

## Van casus-specifieke beleidsevaluatie naar systematische opbouw van kennis en ervaring

**André Knottnerus**

### Aanbevolen citeerwijze bij dit artikel

André Knottnerus, 'Van casus-specifieke beleidsevaluatie naar systematische opbouw van kennis en ervaring', *Beleidsonderzoek Online* mei 2016, DOI: 10.5553/BO/221335502016000005001

### Introductie

Als een beleidsmaker zegt vooruit te willen kijken, heeft hij groot gelijk. Maar als hij erbij zegt dat we het verleden zo snel mogelijk achter ons moeten laten, moet een rood lampje gaan branden. Hij geeft daarmee immers te kennen weinig belangstelling te hebben voor hoe de zaken gelopen zijn en niet gemotiveerd te zijn om van het verleden te leren. Terugblikken, ofschoon niet altijd gemakkelijk, is echter onontbeerlijk om beter vooruit te kunnen.

Deze bijdrage gaat over het evalueren van overheidsbeleid. Om te beginnen wordt benadrukt dat evalueren om te leren zowel vanzelfsprekend als lastig is. Vervolgens wordt ingezoomd op leerpunten en spanningsvelden die uit de evaluatiepraktijk naar voren komen. Aansluitend wordt ingegaan op mogelijkheden om die praktijk te versterken. De kernboodschap is dat beleidsevaluaties niet alleen van betekenis zijn voor het specifieke onderwerp waarop zij primair gericht zijn, maar ook dienen bij te dragen aan een bredere, systematische kennisopbouw ten behoeve van voortgaande beleidsverbetering.<sup>1</sup>

### Evalueren om te verantwoorden en te leren

Evaluatie van beleid kost energie, tijd en geld, maar is onmisbaar voor adequate democratische controle en verantwoording, en om te kunnen leren van meer en minder geslaagd beleid.<sup>2</sup> In diverse wettelijke regelingen is dan ook opgenomen dat met regelmaat beleidsevaluaties moeten worden uitgevoerd, en de Regeling periodiek evaluatieonderzoek stelt eisen ten aanzien van onderzoek naar doeltreffendheid en doelmatigheid, beleidsdoorlichtingen, en subsidie-evaluaties.<sup>3</sup> Er vinden ook veel evaluaties plaats die niet wettelijk verplicht zijn. Deze worden vaak uitgevoerd op initiatief van departementale beleidsdirecties of de Tweede Kamer. De Algemene Rekenkamer, wetenschappelijke instituten en de Nationale Ombudsman kijken 'van buiten' naar het handelen van departementen en uitvoeringsdiensten. En sinds 2000 kennen we Verantwoordingsdag, de derde woensdag in mei waarop het kabinet in de Tweede Kamer verantwoording aflegt over het beleid en de financiën van het afgelopen jaar, aan de hand van evaluatieonderzoek van de Algemene Rekenkamer.

Uitgangspunt bij al deze evaluatie-activiteiten is dat de overheid dient te streven naar zo goed mogelijk beleid en dat zij zich daarover moet verantwoorden. Lukt het te bereiken wat voor ogen stond en is toegezegd? 'Doet' het beleid wat was beoogd, tegen een aanvaardbare prijs? En zo niet, is het beleid met datgene wat het wél oplevert toch te rechtvaardigen? Stuurt de subsidieregeling in de beoogde richting? Beschermt het recht voldoende en in gelijke mate, en tot welke mogelijke averechtse effecten leidt een wet?

Voor het beantwoorden van zulke vragen is goed voorbereid evaluatieonderzoek nodig – ondersteund door gedegen informatieverzameling, zodat de uitwerking van beleid kan worden gevolgd en beoordeeld.<sup>4</sup> Op basis van de resultaten daarvan kan het beleid verbeterd, of anders afgebouwd of vervangen worden. Evalueren is daarom onlosmakelijk verbonden met goede beleidsvoering, net als het ontwerpen, het implementeren, en het uitvoeren ervan.

Daarbij gaat het ook om vertrouwen en geloofwaardigheid. Wie laat zien lessen te trekken uit het verleden en zijn best te doen om gemaakte fouten voortaan te voorkomen, kan meer vertrouwen van derden verwachten.<sup>5</sup> Dat is cruciaal waar verantwoording niet zozeer 'afrekenen' maar vooral verbetering of herkansing als hoofddoel heeft.

### Leren is lastig

Leren van evaluaties is niet gemakkelijk. Veel parlementaire enquêtes en onderzoeken gaan juist ook over de vraag waarom onvoldoende geleerd is van eerdere waarschuwingen en signalen, aangereikt in reguliere evaluaties en toezichtrapporten over al langer bestaande vraagstukken of problemen. Hierbij speelt een rol dat de kennis die uit evaluaties naar voren komt, slechts één van de factoren is die het beleid beïnvloeden.<sup>6</sup> Beschikbare middelen leggen beperkingen op, belangengroepen oefenen druk uit, media dragen oordelen aan, en niet zelden

is er politieke pressie. Soms neemt men daardoor ook te weinig ruimte en tijd om evaluaties voluit te kunnen benutten. Voorts kan politieke opportuniteit een rol spelen, bijvoorbeeld omdat beleidswijziging voldoende draagvlak vereist en dat is moeilijk bereikbaar als sprake is van indringende politieke controverses. Deze ontwikkelingen kunnen leiden tot wat ook wel kennisparadox<sup>7</sup> wordt genoemd: er is relevante kennis beschikbaar, maar deze wordt niet of maar in beperkte mate benut.

Niet alleen het leren van uitgevoerde evaluaties kan lastig zijn, maar ook het vooraf opzetten ervan. Zo kan het tegelijk met het inzetten van beleid optuigen van een gedegen evaluatie worden ervaren als teken van politiek-bestuurlijke onzekerheid en zwakte, of als een politiek risico omdat men dan op gemaakte keuzes kan worden afgerekend. Aldus kan de beslissing om bepaald beleid te evalueren, en hoe en door wie dat gaat gebeuren, gekleurd raken door politiek-bestuurlijke overwegingen. Het gaat dan niet alleen meer om het professioneel en gestructureerd willen leren van hoe het beleid gaat uitwerken, maar ook om wat men 'geanticiperde damage control' zou kunnen noemen.

Dat vele factoren het leren via evaluatie kunnen beïnvloeden en zelfs belemmeren, zien Furubo et al.<sup>8</sup> als een uitdaging: '(...) the fact that mechanisms exist that make learning difficult can at the same time be seen as something that makes it even more important to discuss how evaluation can contribute to learning.' Op het juist te midden van barrières zoeken naar mogelijkheden om leren te bevorderen spitst het vervolg van dit artikel zich toe. Het gaat dan om kwesties die in de evaluatiepraktijk steeds opnieuw aandacht vragen, maar ook om het spanningsveld tussen politisering en verwetenschappelijking. En natuurlijk om de condities die het opbouwen, vasthouden en benutten van kennis uit evaluaties bevorderen.

## Kwesties die steeds opnieuw aandacht vragen

### *Effectiviteit als veelomvattend begrip*

Evaluaties richten zich vaak op doeltreffendheid (effectiviteit) en doelmatigheid (de verhouding tussen de effecten en de kosten) van beleid. Dat levert belangrijke input voor nieuwe of bij te stellen beleidsinterventies, bijvoorbeeld in de gezondheidszorg en het energiedomein. Het is in de praktijk echter niet gemakkelijk om de uiteindelijke effectiviteit van wetgeving en beleid ten aanzien van de beoogde uitkomsten eenduidig en overtuigend vast te stellen. Om dat te kunnen doen zijn, zoals we verderop ook nog zullen zien, specifieke evaluatiedesigns nodig – vaak met een lange looptijd – die niet eenvoudig te realiseren zijn. Daarom is het belangrijk vooraf aandacht te besteden aan de potentiële werkzaamheid van overwogen beleidsinterventies. Onderzoek naar mechanismen die bijdragen aan het bereiken van goede beleidsresultaten<sup>9</sup> en naar de vraag 'waarom iets zou kunnen werken, voor wie en onder welke omstandigheden', helpt daarbij.<sup>10</sup> Beter inzicht in relevante mechanismen, bijvoorbeeld verkregen via experimenten en pilots, maakt het immers mogelijk kritischer te zijn wat betreft het al dan niet invoeren van beleid.<sup>11, 12</sup> Bovendien zijn zulke mechanismen vaak breder van toepassing dan alleen op het specifiek bestudeerde beleid, en inzicht erin kan dan ook naar ander beleid doorwerken. Daarnaast spelen ook factoren als besluitvormingscontext, timing, onderzoekskwaliteit, en het betrekken van stakeholders een belangrijke rol.<sup>13</sup>

In het algemeen geldt dat beleidsmakers via ex ante evaluatie worden geprikkeld om zich vooraf rekenschap te geven van de wetenschappelijke basis van het overwogen beleid.<sup>14</sup> Daarbij is het van belang de beleidstheorie achter het beoogde beleid zichtbaar te maken en de daarin centrale veronderstellingen – ook met betrekking tot gedragsmatige factoren – te onderzoeken op empirische validiteit, waarbij wordt nagegaan wat de kansen zijn dat de beleidsdoelen ook gerealiseerd worden. Vooral bij grotere dossiers kunnen integrale impact-assessments en maatschappelijke kosten-batenanalyses, zinvol zijn, waarbij ook vergelijkingen kunnen worden gemaakt met beleidsalternatieven en de nuloptie.<sup>15, 16</sup> Ex ante evaluatie dient net al ex post evaluatie de vraag te beantwoorden in hoeverre essentiële publieke belangen daadwerkelijk gediend zijn met het (beoogde) beleid. Dit is ook niet voor niets de eerste kernvraag die de overheid zich volgens de WRR steeds moet stellen bij de voorbereiding en evaluatie van het risico- en veiligheidsbeleid.<sup>17</sup>

Ex ante evaluatie komt het verantwoord gebruik van middelen en beleidsinstrumenten ten goede, maar er is op dit punt werk aan de winkel: het aantal ex ante evaluaties in ons land is beperkt, ook internationaal bezien.<sup>13, 18</sup> De OECD schreef in een studie over ons land: 'A significant weakness is the failure so far to implement an effective policy for the ex ante impact assessment of new regulations.'<sup>19</sup>

Voor beoordelaars van effectiviteit en doelmatigheid is het aantrekkelijk zich vooral te richten op reeds beschikbare of relatief gemakkelijk meetbare (indicatoren van) uitkomsten van beleid. Hier kan echter een valkuil schuilgaan als niet of nauwelijks meetbare waarden die voor beleidsmakers en doelgroepen evenzeer van belang zijn, onderbelicht blijven. Denk aan subjectief welzijn, rechtvaardigheid, en maatschappelijke aanvaardbaarheid. Bovendien bestaat het risico dat het gedrag van het te evalueren veld zich vooral op goed meetbare indicatoren gaat richten. Dit wordt ook wel aangeduid als 'pervers effect' van een dergelijke indicatorbenadering.<sup>20, 21</sup>

De associatie van beleidsevaluaties met een focus op wat gemakkelijk meetbaar is ('meten is weten'), in lijn met regelmatig gehoorde kritiek op New Public Management, verklaart mogelijk dat in ambtelijke kring soms een zekere scepsis bestaat ten aanzien van evaluaties. Er wordt in dit verband ook wel gesproken over 'management by measurement'.<sup>22, 23</sup>

Het is al met al belangrijk om effectiviteit en doelmatigheid ruim op te vatten, en in de ontwerpfase van het

beleid en het evaluatieproces ook te kijken naar mechanismen, te verwachten mate van doelbereiking, en belangrijke maatschappelijke waarden die aandacht verdienen.

### *Goede timing*

Een 'lerende overheid'<sup>24</sup> moet zorgen voor tijdige feedback, zodat zonder onnodige vertraging kan worden geleerd. Goede doorwerking van evaluaties in de beleidspraktijk vereist dan ook adequate timing.

Adequate timing vraagt om een geschikt 'time window', dat per onderwerp kan verschillen. Evaluatie moet niet te lang na de invoering van beleid worden verricht omdat de beoogde specifieke bijdrage ervan te midden van bijkomende invloeden dan lastig waarneembaar kan worden. Maar evaluatie kan ook te vroeg komen, als er nog sprake is van 'kinderziekten' of als er redelijkerwijs nog geen effecten mogelijk of meetbaar zijn.<sup>25</sup> Dat levert niet alleen een niet-adequate evaluatie op, het kan ook een methodologisch niet gerechtvaardigde negatieve invloed hebben op de verdere beleidsuitvoering. Berichten over tegenvallende resultaten ('niemand voert de wet uit' of 'er is geen verbetering waar te nemen') kunnen de motivatie om beleid uit te voeren immers ondermijnen. Ryan spreekt zelfs over 'death by evaluation'.<sup>26</sup>

Goede planning is ook belangrijk als een geëvalueerde wet opgaat in een omvangrijk wijzigingsprogramma. Resultaten van een evaluatie van een 'oude wet' worden dan al gauw niet meer relevant geacht, maar daarmee kunnen generieke leerpunten voor de toekomst verloren gaan. In de transitie van oude naar nieuwe regelgeving kan evaluatie van eerder beleid vaak nog meerwaarde toevoegen aan de invulling van een nieuw wettelijk kader, zoals de Omgevingswet.<sup>27</sup>

Gedegen evaluatie, gericht op wetenschappelijk betrouwbare en valide conclusies, vergt ook de nodige tijd. Onredelijke tijdsdruk van de kant van opdrachtgevers en voor aanvaardbare kwaliteit te korte doorlooptijden moeten dan ook niet geaccepteerd worden.<sup>28</sup> Dit punt vraagt bijzondere aandacht als de evaluator afhankelijk is van derde-geldstroomfinanciering en de opdrachtgever haast wil maken.

Kortom, goede timing is bepalend voor het zinvol zijn van evaluatie en voor de kwaliteit en doorwerking ervan. Goede beoordeling van de verhouding tussen behoeften en mogelijkheden, gerichte planning en adequate afstemming tussen opdrachtverstrekker en evaluator zijn daarom cruciaal.

### *Draagvlak en continuïteit*

De evaluatiepraktijk is kwetsbaar als deze afhangt van de inzet van slechts enkele toegewijde personen en als evaluatieboodschappen vaak herhaald moeten worden voordat ze doordringen.<sup>29</sup> Die kwetsbaarheid wordt in de hand gewerkt als het uitvoeren of begeleiden van evaluaties minder status heeft dan beleidsontwikkeling.

Kwetsbaarheid van het evaluatieproces kan worden beperkt door beleidsmedewerkers in een vroeg stadium bij evaluaties te betrekken en regelmatig te laten deelnemen aan kritische reflectie met de onderzoekers die de evaluatie uitvoeren. Dat onderstreept het belang van evaluatie, komt het draagvlak ervoor ten goede, en vergroot de kans dat de bevindingen in de beleidsvorming worden meegenomen. Deze betrokkenheid moet uiteraard zo georganiseerd worden dat de onafhankelijkheid van de evaluatie niet in het geding komt.

Ook is van groot belang dat de politiek-bestuurlijke leiding er geen twijfel over laat bestaan dat gedegen evaluatie essentieel is voor modern, professioneel bestuur en een adequaat bemenste organisatie met continuïteit vereist. Het interdepartementaal verbinden van schaarse evaluatie-expertise biedt daarbij steun.<sup>3</sup>

### *Onafhankelijkheid en toegevoegde waarde*

Wil evaluatie toegevoegde waarde hebben ten opzichte van wat de bij beleidsontwikkeling en -uitvoering betrokkenen al overzien en vinden, dan dient sprake te zijn van voldoende onafhankelijkheid en afstand. Dat is ook nodig om tot een onpartijdige beoordeling te komen die los staat van deelbelangen en geloofwaardig is. Onafhankelijkheid van de evaluator wordt dan ook algemeen belangrijk gevonden.<sup>30</sup>

Dat wil niet zeggen dat de evaluator altijd daadwerkelijk onafhankelijk is of als zodanig kan opereren. Zo rapporteerde de KNAW dat 'de meest concrete voorbeelden van ongewenste druk op de onderzoeker worden gemeld bij opdrachten door de overheid'.<sup>31</sup> Het voorstel van de KNAW dat onderzoekers en opdrachtgevers zich committeren aan de door haar opgestelde 'Verklaring van wetenschappelijke onafhankelijkheid' is daarom ook voor beleidsevaluatie relevant.<sup>32</sup> Daarbij maakt het niet uit of de onderzoekers in dienst zijn van een private, semipublieke of overheidsorganisatie.

Scherp blijven op onafhankelijkheid van evaluaties is dus geen overbodige luxe, maar het streven naar onafhankelijkheid moet niet leiden tot een te grote afstand tussen de geëvalueerden en de evaluatoren. De laatstgenoemden moeten immers voldoende inzicht hebben in de kwesties die spelen, zodat eerstgenoemden zich kunnen herkennen in het proces en de uitkomsten. Dat is ook het beste startpunt voor succesvolle implementatie van aanbevelingen.

Het niet onnodig vergroten van afstand kan ook worden bereikt via een trapsgewijze benadering. Korsten spreekt in dit verband van een 'evaluatieladder'.<sup>29</sup> Hij onderscheidt daarbij interne evaluatie en externe evaluatie, waarbij tot de laatste kan worden besloten als de eerste niet voldoet. Een externe evaluatie kan worden

voorbereid door een gedegen interne evaluatie voor te leggen. Volgens Wallage, voorzitter van de Raad voor het openbaar bestuur (mondelinge mededeling, 2012), behoort het tot de primaire verantwoordelijkheid van bestuurders en beleidsmakers om het eigen werk eerst zelf op effecten en uitwerking te beoordelen (de slager dient als eerste te willen weten of zijn vlees van goede kwaliteit is). Als over die beoordeling intern of extern twijfel bestaat of als sprake is van bijzondere risico's, is onafhankelijke externe evaluatie aangewezen. Via periodieke algemene reviews kan voorts extern getoetst worden of de overheid zijn evaluatieprocessen op orde heeft.

De uitdaging voor evaluatoren is een goede balans te vinden tussen een voldoende mate van onafhankelijkheid, als voorwaarde voor toegevoegde waarde en onpartijdigheid, en voldoende nabijheid om beleidsrelevant te kunnen zijn. Daarbij geldt dat beleidsmakers primair verantwoordelijk zijn voor de effecten van wat zij in gang zetten. Zij kunnen dus ook worden aangesproken op de wijze waarop zij hun verantwoordelijkheid voor adequate evaluatie daarvan invullen.

## Een context van politisering én verwetenschappelijking vraagt om houvast

De zojuist besproken punten onderstrepen dat het realiseren van degelijke en zinvolle beleidsevaluaties geen gemakkelijke opgave is. Dat is te meer het geval gezien de politiek-bestuurlijke context waarbinnen beleidsevaluatie plaatsvindt.

Het streven naar een compacte overheid zet het behartigen van de kennisfunctie op ministeries behoorlijk onder druk. Bezuiniging op kennisdirecties en de roulatie van ambtenaren verminderen zowel de doorlopende dossierkennis bij individuele functionarissen als het 'collectieve geheugen' binnen departementen. Daarmee neemt de behoefte aan het ad hoc invliegen van externe kennis en expertise toe, maar die zijn minder verbonden met beleidservaring. Parallel hieraan signaleren 't Hart<sup>33</sup> en Bekker<sup>34</sup> de trend dat minder gehecht wordt aan inhoudelijke zaakkennis bij ambtenaren, terwijl politieke sensitiviteit en strategische vaardigheden juist vaker van hen worden verwacht.

Het is dan niet verwonderlijk dat ambtenaren een politisering van de ambtelijke cultuur ervaren, gekenmerkt door fenomenen als het uit de wind houden van de minister,<sup>35</sup> het liever vooruit dan achteruit kijken, en gerichtheid op incidenten en de korte termijn.<sup>33</sup> Politisering van de ambtelijke cultuur – vaak gepaard aan een wispelturige politieke agenda, ingegeven door politieke scoringskansen en verdedigingsmechanismen – kan de speelruimte voor evaluatoren beperken en de aandacht voor hun bevindingen verminderen. Ook neemt dan de kans toe dat een evaluatie die vandaag verschijnt, morgen al oud nieuws is.

In dit verband mag een principiële verschil (en fundamentele complementariteit) tussen de politiek en het openbaar bestuur niet uit het oog worden verloren: terwijl de politiek haar bestaansrecht ontleent aan georganiseerde partijdigheid, dient het openbaar bestuur juist onpartijdig te opereren.<sup>36</sup> Dat bestuur draagt daarom een bijzondere verantwoordelijkheid voor regelmatig evalueren op grond van objectieve, valide informatie. Internationaal en nationaal zie we dan ook toenemende aandacht van overheden voor 'verwetenschappelijking van beleid' met een groeiende focus op evidence.<sup>37</sup> Ook de Nederlandse overheid onderkent de betekenis van wetenschappelijke kennis voor beleidsvorming en -ontwikkeling. Dat werd onderstreept door adviesverzoeken over gedragswetenschap en beleid en de kabinetsreactie<sup>38</sup> op rapporten van de Rli,<sup>39</sup> de RMO<sup>40</sup> en de WRR<sup>13</sup> daarover. Voortgaande ontwikkeling van het instrument beleidsevaluatie als belangrijke kennisbron past daarbij.

Ook hier is dus een paradox waarneembaar: enerzijds zijn er tekenen van politisering van de ambtelijke cultuur en anderzijds is er aandacht voor verwetenschappelijking van de beleidsontwikkeling. Deze gelijktijdige maar deels tegengestelde trends zijn op verschillende niveaus verschillend voelbaar. Op departementen kan politisering van het beleid in de hand werken dat men liever niet terugkijkt en bij voorkeur betrokken is bij het ontwikkelen van nieuw beleid. Maar op aanwijzing van diezelfde overheid moet het handelen van het veld, in casu semipublieke organisaties en uitvoeringsdiensten, nauwgezet worden geëvalueerd. Dat mag en moet ook worden verlangd, maar veldpartijen kunnen de gerechtvaardigde vraag stellen of de overheid evenveel belangstelling heeft voor evaluatie van haar eigen beleid als voor evaluatie van de prestaties van andere organisaties met publieke taken.

Het spanningsveld tussen politisering en verwetenschappelijking kan zich ook voordoen in de verhouding tussen ministeries en politiek. Zo hecht het ministerie van OCW eraan om evidence-informed beleid te voeren. Dat leidde tot een voorstel van de staatssecretaris en de Kinderombudsman om scholen verplicht te laten kiezen voor een erkend bewezen effectief anti-pestprotocol.<sup>41</sup> Een belangrijk deel van de politiek en het veld bleek dit echter op te vatten als een beperking van de vrijheid van onderwijsinstellingen om een eigen aanpak te ontwikkelen.

De spanning tussen politisering en verwetenschappelijking kan worden verminderd door een evidence-informed benadering op basis waarvan het verantwoordelijke politiek bestuur goed onderbouwde keuzes kan maken. Het gedegen evalueren van het gekozen beleid, met een ruime definitie van het begrip effectiviteit en rekening houdend met zowel goed als minder goed meetbare uitkomsten, biedt daarbij houvast.

## Een zo hoog mogelijk niveau van bewijskracht

Voor het ontwikkelen van een evaluatieaanpak om beleid valide te beoordelen en van evidence-informed aanbevelingen te voorzien, is het thema bewijskracht van groot belang. Om met voldoende zekerheid iets te

kunnen zeggen over doelbereiking en over werkzame mechanismen, is het wenselijk dat evaluaties zo hoog mogelijk scoren op de Maryland Scientific Methods Scale (MSMS). Bij voorkeur gaat het ten minste om het quasi-experimentele design (zie ook box 1).<sup>42, 43, 44</sup> Nog hoger scoort de randomized controlled trial (RCT), waarbij te evalueren nieuw beleid via random allocatie wordt vergeleken met een andere – veelal de al bestaande – aanpak. In beide gevallen zijn naast een interventie- ook een controleconditie en een voor- en een nameting nodig. Indien er ten aanzien van een bepaalde toepassing meerdere RCT's zijn gedaan, kan daarenboven een systematische review worden uitgevoerd waarbij alle studies op kwaliteit worden beoordeeld en vervolgens worden samengevat.<sup>45</sup> Ook niet-gerandomiseerd onderzoek kan systematisch worden gereviewd.

Box 1: Vormen van effectiviteitsonderzoek\*

- Als hoogst haalbare methode van bewijsvoering voor effectiviteit geldt experimenteel onderzoek in de vorm van de randomised controlled trial (RCT), waarin de effecten van het al dan niet toepassen van een bepaalde interventie worden vergeleken. De RCT is in beleidsonderzoek niet gemakkelijk toe te passen, en is vooral bruikbaar is voor de evaluatie van relatief eenvoudige, redelijk te blinderen en op de korte termijn gerichte interventies. Bijvoorbeeld als in een goed gedefinieerde doelgroep een scherp omschreven gedragsverandering wordt onderzocht. Onder bepaalde voorwaarden kan het N=1 design interessant zijn, met name als het gaat om kleine aantallen.\*\* De RCT is niet of nauwelijks geschikt om te beoordelen of complexe, niet te blinderen beleidsinterventies de langetermijntoekomst verbeteren. Daarvoor komen dikwijls andere, meer observationele en kwalitatieve onderzoeksvormen in aanmerking zoals:
- Quasi-experimentele studies (met een niet-gerandomiseerde controleconditie) of observationele follow-up studies. Dergelijk onderzoek is bijvoorbeeld geschikt voor het vergelijken van regio's respectievelijk van trends in beschikbare gegevensbestanden. Met voor-na-vergelijkingen kan meer of minder aannemelijk worden gemaakt dat de gevonden verschillen het gevolg zijn van beleidsinterventies. Zulk onderzoek kent veel haken en ogen, maar kan bijdragen aan het inschatten van de effecten van beleidsinterventies.
- Cross-sectionele studies waarmee correlaties tussen uitgerold beleid en uitkomstmaten kunnen worden beoordeeld. De richting van het verband is hiermee echter lastig te beoordelen.
- Case-series, bijvoorbeeld bestaand uit systematische analyse van vergelijkbare case-ervaringen in eenzelfde beleidsdomein of in vergelijkbare domeinen, kunnen bijdragen aan het cumuleren van ervaringen en het inventariseren van leerpunten.
- Diepgaande individuele casestudies zijn vooral geschikt voor analyse van min of meer unieke situaties, of van situaties die snelle evaluatie vereisen, als concrete leerervaring.
- Kwalitatief onderzoek zoals interviews met actoren, zodat inzicht ontstaat in uitwerking en beleving van beleid.

Systematische reviews (SRs) zijn in principe geschikt voor de kwantitatieve en kwalitatieve evaluatie van alle bovengenoemde onderzoeksvormen, mits voldoende reproduceerbaar gerapporteerd.

\* Zie ook: Wetenschappelijke Raad voor het Regeringsbeleid. (2013). *Toezien op publieke belangen*.

\*\* Witvliet, M., & Timmermans, M. (2015). Het N=1 design toegepast in een effectonderzoek naar de justitiële gedragsinterventie. *Beleidsonderzoek Online*. DOI:10.5553/BO/221335502015000021001.

Een vergelijkende evaluatie vereist methodologische diepgang en kan zich daarom niet op alle mogelijke contrasten tegelijk richten. Bovendien moet ook de evaluatie zelf doelmatig zijn en de beschikbare middelen inzetten op de vragen waar het echt om gaat.<sup>36</sup> Daarom moeten keuzes worden gemaakt, en moet de vergelijkende evaluatie worden gericht op de essentiële kwesties waarover vooraf controversen dan wel onzekerheid bestaat en die niet door middel van een ex ante beoordeling valide op te lossen zijn.<sup>46</sup> Daarmee wordt voorkomen dat men de te oogsten empirie onbenut laat en na verloop van jaren weer met dezelfde onbeantwoorde vragen en onzekerheden geconfronteerd wordt. Deze kwesties kunnen de bij het beleid betrokkenen het beste gezamenlijk formuleren. Dat dwingt tot maximale scherppte.

Het is tegelijkertijd duidelijk dat (quasi)experimentele designs in de beleidswereld heel wat moeilijker te realiseren zijn dan in bijvoorbeeld medisch onderzoek. Denk daarbij aan het definiëren van een adequate controleconditie – zoals het (al dan niet at random) uitrollen van beleid in een deel van de uiteindelijke doelgroep en in een ander deel (nog) niet – en de follow-up tijd die minimaal nodig is om het uiteindelijke effect te kunnen vaststellen.

Toch worden vergelijkende beleidsevaluaties internationaal steeds vaker uitgevoerd en systematisch gedocumenteerd, samengevat en beoordeeld, zoals blijkt uit de zich gestaag ontwikkelende Campbell Collaboration Library of Systematic Reviews.<sup>47, 48</sup> Deze *evidence base* dekt echter nog maar een beperkt deel van de diverse beleidsvelden, en vereist continu aanvulling op basis van 'evidence chase'. Hier ligt een belangrijke verantwoordelijkheid voor evaluatie-onderzoekers.

Regelmatig wordt de vraag gesteld hoe de door middel van wetenschappelijke methoden verkregen kennis zich verhoudt tot de kennis en ervaring die professionals bij de uitoefening van hun vak opdoen. Moeten praktische

wijsheid en ervaring geen voorrang krijgen? Er is echter geen principiële tegenstelling tussen kennisverwerving via methodologisch doordachte evaluatiedesigns en het alledaagse praktijkleren van professionals. Integendeel, deze vormen van leren vullen elkaar aan. Dat past goed in een mixed-methods-benadering waarin diverse onderzoeksdesigns, zowel kwantitatief als kwalitatief van aard, complementair worden ingezet.<sup>41</sup> Naast vergelijkende studies, surveys, bestandsanalyses en documentanalyses is het oogsten van praktijkervaring via interviews, focusgroepen en praktijkobservaties onmisbaar.

Al met al heeft, juist ook in een complexe politiek-bestuurlijke context, evaluatie op basis van gedegen evidence een onmisbare meerwaarde. De focus moet daarbij gericht zijn op de essentiële vragen waarover controversen dan wel onzekerheid bestaat. Gestreefd dient te worden naar een – voor het type vraag dat aan de orde is – zo hoog mogelijk niveau van bewijskracht, met een brede benadering van het begrip evidence. Een mixed-methods-benadering, met aandacht voor praktijkkennis, is daarbij vaak aangewezen.

## Breder leren biedt extra kansen

Management in het publieke domein moet kunnen inspelen op nieuwe, soms onverwachte beleidsuitdagingen en vraagt om het zoeken en combineren van steeds nieuwe kennis.<sup>49</sup> In de woorden van 't Hart: 'In een wereld van accelererende kennisproductie en -diffusie is stilstand achteruitgang. En zonder actuele en meervoudige kennis kun je nauwelijks nog waarde toevoegen in de netwerken waarin bestuur gestalte krijgt.'<sup>32</sup>

Voor het kunnen inspelen op actuele en toekomstige ontwikkelingen lijken evaluaties op het eerste gezicht een minder geschikte kennisbron omdat ze op eerder ontworpen beleid gericht zijn. Evaluaties kunnen echter belangrijke leerpunten opleveren ten behoeve van preventie, anticipatie, of een betere beleidsrespons in vergelijkbare toekomstige situaties. En op basis van opgedane inzichten in mechanismen en causale relaties zal men beter voorbereid zijn op nieuwe ontwikkelingen waarin die mechanismen ook een rol spelen. Zulke inzichten zijn bovendien bruikbaar voor ex ante evaluatie van voorgenomen beleid.<sup>37</sup> Natuurlijk kunnen niet alle opkomende kwesties vooraf geëxploreerd worden. Maar een goed gekozen evaluatieportfolio kan via een 'pars-pro-toto-effect' bijdragen aan een kennisklimaat dat ook de beleidskwaliteit op relatief minder belichte terreinen ten goede komt.

### Box 2: Het belang van double loop leren en een brede blik volgens King en Crewe\*

*'The National Audit Office, the House of Commons Public Accounts Committee and some, though not all, other parliamentary committees are admirable bodies, and we frequently cite their reports. That said, however, their reports and the investigations that lead up to them typically suffer from two limitations, both to some extent self-imposed. One is that, partly out of a desire to operate on a non-partisan, dispassionate basis, they largely focus on the 'what' questions and tend to neglect the 'why' questions. They say that something went wrong, describe what went wrong and usually say what they think should be done to avoid the same kind of thing going wrong in future; but they seldom delve deeply into the causes of whatever went wrong. In particular, they seldom explore the decision making by ministers and officials that led to the committing of the blunder in question. Secondly, the various investigative bodies typically operate on a case-by-case basis. They only rarely step back and try to discern patterns of behaviour of the kind that we try to identify here. Just as much of British government is not, as they say, 'joined up', so most of the research of these other bodies tends to be highly segmented, without lines being drawn between the dots.'*

\* King A., & Crewe, I. (2013). *The Blunders of Our Governments*. London: Oneworld Publications.

We hebben het dan in feite ook over 'tweede orde'- of 'double loop'-leren,<sup>50</sup> waarmee men uitstijgt boven afzonderlijke evaluaties (zie ook box 2). Nagaan waarom een bepaalde aanpak al dan niet werkt dan wel beter zou kunnen werken, kan ook in andere domeinen tot betere werkwijzen leiden. Tweede orde leren helpt om buiten bestaande kaders te denken,<sup>51</sup> en kan tegenwicht bieden tegen kortetermijngerichte prikkels voortkomend uit politieke opportuniteit.<sup>52</sup> Informatie over wat wel en niet aanslaat in het dagelijkse politieke spel tussen departementen en parlement is immers gemakkelijker voorhanden dan 'langzame' informatie over wat bezien over een aantal jaren wel en niet blijkt te werken: 'political learning' gaat sneller dan 'policy learning'.<sup>53</sup> Maar juist die laatste vorm van leren is nodig om goed onderbouwd beleid te kunnen ontwikkelen en onderhouden, en om nieuwe situaties tegemoet te treden.

Alles overziend moge het duidelijk zijn dat evaluaties niet alleen op hun onderwerp-specifieke meerwaarde moeten worden beoordeeld, maar ook naar de mate waarin ze bijdragen aan meer generieke kennis, vaardigheden en werkwijzen. De evaluatiepraktijk kan daarop worden ingericht door vooraf telkens de vraag te stellen: hoe draagt de uit te voeren evaluatie bij aan de algemene 'body of knowledge'? Dat vereist dan wel een context van systematische kennisopbouw, een evenwichtige en gemengde evaluatieportfolio, permanente aandacht voor de bruikbaarheid van evaluaties en daaruit voortvloeiende aanbevelingen, en – als randvoorwaarde – zowel afstand als betrokkenheid van evaluatoren.

## Systematische kennisopbouw als doorlopende kennisbron

Systematische kennisopbouw maakt het mogelijk om informatie uit evaluaties ook voor andere dossiers en voor de langere termijn beleidsrelevant te maken en te houden (zie ook box 3). Dat is een voorwaarde voor doelmatige

benutting van kennis en expertise en hoort thuis in de kennisinfrastructuur ten behoeve van het openbaar bestuur en de rijksdienst.<sup>33,35</sup>

In dit verband pleitte de WRR voor een sterkere rijksbrede verankering van (gedrags)wetenschappelijke kennis voor beleid en de daarvoor benodigde capaciteit.<sup>13</sup> Zulke kennis is onontbeerlijk voor het voorbereiden van beleid, het toetsen van beleidsveronderstellingen, (ex ante) impact assessments, en beleidsevaluatie. Immers, beleid is vrijwel altijd gericht op het veranderen van het gedrag van mensen. Daarom zijn gedragswetenschappelijke inzichten ook niet alleen toepasbaar op de doelgroepen van het beleid maar ook op beleidsverantwoordelijken zélf.<sup>54</sup> Gedragskennis kan bijdragen aan het evalueren van het functioneren van (mensen in) organisaties en daarmee leiden tot nieuwe werkwijzen en verbeteringen in en van overheidsorganisaties.

Box 3: Het belang van systematische kennisopbouw via evaluaties volgens Furubo et al.\*

*'Instead of a model in which evaluations are regarded as knowledge channeled directly to the political decision makers and other stake holders, an alternative way of looking at things is to use the metaphor of a knowledge bank with information deposits. The officials of the bank, to continue the metaphor, interpret the information delivered in different studies, rearrange the information, and relate it to earlier knowledge in the field.'*

*'The immediate users of evaluations are not only the decision makers, but also the officials of our metaphorical bank (...). So, the extent to which the information, which we have gained in connection with earlier governmental interventions, will actually be channeled into the political and administrative system depends on to what degree it also can contribute to the building of more general knowledge.'*

\* Furubo, J.E., Rist, R.C., & Speer, S. (2013). *Evaluation and Turbulent Times: Reflections on a Discipline in Disarray*. New Jersey: Transaction Publishers.

Systematische kennisopbouw maakt het mogelijk de evaluatiefunctie te koppelen aan een informatiefunctie waarmee beleidsmakers op de hoogte kunnen worden gehouden van de internationale stand van kennis, relevante systematische reviews<sup>44,46</sup> en andere wetenschappelijk gedocumenteerde ervaringen. Dat kan gepaard gaan met vergelijkende internationale analyse, gericht op overeenkomsten en verschillen met Nederland. Waar beleidsmakers in de 'heat of the moment' niet gauw naar de laatste wetenschappelijke gegevens grijpen, kan zo'n informatievoorziening relevante inzichten en conclusies actief aanreiken. Welke wetenschappelijke benaderingen zijn er? Welk type kennis leveren ze op? Wat is daarvan bruikbaar in de nationale beleidscontext?

Systematische opbouw van kennis en ervaring dient ook leerpunten over minder geslaagd beleid te omvatten, want juist ook daarvan is veel te leren. Dat geldt te meer naarmate de redenen waarom iets niet werkte, kunnen worden verbonden met actuele beleidsuitdagingen. De waarde die terecht gehecht wordt aan best practices, mag daarom niet leiden tot het weggijken van niet-succesvolle regelgeving en beleidspraktijken en de factoren die daaraan hebben bijgedragen.

## Het belang van een evenwichtige evaluatieportfolio

Systematische kennisopbouw is gebaat bij een evenwichtige evaluatieportfolio naar thematiek en methode. Daarbij is het – zoals eerder toegelicht – van belang om doelmatigheid en effectiviteit ruim op te vatten wat betreft te onderzoeken waarden, grote naast kleine evaluaties te programmeren, ex ante naast ex post evaluaties te verrichten, en verschillende, elkaar aanvullende en eventueel tegensprekende bronnen van kennis aan te boren.

Een evenwichtige evaluatieportfolio, waarin individuele evaluaties worden gezien als deel van een geheel, bevordert dat specifieke evaluaties input leveren voor het geheel aan kennis (de body of knowledge) op verschillende niveaus van beleid: feitelijke details (bijvoorbeeld ten behoeve van de verbetering van specifieke interventies), grote lijnen (van bijvoorbeeld wetgeving of beleidsprogramma's), en de samenhang tussen grote lijnen, de bredere beleidsstrategie. Steeds moet de vraag worden gesteld hoe een bepaalde evaluatie aan systematische kennisopbouw kan bijdragen, naast de kennis die nodig is voor het specifieke beleidsdossier. Zo zou aan elke specifieke (wets)evaluatie een standaarddeel gekoppeld kunnen worden, gericht op een meer generieke kennisopbouw. Het is dan zaak ook de methodologie en de rapportage van evaluaties zo in te richten dat ze valide en reproduceerbare informatie kunnen leveren voor systematische reviews.<sup>44,46</sup>

Binnen een op systematische ontwikkeling gerichte, gemengde portfolio kunnen bepaalde algemene waarden – bijvoorbeeld vanuit het parlement aangedragen – aandacht krijgen, ook gespreid over meerdere evaluaties. Specifieke doelen worden immers vaak mede gesteld om een 'algemenere waarde' te realiseren. Effectiviteitsonderzoek en waardenonderzoek liggen daarbij in elkaars verlengde: waarden zijn richtinggevend voor het belang dat men aan bepaalde effecten hecht, en effecten ontlenen hun relevantie aan hun implicaties voor essentiële waarden.

Dikwijls zijn bepaalde waarden in het doel van een wet of beleidsinitiatief verankerd. Zo onderzoeken evaluaties van de Algemene wet gelijke behandeling de mate waarin naleving en handhaving van de wet bijdragen aan (het ervaren van) gelijke behandeling en het bestrijden van discriminatie en ongelijke behandeling. Zulke waarden

kunnen soms ook uit doelen van wetgeving en beleid worden afgeleid als ze daarin niet expliciet zijn vermeld.

Het vooraf in beeld brengen van belangrijke waarden en het evalueren van de realisatie daarvan geeft meer sturing en is beter haalbaar dan 'goal-free' evaluation<sup>55</sup> (waarbij de expliciet geformuleerde beleidsdoelen en effecten niet bepalend zijn voor het evaluatieonderzoek maar geprobeerd wordt alle relevante effecten en neveneffecten in kaart te brengen), maar is minder beperkend dan gericht effectiviteitsonderzoek (dat zich geheel toespitst op dossier-specifieke (bijvoorbeeld budgettaire) doelbereiking). Hierbij kunnen verschillende perspectieven van gebruikers en geadresseerden aan het licht komen en wordt de kans op draagvlak onder beleidsmakers en op doorwerking van lessen vergroot.

Vooraf bij complexe en risicovolle dossiers met veel onzekerheden is het nuttig expliciet te onderzoeken of er belangrijke waarden in het geding komen als neveneffect van beleid. Dat kan belangrijke input bieden voor bijstelling of herziening van lopend beleid. Daarbij valt te denken aan de evaluatie van veelomvattende – en qua mogelijke neveneffecten voor burgers en professionals vooraf lastig te overziene – interventies als stelselherzieningen en decentralisatieoperaties. Het kan dan bijvoorbeeld gaan om zaken als kwaliteit van de thuiszorg, belemmerende bureaucratie voor professionals, of rechtsongelijkheid qua beschikbare voorzieningen in verschillende gemeenten.

Systematische kennisopbouw op basis van een evenwichtige evaluatieportfolio moet gebruik kunnen maken van een veelheid aan informatiebronnen, inclusief informatie van communicatie-afdelingen van ministeries, toezichthouders, planbureaus en monitors. Hierbij spelen de snelle ontwikkelingen in de 'digital society' – en de daaraan gerelateerde mogelijkheden en vereisten op het gebied van de beschikbaarheid van en de koppeling tussen data en het gebruik daarvan ('open data' en big data) – een toenemende rol.<sup>56, 57, 58</sup> Ook het bevragen van burgers en doelgroepen, inclusief het volgen van trends in hun opvattingen, is een belangrijke informatiebron.<sup>59</sup>

Tussen het streven naar een evenwichtige evaluatieportfolio en de meer gefocusseerde sturing van de evaluatieagenda vanuit actuele beleidsrelevantie kan spanning optreden: komen alle specifieke evaluatiedoelen wel aan bod? Inbedding in een breder kennisinfrastructureel kader, met een grote thematische diversiteit, biedt echter juist ook de mogelijkheid om beide doeleinden tegelijk te behartigen. Er is dan een gevarieerd repertoire voorhanden van waaruit flexibel kan worden ingespeeld op – deels niet te voorziene – beleidsactualiteit. Het hanteren van een brede en overzichtelijke portfolio helpt bovendien om overmatig evalueren in vergelijkbare dossiers tegen te gaan.

## Bruikbaarheid in de praktijk als toetssteen

Wets- en beleidsevaluaties gaan vaak over praktijken die op grote afstand van de beleidsmaker en de politicus plaatsvinden. Daarom is het zinvol dat evaluaties ook een bottom-up perspectief bieden op de uitvoering van en ervaring met regels. Evaluaties die alleen het top-down perspectief belichten (heeft men gedaan en bereikt wat in het beleid is afgesproken en zo nee, hoe kan dat alsnog bereikt worden?) en onvoldoende feedback vanuit het werkveld inbouwen, brengen het risico met zich mee dat onwenselijke neveneffecten onderbelicht blijven en dat men aan onhaalbaar beleid blijft vasthouden. Het helpt als evaluatierapporten een heldere beschrijving geven van de beleidspraktijk en de knelpunten die zich daarin voordoen, en dat daarbij geen cruciale details verloren gaan.

Dat juist degenen die eindverantwoordelijkheid dragen, de hiërarchisch hoger geplaatsten die relatief ver van het praktijkveld af staan, weinig gedetailleerde informatie krijgen, is begrijpelijk want zij hebben beperkte leestijd. Maar tegelijkertijd moeten voor hen zowel de problemen als de mogelijkheden voor kansrijke beleidsaanpassing voldoende concreet zichtbaar zijn. Het is dan ook de kunst rapportages weliswaar beknopt te houden maar zo op te zetten dat naast de abstractere hoofdlijnen de essentialia van de werkvloersituatie worden vermeld, met de 'problemdiagnose' en daaraan gerelateerde mogelijke oplossingsrichtingen.

Gevoelige snaren moeten daarbij niet worden vermeden. Het herschrijven van stukken totdat ze aanvaardbaar zijn voor alle betrokkenen, kan leiden tot zodanige algemeenheden dat niet meer duidelijk is of aangedragen oplossingen wel passen bij de onderliggende problemen. In dit verband moet worden gewaarschuwd voor bureaucratisch schrijven dat leidt tot een 'writing that is written by many hands' (naar Vinning).<sup>60, 61</sup> Hierop moet ook worden gelet bij het betrekken van belanghebbenden bij het beoordelen van conceptrapportages van bevindingen.

De hamvraag is of de oplossingen die door de evaluator worden aangereikt, de ontvanger van de evaluatie verder zullen helpen. Of het nu gaat om lessen omtrent ongemakken, herstructurering van afdelingen of het verleggen van verantwoordelijkheden: gaan de aangedragen oplossingen ook echt werken? Om kansrijke oplossingen te kunnen ontwikkelen is gedegen kennis nodig van het praktijkveld, de gewoonten van de doelgroep en de invloed van de omgeving.<sup>62</sup> Zo moet men bij het ontwerpen van beleid met betrekking tot uitkeringsfraude inzicht hebben in de motieven van betrokkenen,<sup>63</sup> en is preventiebeleid gericht op duurzame gedragsverandering alleen kansrijk als men rekening houdt met omgevingsfactoren.<sup>64</sup>

Ook is van belang dat de evaluator handelingsperspectieven aanreikt die aansluiten op het relevante belevingsniveau. Veel evaluaties bevatten echter nogal algemene aanbevelingen die bovendien vaak raken aan fenomenen die in de bestuurskunde als notoir lastig bekendstaan, zoals het aanpakken van 'coördinatieproblemen' of 'eerder opschalen'.

Als een evaluator onvoldoende zicht kan krijgen op een specifieke uitvoeringspraktijk om voldoende concrete



aanbevelingen te kunnen doen, is een gedetailleerde beschrijving van knelpunten – zonder al uitgewerkte aanbevelingen aan te dragen – een nuttig alternatief. De evaluatie heeft dan primair een signalerende en agenderende functie, met als doel beleids- en velddeskundigen uit te dagen tot een door henzelf uit te werken aanpak. Dat is zinvoller dan het doen van aanbevelingen die als vaag, onwerkbaar of onhaalbaar worden ervaren, en waarbij het risico bestaat dat niet alleen deze aanbevelingen maar ook de daaraan voorafgaande analyse genegeerd worden. Het uitdagen van de eigen creativiteit uitgaande van die analyse biedt dan meer perspectief.

## Zowel afstand als betrokkenheid nodig

Onafhankelijkheid en onpartijdigheid zijn essentieel voor de validiteit en toegevoegde waarde van de evaluatie én voor het publiek en politiek-bestuurlijk vertrouwen in de resultaten ervan.

Onafhankelijkheid is geen abstract gegeven maar moet worden geborgd door een daarbij passende positionering van de evaluator.<sup>36,41</sup> Maar onafhankelijkheid schuilt ook in de personen die de evaluatie uitvoeren, en in de wijze waarop wordt omgegaan met mogelijke belangen die evaluatoren hebben bij het object van evaluatie en met relaties met actoren in het beleidsproces. Evenzeer van belang zijn de wijze waarop de opdracht wordt geformuleerd, de samenstelling van begeleidingscommissies, de gehanteerde procedure inzake belangenverstrengeling,<sup>65</sup> de vormgeving van het evaluatieproces, de toegang tot alle relevante informatie, de wijze waarop belanghebbenden inbreng leveren, de mogelijkheid van de evaluator om onderzoeksresultaten vrijelijk en volledig te publiceren, en de publieke transparantie van al deze aspecten. De vrijheid van de evaluator om te publiceren over aard, vormgeving en resultaten van het onderzoek maakt het evaluatieproces openbaar navolgbaar en toetsbaar. De evaluator kan daarmee ook verantwoording afleggen over de ervaren (on)mogelijkheden om zijn werk naar behoren te doen.

Onafhankelijkheid moet voortdurend, in de context van de dagelijkse evaluatiepraktijk, worden waargemaakt. Zij moet echter niet verworden tot een toestand van splendid isolation.<sup>41</sup> Terwijl voldoende afstand een voorwaarde is voor de onafhankelijkheid van het evaluatieproces, kunnen te weinig betrokkenheid en een te grote afstand tussen evaluatieonderzoekers en beleidsmakers juist weer barrières vormen voor de bruikbaarheid en acceptatie van resultaten uit evaluatiestudies. In dit verband hebben aan de overheid gelieerde onderzoekseenheden, mits onafhankelijk gepositioneerd ten opzichte van de te evalueren beleidsvoorbereiding en -uitvoering zoals bij het WODC het geval is,<sup>66</sup> een voordeel ten opzichte van bijvoorbeeld academische evaluatoren: zij kunnen inhoudelijke onafhankelijkheid combineren met inhoudelijke nabijheid ten opzichte van beleid en met gevoel voor politiek-bestuurlijke verhoudingen. Effectieve onafhankelijkheid betreft dan niet alleen de te hanteren onderzoeksmethoden, maar ook de evaluatieagenda. Dat wil zeggen dat een evaluatie-eenheid ook uit eigen beweging onderzoek moet kunnen entameren indien zij daartoe, juist gegeven haar inzicht in de beleidscontext, aanleiding ziet.

Bij het uitvoeren van evaluatieonderzoek biedt samenwerking met universiteiten en andere externe onderzoeksinstituten – inclusief de instelling van op evaluatie gerichte leerstoelen – specifieke meerwaarde. Die samenwerking vergemakkelijkt de toegankelijkheid van up-to-date wetenschappelijke kennis en kan de onafhankelijkheid van evaluatieonderzoek versterken en extern verankeren.

## Beschouwing en conclusies

Evaluatie van beleid hoort bij goede beleidsvoering, net als het ontwerpen, de implementatie, en de uitvoering ervan. Daarbij gaat het om zowel verantwoordten als leren. Het verantwoordten draagt bij aan het leren, voor zover het gericht is op verbetering en herkansing. De kernboodschap van dit artikel is dat, waar het gaat om leren, beleidsevaluaties niet alleen van betekenis zijn voor het specifieke onderwerp waarop zij primair gericht zijn, maar ook moeten bijdragen aan systematische en evenwichtige kennisopbouw ten behoeve van voortgaande beleidsverbetering. Het interdepartementaal verbinden van expertise is hierbij van belang.

Bij het uitvoeren van evaluaties moet ermee rekening worden gehouden dat vele factoren het leren van beleidsevaluaties kunnen bevorderen of belemmeren. Gedegen kennis daarover biedt aangrijpingspunten om kostbare evaluatiekennis beter te benutten. Bij het beoordelen van effectiviteit en doelmatigheid van beleid mogen lastig meetbare maar essentiële waarden – zoals subjectief welzijn, rechtvaardigheid en maatschappelijke aanvaardbaarheid – niet onderbelicht blijven. Een goede timing van evaluaties is van groot belang, evenals voldoende capaciteit en continuïteit ten behoeve van de uitvoering en benutting van evaluaties.

De overheid onderkent de behoefte aan gedegen onderbouwde evaluatiekennis, maar de expertise en middelen om die kennis aan te reiken staan onder druk. Er is sprake van een paradoxaal spanningsveld tussen de toenemende aandacht voor verwetenschappelijking van beleid enerzijds en de groeiende vraag naar politieke sensitiviteit en beleidsstrategische vaardigheden van ambtenaren anderzijds. Binnen deze politiek-bestuurlijke context moet beleidsevaluatie stevig verankerd zijn om haar onafhankelijke en onpartijdige rol te kunnen vervullen. Dat is nodig voor de validiteit en toegevoegde waarde van evaluaties, maar ook voor het publiek en politiek-bestuurlijk vertrouwen daarin. Ook publieke transparantie is daarvoor essentieel.

Er hoeft niet altijd een externe evaluator te worden ingeschakeld; men kan te werk gaan volgens een getrapte benadering. Net als een slager als eerste moet willen weten of zijn vlees van goede kwaliteit is, ligt de primaire evaluatieverantwoordelijkheid bij het politiek bestuur. Dat dient te zorgen voor een adequate beleidsevaluatie, onafhankelijk aangestuurd door een daartoe geëquiperde eenheid. Als de resultaten of de onpartijdigheid daarvan twijfels oproepen, of als sprake is van bijzondere risico's of controverses, moeten externe onafhankelijke evaluatoren worden ingeschakeld. Er zijn, via samenwerking tussen overheid en externe (wetenschappelijke)

experts, ook tussenvormen mogelijk. Voorzien moet worden in state-of-the-art conflict-of-interest procedures.<sup>67</sup> Via periodieke algemene reviews kan extern getoetst worden of de overheid zijn evaluatieprocessen op orde heeft.

Omdat het onmogelijk is alle elementen van een beleidsinterventie wetenschappelijk te evalueren, verdient het aanbeveling zich te concentreren op de essentiële kwesties waarover vooraf discussie of onzekerheid bestaat. Goede methodologische kwaliteit is voor beleidsevaluatie essentieel, waarbij vaak behoefte is aan een mixed-methods-benadering met ruimte voor zowel kwantificerend vergelijkend als kwalitatief onderzoek. Ook dient er aandacht te zijn voor ex ante evaluatie, met zoveel mogelijk inzicht in werkingsmechanismen van beleidsmaatregelen, gericht op de vraag of verwacht mag worden dat het publiek belang inderdaad met het (beoogde) beleid wordt gediend.

Bij beleidsevaluatie moet gestuurd worden op een bredere beleidsimpact dan alleen dossier-specifieke doelbereiking, en op bijdragen aan meer generieke kennisontwikkeling. In dit verband is systematische opbouw van kennis en ervaring, gekoppeld aan een beleidsrelevante onderzoeksagenda cruciaal. Dit sluit aan bij eerdere pleidooien voor het toegankelijk bijeenbrengen van de evaluatiebevindingen die de afzonderlijke wets- en beleidsterreinen overstijgen en de aanzetten die hiertoe zijn gegeven in het kader van het Clearing House voor Wetsevaluatie.<sup>24,68</sup>

Het kan daarbij helpen evaluaties te voorzien van een specifiek en een breder geldend algemeen deel. In het kader van een brede opbouw van kennis en ervaring om maximaal te kunnen leren is het ook wenselijk te sturen op de ontwikkeling van een evenwichtige, gemengde evaluatieportfolio per beleidsterrein, met aandacht voor diverse ex post en ex ante methoden, uiteenlopende waarden, en verschillende informatiebronnen. Inbreng vanuit gedragswetenschappelijk perspectief is onontbeerlijk. Niet alleen positieve ervaringen (inclusief best practices) maar ook minder geslaagd en mislukt beleid moeten in beeld worden gebracht. Het mijden en wegschrijven van gevoelige snaren is qua leerproces contraproductief.

Het bottom-up perspectief dient in evaluaties vertegenwoordigd te zijn, qua inbreng van informatie en ervaring maar ook om te kunnen komen tot bruikbare aanbevelingen voor de praktijk. Als de evaluator onvoldoende thuis is op een zeer specifiek gebied waarover maar weinigen praktijkkennis hebben, kan deze zich eventueel beperken tot het aangeven van de hoofdproblemen en het werkveld uitdagen om zelf verbeterpunten op te stellen.

Evaluatie van beleid is een qua beschikbare expertise en doorwerking kwetsbaar kernelement van de beleidscyclus, en moet aantrekkelijk blijven voor talentvolle medewerkers. Dat maakt aandacht, prioriteit en waardering voor beleidsevaluatie, en benutting en kwaliteitsborging ervan, extra belangrijk. Dat vereist ook goede verankering en positionering binnen de kennisinfrastructuur van de overheid en een effectieve samenwerking met voor beleidsevaluatie belangrijke kennisinstellingen.

Binnen die kennisinfrastructuur dient, ten slotte, ook steeds evaluatie-aandacht te zijn voor het up-to-date zijn en de meerwaarde van het evalueren zelf.<sup>36</sup> Het kan dan bijvoorbeeld gaan om de vraag of dataverzamelings-, monitoring- en evaluatiesystemen nog wel bij de tijd en voldoende dekkend en doelmatig zijn, en om follow-up onderzoek naar de vraag of de gegeven aanbevelingen zijn opgevolgd en in hoeverre dit ook daadwerkelijk tot verbetering heeft geleid. Dat is cruciaal voor het 'evaluatievermogen' van de lerende overheid, dat immers niet alleen het verrichten maar ook het gebruiken van evaluaties omvat.<sup>69</sup>

## Dankwoord

*Ik dank dr. Petra Jonkers, senior wetenschappelijk medewerker van de WRR, in het bijzonder voor het aanleveren van stukken tekst en de gedachteswisseling over systematische kennisopbouw en een evenwichtige evaluatieportfolio.*

## Noten

**1** Bij het schrijven van dit artikel is gebruikgemaakt van de wetenschappelijke en vakliteratuur en van bevindingen in het kader van diverse WRR-studies, met name over de lerende overheid, toezicht, gedragskennis en beleid, risico- en veiligheidsbeleid, en beleidsevaluatie.

**2** Bovens, M., & Schillemans, T. (2009). *Handboek publieke verantwoording*. Den Haag: Lemma.

**3** Regeling Periodiek Evaluatieonderzoek, Staatscourant 2012, nr. 18352.

**4** Knaap, P. van der. (2010). Veiligheidsbeleid: onderbouwd en effectief? De meerwaarde van beleidstheorieën voor beleid en beleidsevaluatie. *Tijdschrift voor Veiligheid*, 9(1), 6-21.

**5** Jonkers, R. (2013). Zet transparantie liever in voor bekritisbaarheid dan voor vertrouwen. In D. Broeders, J.E.J. Prins, H. Griffioen, P. Jonkers, M. Bokhorst & M. Sax (red.), *Speelruimte voor transparantere rechtspraak* (pp. 449-471). Wetenschappelijke Raad voor het Regeringsbeleid. Amsterdam: Amsterdam University Press.

**6** Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. New York: Oxford University Press.

- 7** Rouw, R. (2011). *Gevoel voor bewijs. Naar vloeiende verbindingen tussen kennis en beleid*. Atelierreeks. Den Haag: NSOB.
- 8** Furubo, J.E., Rist, R.C., & Speer, S. (2013). *Evaluation and turbulent times: Reflections on a discipline in disarray*. New Brunswick, NJ: Transaction Publishers.
- 9** Leeuw F., & Mevissen, J. (2015). Beleidstheorieën, gedragsmechanismen en evaluatieonderzoek. In: P.H.M. van Hoesel, J. Mevissen & L.H. Dekker (red.), *Kennis voor beleid. Beleidsonderzoek in Nederland*. Assen: Van Gorcum.
- 10** Nelen, H. (2008). *Evidence maze; het doolhof van het evaluatieonderzoek*. Oratie Universiteit Maastricht.
- 11** Nelen, H., Leeuw, H.B.M., Bakker, F., & Herrenberg, T. (2012). *In de war en uit de bocht; ex-ante evaluatie van de aanpak van solistische dreigers binnen de pilot Dreigingsmanagement*. Den Haag: Boom Lemma uitgevers.
- 12** Leeuw, F.L. (2008). *Gedragsmechanismen achter overheidsinterventies en rechtsregels*. Intreerede Universiteit Maastricht.
- 13** Klein Haarhuis, C., Smit, M., & Keulemans, S. (2014). Ex ante onderzoek in beeld: over aard, aantal en gebruik van ex ante onderzoek bij beleidsvoorbereiding. *Beleidsonderzoek Online*, DOI:10.5553/Beleidsonderzoek.000041.
- 14** Wetenschappelijke Raad voor het Regeringsbeleid. (2014). *Met kennis van gedrag beleid maken*. Amsterdam: Amsterdam University Press.
- 15** Rienstra, S. (2008). *De rol van kosten-batenanalyse in de besluitvorming*. Den Haag: Kennisinstituut voor Mobiliteitsbeleid.
- 16** Hanemaayer, D, Gucht, H. de, & Gerritsen, M. (2010). *Evaluatie maatschappelijke kosten-batenanalyses Nota Ruimte Budget projecten*. Den Haag: B&A Consulting.
- 17** Wetenschappelijke Raad voor het Regeringsbeleid. (2014). *Consistent maatwerk – handreikingen voor dossieroverstijgend risico- en veiligheidsbeleid*. Den Haag: WRR.
- 18** Hanemaayer, D. (2012). *Ex ante evaluatie in Nederland: de stand van zaken*. Den Haag: Boom Lemma uitgevers.
- 19** OECD. (2010). *Better regulation in the Netherlands*. Paris: Organisation for Economic Cooperation and Development.
- 20** Gezondheidsraad/Raad voor de Volksgezondheid & Zorg. (2006). *Vertrouwen in verantwoorde zorg? Effecten van en morele vragen bij het gebruik van prestatie-indicatoren*. Signalering ethiek en gezondheid 2006/1. Den Haag: Centrum voor ethiek en gezondheid.
- 21** Gezondheidsraad. (2013). *Publieke indicatoren voor kwaliteit van curatieve zorg. De stand van de discussie*. Publicatienr. 2013/29. Den Haag: Gezondheidsraad.
- 22** Noordergraaf, M., & Wit, B. de. (2012). *Van maakbaar naar betekenisvol bestuur. Een achtergrondstudie naar (keten)governance en (nieuw) publiek management en de gevolgen voor toezicht en evaluatie*. WRR-webpublicatie nr. 63. Den Haag: WRR.
- 23** Noordergraaf, M., & Abma, T. (2003). Management by measurement? Public management practices amidst ambiguity. *Public Administration*, 81(4), 853-871.
- 24** Wetenschappelijke Raad voor het Regeringsbeleid. (2006). *Lerende overheid: een pleidooi voor probleemgerichte politiek*. Amsterdam: Amsterdam University Press.
- 25** Klein Haarhuis, C.M., & Niemeijer, E. (2008). *Wet en werkelijkheid. Bevindingen uit evaluaties van wetten*. Den Haag: Boom Juridische uitgevers.
- 26** Ryan, B. (2003). 'Death by evaluation'?: Reflections on monitoring and evaluation in Australia and New Zealand [online]. *Evaluation Journal of Australasia*, 3(1), 6-16.
- 27** RIVM. (2014). *Gezondheid en veiligheid in de Omgevingswet. Doelen, normen en afwegingen bij de kwaliteit van de leefomgeving*. RIVM Rapport 2014-0138, hoofdrapport. Bilthoven: RIVM.
- 28** Nelen, J.M. (2000). *Gelet op de Wet, de evaluatie van strafwetgeving onder de loep*. Amsterdam: Vrije Universiteit Amsterdam.

- 29** Gils, G.H.C. van, Leeuw, F., Gildemyn, M., & Inberg, L. *Leren van Evalueren*. (2010). *Onderzoek naar het gebruik van evaluatieonderzoek bij het ministerie van Buitenlandse Zaken*. Den Haag: Ministerie van Buitenlandse Zaken, Directie Financieel Economische Zaken.
- 30** Korsten, A.F.A. (2013). *Wat ervan terecht komt. Zicht op beleidsevaluatie*. <http://www.arnokorsten.nl/PDF/Onderzoek/ABC%20van%20beleidsevaluatie.pdf>
- 31** KNAW-werkgroep opdrachtonderzoek. (2005). *Wetenschap op bestelling. Over de omgang tussen wetenschappelijk onderzoekers en hun opdrachtgevers*. Amsterdam: KNAW.
- 32** Enthoven, G.M.W. (2011). *Hoe vertellen we het de Kamer? Een empirisch onderzoek naar de informatierelatie tussen regering en parlement*. Proefschrift Universiteit van Tilburg.
- 33** Hart, P. 't. (2014). *Ambtelijk Vakmanschap 3.0: Zoektocht naar het handwerk van de overheidsmanager. Essay in opdracht van de Vereniging voor OverheidsManagement, Vereniging van Gemeentesecretarissen, stichting IKPOB*. Departement Bestuurs- en Organisationswetenschap Universiteit Utrecht & Nederlandse School voor Openbaar Bestuur.
- 34** Bekker, E.R. (2014). Afscheidscollege 'Een toekomst zonder ministeries'. Albeda Leerstoel Universiteit Leiden/CAOP. <http://www.deleerstoelen.nl/fileadmin/caop/data/Actueel/2014/Verslag-afscheidscollege-Roel-Bekker-en-Roel-Nieuwenkamp.pdf>
- 35** Nieuwenkamp, R. (2001). *De prijs van het politieke primaat; wederzijds vertrouwen en loyaliteit in de verhouding tussen bewindspersonen en ambtelijke top*. Proefschrift. Delft: Eburon.
- 36** Knottnerus, J.A. (2014). *Voordracht naar aanleiding van de afscheidsredes van Roel Bekker en Roel Nieuwenkamp*. Rijksuniversiteit Leiden, CAOP. <http://www.deleerstoelen.nl/fileadmin/caop/data/Actueel/2014/Verslag-afscheidscollege-Roel-Bekker-en-Roel-Nieuwenkamp.pdf>
- 37** *Blueprints for informed policy decisions: A review of laws and policies requiring routine evaluation*. Report from Kunnskapssenteret (Norwegian Knowledge Centre for the Health Services). No 16-2012 Review. Oslo, 2012.
- 38** Kabinetsreactie op adviesrapporten van Rli, RMO en WRR over de benutting van gedragswetenschappelijke kennis in beleid. Den Haag: Ministerie van Economische Zaken, 4 december 2014.
- 39** Rli. (2014). *Doen en laten. Effectiever beleid door mensenkennis*. Den Haag: Raad voor de leefomgeving en infrastructuur.
- 40** RMO. (2014). *De verleiding weerstaan; grenzen aan beïnvloeding van gedrag door de overheid*. Den Haag: Raad voor Maatschappelijke Ontwikkeling.
- 41** <https://www.rijksoverheid.nl/onderwerpen/veilig-leren-en-werken-in-het-onderwijs/vraag-en-antwoord/aanpak-pest-en-school> (geraadpleegd d.d. 17 december 2015)
- 42** Wetenschappelijke Raad voor het Regeringsbeleid. (2013). *Toeziën op publieke belangen. Naar een verruimd perspectief op rijkstoezicht*. Amsterdam: Amsterdam University Press.
- 43** Sherman, L.W., Gottfredson, D.C., MacKenzie, D.L., Eck, J., Reuter, P., & Bushway, S.D. (1998). *Preventing crime: What works, what doesn't, what's promising*. Research in Brief, National Institute of Justice, pp. 1-19. <https://www.ncjrs.gov/pdffiles/171676.PDF>
- 44** Farrington, D., Gottfredson, D., Sherman, L., & Welsh, B. (2002). The Maryland Scientific Methods Scale. In L. Sherman, D. Farrington, B. Welsh & D. Mackenzie (Eds.), *Evidence-based crime prevention*. London: Routledge.
- 45** Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden: John Wiley & Sons.
- 46** Petticrew, M., McKee, M., Lock, K., Green, J., & Phillips, G. (2013). In search of social equipoise: A failure to acknowledge uncertainty about the effectiveness of social interventions is a major barrier to evidence based public policy making. *BMJ*, 347, [4016]. doi:10.1136/bmj.f4016
- 47** <http://www.campbellcollaboration.org/>
- 48** Filges, T., Smedslund, G., Due Knudsen, A.S., & Klint Jorgensen, A.M. (2015). Active labour market programme participation for unemployment insurance recipients: A systematic review. *Campbell Systematic Reviews*, 2, January 2.
- 49** Noordegraaf, M., Geuijen, K., & Meijer, A. (red.). (2011). *Handboek publiek management*. Den Haag: Boom

Lemma uitgevers.

**50** Argyris, C. (1977). Double loop learning in organizations. *Harvard Business Review*, September-October, 115-125.

**51** Asselt, M.B.A. van, Faas, A., Molen F. van der, & Veenman, S.A. (red.). (2010). *Uit zicht. Toekomst verkennen met beleid*. Wetenschappelijke Raad voor het Regeringsbeleid. Amsterdam: Amsterdam University Press.

**52** Ringeling, A.B. (2004). Wie vertegenwoordigt wat? en hoe? In Raad voor het Openbaar Bestuur, *Democratische vergezichten. Essays over de representatieve democratie in Nederland*. Den Haag.

**53** May, P.J. (1992). Policy learning and failure. *Journal of Public Policy*, 12(4), 331-354.

**54** Sunstein, C.R. (2013). *Simpler: The future of government*. New York: Simon & Schuster.

**55** Scriven, M. (1972). Pros and cons about goal-free evaluation. *Journal of Educational Evaluation*, 3(4), 1-7.

**56** Algemene Rekenkamer. (2014). *Tendrapport open data*. Den Haag.

**57** Janssen, R. (2015). *The art of audit*. Algemene Rekenkamer. Amsterdam: Amsterdam University Press.

**58** Wetenschappelijke Raad voor het Regeringsbeleid. (2016). *Big Data in een vrije en veilige samenleving*. Amsterdam: Amsterdam University Press.

**59** Sociaal en Cultureel Planbureau. *Continu onderzoek burgerperspectieven (COB) – SCP*. [http://www.scp.nl/Onderzoek/Bronnen/Beknopte\\_onderzoeksbeschrijvingen/Continu\\_onderzoek\\_burgerperspectieven\\_COB](http://www.scp.nl/Onderzoek/Bronnen/Beknopte_onderzoeksbeschrijvingen/Continu_onderzoek_burgerperspectieven_COB)

**60** Drenth von Februar, M. (2006). De rapporterende burger. Over incidentalisme versus cerebralisering. In P.L. Meurs, E.K. Schrijvers & G.H. de Vries (red.), *Leren van de praktijk. Gebruik van lokale kennis en ervaring voor beleid*. Wetenschappelijke Raad voor het Regeringsbeleid. Amsterdam: Amsterdam University Press.

**61** Vining, J. (1986). *The authoritative and the authoritarian*. Chicago: University of Chicago Press.

**62** Korsten, A.F.A., & Mevissen, J. (2015). Risico's en remedies. Kritiek op beleidsonderzoek en wat ervan te leren. In P.H.M. van Hoesel, J.W.M. Mevissen & L.H. Dekker (red.), *Kennis voor beleid – Beleidsonderzoek in Nederland*. Assen: Van Gorcum.

**63** Brummelkamp, G.W., Kerckhaert, A.C., & Engelen, W.M. (2013). *Het verhaal achter notoire uitkeringsfraude*. Zoetermeer: Panteia.

**64** Berg, M. van den, Post, N.A.M., Hamberg-van Reenen, H.H., Baan, C.A., & Schoemaker, C.G. (red.). (2014). *Preventie in de zorg*. Themarapport Volksgezondheid Toekomst Verkenning 2014. Bilthoven: RIVM.

**65** Code ter voorkoming van oneigenlijke beïnvloeding door belangenverstremeling. Koninklijke Nederlandse Akademie van Wetenschappen, Koninklijke Nederlandsche Maatschappij tot bevordering der Geneeskunst, Gezondheidsraad, Centraal BegeleidingsOrgaan, Nederlands Huisartsen Genootschap, Orde van Medisch Specialisten, 2012.

**66** Missie en kerntaken WODC, <https://www.wodc.nl/organisatie/>, 28 december 2015.

**67** Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie. (2009). *Evaluatiebeleid en richtlijnen voor evaluaties*. Den Haag: Ministerie van Buitenlandse Zaken.

**68** Veerman, G.J., m.m.v. R.J. Mulder & E.S.M. Meijsing. (2013). *Een empathische wetgever. Meta-evaluatie van empirisch onderzoek naar de werking van wetten*. Den Haag: Sdu.

**69** Klein Haarhuis, C.M., & Parapuf, A. (2016). *Evaluatievermogen bij beleidsdepartementen. Praktijken rond uitvoering en gebruik van ex post beleids- en wetsevaluaties*. Concept Eindrapportage. Den Haag: WODC.