.E. Goodin (eds.), *The Oxford Handbook of Public Policy*, Oxford University Press, 2006, pp. 425-448.

Ringeling, A.B., De instrumenten van beleid, in: Lehning, P.B. en J.B.D. Simonis (red.), *Handboek beleidswetenschap*, Boom, Meppel, 1987, pp. 105-116.

Ringeling, A.B., *Het imago van de overheid*, Vuga, Den Haag, 1993.

Ritzer, G., *The McDonaldization of Society*, Sage, Thousand Oakes, 1996.

Rose, R. & Ph. L. Davies, *Inheritance in Public Policy- Change without Choice in Britain*, Yale University Press, New Haven, 1994.

Rossi, P.H. & H.E. Freeman, *Evaluation – A systematic approach*, Sage, Newbury Park, 1989.

Rossi, P.H., M.W. Lipsey & H.E. Freeman, *Evaluation – A systematic approach*, Sage, Londen, 2004 (7th edition).

Rynck, Ph. De, J. Voet en W. van Dooren, Het beoordelen van de prestaties van netwerken: aanzet tot een analysekader, in: *Bestuurswetenschappen*, jrg. 64, 2010, nr. 1, pp. 35-56.

Schaardenburgh-Verhoeve, K., Kenmerken van een goed ongevalsonderzoeksrapport: wat beïnvloedt de kwaliteit en hoe evalueer je die kwaliteit?, in: Mertens, F., K. van Schaardenburgh-Verhoeve en S. Sillem (red.), *Methodische aspecten van het onderzoek naar ongevallen*, Eburon, Delft, 2012, pp. 209-227.

Schedler, K., Public Management und Public Governance, in: Benz, A., S. Lütz, U. Schimank & G. Simonis (Hrsg.), *Handbuch Governance*, VS Verlag, Wiesbaden, 2007, pp. 253-269.

Scheele, J.P., P.A.M. Maassen & D.F. Westerheijden (eds.), *To be Continued Follow-UP of Quality Assurance in Higher Education*, Elsevier/De Tijdstroom, Maarssen, 1998.

Schillemans, Th. en M. Bovens, Publieke verantwoording 2.0: sober maar scherp, in: Bovens, M. en Th. Schillemans (red.), *Handboek publieke verantwoording*, Lemma, Den Haag, 2009, pp. 275-295.

Schillemans, Th., *Verantwoording in de schaduw van de macht - Horizontale verantwoording bij zelfstandige uitvoeringsorganisaties*, Lemma, Den Haag, 2007.

Schillemans, Th., Visitaties als vorm van verantwoording, in: Bovens, M. en Th. Schillemans (red.), *Handboek publieke verantwoording*, Lemma, Den Haag, 2009, pp. 161-181.

Schillemans, Th., Visitaties als vorm van verantwoording, in: Bovens, M. en Th. Schillemans (red.), *Handboek publieke verantwoording*, Boom Lemma, Den Haag, 2009, pp. 161-180.

Schoenmaker, M., *Bestuurlijk gedonder*, Wolff Publ., Nijmegen, 2011.

Schön, D.A. & M. Rein, *Frame Reflection – Towards the Resolution of Intractable Policy Controversies*, Basic Books, New York, 1994.

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.

Schuyt, C.J.M., *Tegendraadse werkingen – Sociologische opstellen over de onvoorziene gevolgen van verzorging en verzekering*, Amsterdam University Press, Amsterdam, 1995.

Schwarz, S. & D.F. Westerheijden (eds.), *Accreditation and Evaluation in the European Higher Education Area*, Kluwer Academic Publ., Dordrecht, 2004.

Scott, J., *Seeing like a State – How certain schemes to improve the human condition have failed*, Yale University Press, New Haven, 1996.

Scriven, M., Pros and cons about goal free evaluation, in: *Evaluation comment: The Journal of Educational Evaluation*, vol. 3, 1972, nr. 4, pp. 1-7.

Shulba, L.M. & J.B. Cousins, Evaluation use: theory, research and practice since 1986, in: *Evaluation Practice*, vol. 18, 1997, nr. 3, pp. 195-208.

Slob, A., Sturen op kwaliteit van evaluatieonderzoek, in: Knaap, P. van der, en A. Schilder (red.), *Resultaatgericht sturen en evalueren*, Sdu, Den Haag, 2004, pp.139-145.

Slob, M. en J. Staman, *Beleid en het bewijsbeest – Een verkenning van verwachtingen en praktijken rond evidence based policy*, Rathenau Instituut, Den Haag, 2012.

Slotboom, A. en C. Wiebrens, Opsluiten of sleutelen? Kosten en baten van detentie en resocialisatie, in: *Justitiële Verkenningen*, jrg. 29, 2003, nr. 9, pp.40-54.

Smid, G., *Goed werk vraagt om goed besturen*, Open Universiteit, Heerlen, 2012 (oratie).

Smith, H. e.a., *Benchmarking and Threshold Standards in Higher Education*, Kogan Page, London, 2000.

Smith, P., On the unintended consequences of publishing performance data in the public sector, in: *International Journal of Public Administration*, 1995, nr. 18, pp. 377-e.v.

Snellen, I.Th.M., *Boeiend en geboeid*, Samsom, Alphen, 1987.

Steen, M. van der, R. Peeters en M. van Twist, *De boom en het rizoom- Overheidssturing in een netwerksamenleving*, Wonen, wijken & Integratie/Ministerie van VROM, Den Haag, 2009.

Steenbeek, D., *Ex ante evaluatietoetsen bij de rijksoverheid- Een stageverslag*, Ministerie van Justitie, Den Haag, juli 1997.

Steenbeek, D., *Leren met de ex ante evaluatie-toets bij de rijksoverheid*, Heerlen/A'dam, 1998 (scriptie).

Steijn, B., *Carrièrejager of dienaar van de publieke zaak – Over ambtenaren en hun motivatie*, Erasmus universiteit, Rotterdam, 2006.

Steijn, B., Er leeft onvrede over managers in de publieke sector, in: Brink, G. van den, Th. Jansen en D. Pessers (red.), *Beroepszeer*, Boom, Amsterdam, 2005, pp. 132-143.

Steunenbergh, B., M. Beerkens, C. van den Berg e.a., 'Wij doen het zo'- *Onderzoek naar de grenseffecten van Europees beleid*, Universiteit Leiden, Den Haag, 4 aug. 2012.

Stokkom, B. van, en J. Terpstra, Probleemgericht werken in lokale veiligheidsnetwerken, in: Meurs, P.L., E.K. Schrijvers en G.H. de Vries (red.), *Leren van de praktijk – Gebruik van lokale kennis en ervaring voor beleid*, Amsterdam University Press, Amsterdam, 2006, pp. 91-119.

Suurmond, G. en B.C.J. van Velthoven, Brandveiligheid in de horeca: toepassing van kosten-batenanalyses op justitieel gebied, in: *Justitiële Verkenningen*, jrg. 29, 2003, nr. 9, pp. 8-30.

Swanborn, P.G., *Evalueren*, Boom, Amsterdam, 1999/2002.

Swanborn, P.G., *Methoden van sociaal-wetenschappelijk onderzoek*, Boom, Meppel, 1987.

Swieringa, J. en A.F.M. Wierdsma, *Op weg naar een lerende organisatie*, Wolters Noordhoff, Groningen, 1990.

Swierstra, T. en E. Tonkens (red.), *De beste de baas? Prestatie, respect en solidariteit in een meritocratie*, Amsterdam University Press, Amsterdam, 2008.

Taylor, J., Factors Influencing the Use of Performance Information for Decision Making in Australian State Agencies, in: *Public Administration*, vol.89, 2011, nr. 4, pp. 1316-1335.

Termeer, C.J.A.M., H. Mastik, R.G. Scalzo en R.J. in 't Veld, Spelsimulatie als methode voor de ex ante evaluatie van wetgeving, in: *Beleidwetenschap*, jrg. 9, 1995, nr. 3, pp. 225-241.

Termeer, C.J.A.M., *Vitale verschillen – Over publiek leiderschap en maatschappelijke innovatie*, Wageningen, 2006 (oratie).

Termeer, C.J.A.M., Vitaliseren van gestagneerde organiseerprocessen, in: *M&O*, maart/april 2006, nr. 2, pp. 26-40.

Terpstra, J., Netwerken en samenwerking bij de uitvoering van beleid, in: *Beleidswetenschap*, jrg. 15, 2001, nr. 2, pp. 141-169.

Thiel, S. van, en F.L. Leeuw, De prestatieparadox in de publieke sector, in: *Beleidswetenschap*, jrg. 17, 2003, nr. 2, pp. 123-144.

Tijdelijke adviescommissie Beoordeling Projectvoorstellen Studeerbaarheidsfonds (BPS), *3000x beter – Slotrapport Studeerbaarheidsfonds*, Ministerie van OC en W, Zoetermeer, 1999.

Tijssen, I., *Kwaliteit noodt tot meer gebruik- Bruikbaarheid van sociaal-wetenschappelijke onderzoeksresultaten volgens arts-bestuurders*, Quickprint, Nijmegen, 1988.

Tillema, S. & H.J. ter Bogt, *Dutch Local Government Audit Committees*, RUG, Groningen, 2009 (paper).

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.

Tonkens, E., *Mondige burgers, getemde professionals*, Van Gennip, Amsterdam, 2008.

Tops, P., *Veiligheid: iedereen een zorg? Visitatiecommissie veiligheidsbeleid in Tilburg*, Gemeente Tilburg, Tilburg, 2006.

Torenvlied, R., *Besluiten in uitvoering*, Groningen, 1996.

Twist, M.J.W. van, De beleidstheorie vanuit de wetenschapspraktijk: van 'hoe ver?' naar 'hoe verder?', in: *Beleidswetenschap*, jrg. 7, 1993, nr. 1, pp. 34-48.

Twist, M.J.W. van, R. Bagchus en F.O.M. Verhaak (red.), *Kerndepartementen op afstand?*, Eburon, Delft, 1996.

Universiteit Twente, Faculteit Bestuurskunde, *Zelfevaluatie Opleiding Bestuurskunde*, mei 1997.

Valk, Th. de, en M.S. de Vries, Criteria voor milieu-effectrapporten, in: *Beleidswetenschap*, 1994, nr. 3, pp. 274-304.

Vall, M. van de, en B. van Dijkum-De Jong, *Het rendement van sociaal beleidsonderzoek – Een onderzoek naar de voorwaarden voor beleidsrelevantie*, Samsom, Alphen a.d. Rijn, 1979.

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

Veld, R.J. in 't, De sturingsconceptie van paars, in: Becker, F., W. van Henekeler, M. Sie Dhian Ho en B. Tromp (red.), *Zeven jaar paars*, De Arbeiderspers, Amsterdam, 2001a, pp. 164-197.

Veld, R.J. in 't, *Een vlucht naar Isfahan?*, Vuga, Den Haag, 1984.

Veld, R.J. in 't, en P. van der Knaap (red.), *Dynamische bestuurskunde*, Phaedrus, Den Haag, 1994.

Veld, R.J. in 't, *Kennisdemocratie- Opkomend stormtij*, Academic Service/ Sdu, Den Haag, 2010.

Veld, R.J. in 't, Onderwijs: bevrijding uit de beleidsgevangenis, in: Bussemaker, J. en R. van der Ploeg (red.), *Leven na Paars?*, Prometheus, Amsterdam, 2001b, pp. 85-97.

Verhoeven, J. en K. de Wit, Hoger onderwijs en beleid in Europa, in: *Bestuurskunde*, jrg. 8, 1999, nr. 8, pp. 346-357.

Verlet, D. & C. Devos (red.), *Over beleidsevaluatie: van theorie naar praktijk en terug*, SVR Studie 2, Gazelle, Brussel, 2008.

Versteden, C.J.N., De cafébrand in Volendam en de kwaliteit van het openbaar bestuur, in: *Bestuurskunde*, jrg. 10, juli 2001, nr. 4, pp. 142-152.

Visitatiecommissie bestuurskracht Limburg, *Weging van de bestuurskracht van de provincie Limburg*, Maastricht, 1 dec. 2006.

Visitatiecommissie bestuurskracht van de provincie Zeeland, *Ruimte voor verbeelding*, B&A, Den Haag, februari 2011.

Visitatiecommissie Rechtspraak, *Rapport visitatie gerechten*, Raad voor de Rechtspraak, Den Haag, 2006 en 2010.

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).

Voogd, H., Multicriteria-evaluatie: praktijkervaringen gewogen, in: *Beleidsanalyse*, 1988, nr. 2, pp. 11-16.

Voogd, J.H., *Multicriteria Evaluation for Urban and Regional Planning*, Pion, Londen, 1983.

Vries, B. de, A.F.A. Korsten en W. Gooren, *Evaluatie van de totstandkoming van het Innovatieprogramma biotechnologie*, Den Haag, 1983.

Vries, J. de, 'Bildung is iets voor in het weekend'- Interview met Grahame Lock, in: *De Groene Amsterdammer*, 22 augustus 2008.

Vries, J. de, *Paars en de managementstaat - Het eerste Kabinet-Kok (1994-1998)*, Garant, Amersfoort, 2002.

Vries, M.S. de, *Calculeren met beleid*, Van Gorcum, Assen, 1994.

Vries, M.S. de, *De ontmythologisering van het openbaar bestuur*, Radboud Universiteit, Nijmegen, 2005.

Vries, M.S. de, Prioriteiten (s)tellen: multicriteria evaluatie in theorie en praktijk, in: *Beleidswetenschap*, 1992, nr. 2, pp. 126-154.

VSNU, *Assessment of Research Quality - Protocol 1998*, VSNU, Utrecht, 1998.

VSNU, *Assessment of Research Quality- Public Administration - Political Science - Communication Science*, Utrecht, 2002.

VSNU, *Assessment of Research Quality- Rechtsgeleerdheid*, VSNU, Utrecht, 2002.

VSNU, *Onderwijsvisitatie Economie*, Utrecht, juli 2003.

VSNU, *Quality Assessment of Research - Onderzoekbeoordeling - Protocol 1994*, VSNU, Utrecht, 1994.

VSNU, *Quality Assessment of Research - Rechtsgeleerdheid*, VSNU, Utrecht, 1996.

VSNU, *Quality Measurement Made tot Measure: Protocol for External Assessment of Educational Programs 2000-2005*, Utrecht, 1999.

VSNU-visitatiecommissie Bestuurs-, Beleids- en Politieke Wetenschappen, *Onderwijsvisitatie - Bestuurs-, beleids- en politieke wetenschappen*, VSNU, Utrecht, 1992.

VSNU-visitatiecommissie Bestuurskunde, *Onderwijsvisitatie - De kwaliteit van het universitair onderwijs - Bestuurskunde*, VSNU, Utrecht, 1998.

Vught, F.A. van, en D.F. Westerheijden, Toward a general theory of quality assessment in higher education, in: *Higher Education*, jrg. 28, 1994, pp. 355-371.

Waal, A.A. de, en L.A.F.M. Kerklaan, *De resultaatgerichte overheid*, Sdu, Den Haag, 2004.

Walle, S. van de, & A. Roberts, Publishing Performance Information: An Illusion of control?, in: Dooren, W. van, en S. van de Walle (eds.), *Performance Information in the Public Sector- How it is used*, Palgrave MacMillan, Londen, 2008, pp. 211-227.

Walle, S. van de, *Perceptions of Administrative Performance: The Key to Trust in Government?*, KULeuven, Leuven, 2004.

- Walle, S. van de, Zullen beter werkende overheidsdiensten bijdragen tot vertrouwensherstel?, in: Korsten, A. en P. de Goede (red.), *Bouwen aan vertrouwen in het openbaar bestuur – Diagnoses en remedies*, Elsevier, Den Haag, 2006, pp. 125-139.
- Wansink, H., Nogmaals: regeren is achteruit zien, in: *De Volkskrant*, 20 juli 2002.
- Webber, D.J., The Distribution and Use of Policy Knowledge in the Policy Process, in: *Knowledge and Policy*, vol. 4, 1991, nr. 4, pp. 6-35.
- Webbink, H.D., Causale effecten van beleid: over gecontroleerde en natuurlijke experimenten, Weibel, A., K. Rost & M. Osterloh, Pay for performance in the public sector- Benefits and (hidden) costs, in: *Journal of Public Administration Research and Theory*, vol. 20, 2009, nr. 2, pp. 387-412.
- Weiss, C.H. (ed.), *Using Social Science Research in Public Policy-Making*, Lexington Books, Lexington, 1977.
- Weiss, C.H. (with M. Bucuvalas), *Social Science Research and Decision-making*, Columbia University Press, New York, 1980.
- Weiss, C.H., E. Murphy-Graham & S. Birkeland, An Alternative Route to Policy Influence, in: *American Journal of Evaluation*, vol. 26, maart 2005, nr. 1, pp. 12-30.
- Weiss, C.H., Evaluation Research in a Political Context, in: Struening, E. & M. Guttentag (eds.), *Handbook of Evaluation Research*, Sage, Beverly Hills, 1976, pp. 13-27.
- Weiss, C.H., *Evaluation research*, Prentice Hall, Englewood cliffs, 1972.
- Weiss, C.H., *Evaluierungsforschung – Methoden zur Einschätzung von sozialen Reformprogramme*, Westdeutscher Verlag, Opladen, 1974.
- Weiss, C.H., Knowledge Creep and Decision Accretion, in: *Knowledge: Creation, Diffusion, Utilization*, 1980, pp. 381-404.
- Weiss, C.H., Research for Policy's Sake: the Enlightenment Function of Social Science, in: *Policy Analysis*, vol. 3, 1977, nr. 4, pp. 531-545.
- Weiss, C.H., The Many Meanings of Research Utilization, in: *Public Administration Review*, sept/oct. 1979, pp. 426-431.
- Werle, R., Pfadabhängigkeit, in: Benz, A., S. Lütz, U. Schimank & G. Simonis (Hrsg.), *Handbuch Governance*, VS Verlag, Wiesbaden, 2007, pp. 119-132.
- Wesemann, P., Kosten en baten van alcoholcontroles in het verkeer, in: *Justitiële Verkenningen*, jrg. 29, 2003, nr. 9, pp. 30-40.
- Westerheijden, D.F. e.a., *Changing contexts of quality assessment, Recent trends in West European Higher Education*, CHEPS/Lemma, Utrecht, 1994.
- Westerheijden, D.F., A solid base for decisions: Use of the VSNU research evaluations in Dutch universities, in: *Higher Education*, vol. 33, 1997, nr. 4, pp. 397-413.
- Westerheijden, D.F., B. Stensaker en M.J. Rosa (red.), *Quality Assurance in Higher Education, Trends in Regulation, Translation and Transformation*, Springer, Dordrecht, 2007.
- Westerheijden, D.F., Quality assessment in Dutch higher education: Balancing improvement and accountability, in: *European Journal for Education Law and Policy*, vol. 1, 1997, pp. 81-90.
- Weusthof, P.J.M., *De interne kwaliteitszorg in het wetenschappelijk onderwijs*, Lemma, Utrecht, 1994.
- Wholey, J. e.a. (eds.), *Handbook of practical program evaluation*, Jossey-Bass, San Francisco, 1994.
- Wildavsky, A., The Self-Evaluating Organization, in: Kearny, R. & E.V. Berman (ed.), *Public Sector Performance – Management, Motivation and Measurement*, ASPA, Westview Press, Boulder, 1999, pp. 273-293.
- Winter, H.B. en M. Herweijer, Evaluatie van beleid, in: *Strategie en beleid in de publieke sector*, Samsom, Alphen, dec. 1995, F2120, pp. 1-26.
- Yin, R.K. & M.K. Gwaltney, Knowledge Utilization as a Networking Process, in: *Knowledge: Creation, Diffusion, Utilization*, vol. 2, 1981, nr. 4, pp. 555-580.
- Zwan, A. van der, en M. Noordegraaf, Onderzoek van een complexe casus – Een gemeentevervoerbedrijf als voorbeeld, in: *Bestuurskunde*, jrg. 7, 1998, nr. 2, pp. 62-74.

De auteur

Prof.dr. Arno F.A. Korsten is honorair hoogleraar aan de Universiteit Maastricht en emeritus hoogleraar bedrijfs- en bestuurswetenschappen, in het bijzonder bestuurskunde aan de Open Universiteit. Was vanaf medio 1986 – medio 2010 gewoon hoogleraar aan de OU en vanaf medio 1991 bijzonder hoogleraar aan de UM.

Was onder meer lid van de visitatiecommissie die de provincie Limburg visiteerde en de commissie die de provincie Zeeland doorlichtte. Voorzitter van de visitatiecommissie ter beoordeling van hbo-opleidingen bestuurskunde/overheidsmanagement.

Was mede-oprichter van het tijdschrift Bestuurskunde. Eerder ook informateur dan wel formateur in Heerlen, Den Bosch en Maasdriel. Voormalig lid van de Raad voor het openbaar bestuur. Erelid van de Vereniging voor Bestuurskunde. Officier in de Orde van Oranje-Nassau wegens inzet en verdiensten op het vlak van de bestuurskunde en het openbaar bestuur.