

G.J. Veerman

EVALUATIE VAN WETGEVING? BOUWSTENEN VOOR BESLISSINGEN



Leiden 2014

Samenvatting

Met grote regelmaat ligt bij 'het beleid' of 'de wetgeving' de vraag voor of nader onderzoek nodig is alvorens het wetsvoorstel wordt opgesteld of gelanceerd (ex ante analyse) dan wel of het nodig zal zijn de werking en doelbereiking van het tot wet gepromoveerde wetsvoorstel met behulp van empirisch onderzoek vast te stellen (ex post evaluatie). De beslissing een ex post evaluatie te houden kan worden genomen zowel in de fase van het wetsvoorstel als in de fase dat de wet al in werking is. Bij dergelijke beslissingen spelen diverse vragen een rol: waarom, wat, wanneer, waar, wie en hoe. Op deze vragen wordt ingegaan en de overwegingen die daarbij worden genoemd, kunnen worden gebruikt als bouwstenen voor die beslissing. De belangrijkste vraag is het 'wat', de overwegingen bij de andere vragen betekenen in veel gevallen beperkingen van de mogelijkheid het 'wat' te onderzoeken. Voor het 'wat' valt geen algemeen antwoord te geven, daarom noem ik diverse mogelijkheden. Belangrijk is dat de opdrachtgever een beeld heeft van wat hij wil weten. Want naarmate hij minder weet wat hij *wil* krijgen, weet hij ook minder wat hij *zal* krijgen.

Een samenvatting van de overwegingen geef ik hieronder schematisch weer.

	Ex ante evaluatie	Ex post evaluatie
Waarom wetsevaluatie	Reductie van onzekerheid; Potentieel betere voorspellingen en betere werking van de voorgenomen wet; beter 'evidence based' beleid; Strategisch tbv draagvlak voor wetsvoorstel.	Reductie van onzekerheid; Inzicht in werking/ doelbereiking; Onderbouwde aanpassing vd wet; Strategisch tbv compromis over wetsvoorstel; Reservoir van relevante kennis vullen.
Wat te evalueren	IAK-vragen; Beleids­theorie (doel & middelen); Het Probleem (wat voor wie) Middelen (normadressaten moeten "weten, willen, kunnen"); Kosten-baten, economische randvoorwaarden; Juridische randvoorwaarden; Mogelijke knelpunten. P.M. onderzoek naar meer alternatieven, of naar een alternatief.	Juridische werking; Doelbereiking; Effecten; Doeltreffendheid; Uitvoeringsprocessen & mechanismen; Knelpunten; Relatie met ex ante analyse. PM: thema's (evaluatie van meer wetten).
Wanneer evalueren	Niveau van onzekerheid; Tijd en kosten; Onderzoekbaarheid; Verwacht gebruik/ opbrengst.	Niveau van onzekerheid; Tijd en kosten; Onderzoekbaarheid; Verwacht gebruik/ opbrengst.
Waar onderzoek	Desk research; Veldwerk igv o.a. probleemanalyse of knelpunten.	Deskresearch en veldwerk bij diverse groepen en op verschillende niveaus, triangulatie.
Wie evalueren/ organisatie	Bvk onderzoekers, begeleidingscommissie.	Bvk onderzoekers, begeleidingscommissie; evt stuurgroep.
Hoe evalueren	o.a.: Delphi, scenario, simulatie;	Quasi-experimenteel design, T)-T1

	extrapolatie; experiment.	design, T1 design; Contextueel design, gevalsstudies; experimenten; monitoring adhv indicatoren.
--	------------------------------	---

Inhoudsopgave

Samenvatting	2
Inhoudsopgave	4
1. Inleiding	5
2. Waarom wetsevaluatie?	7
3. Wat evalueren?	11
4. Wanneer wetsevaluatie?	22
5. Wie evalueren?	30
6. Waar evalueren?	34
7. Hoe evalueren?	36
Literatuur	43
Over de schrijver	44

1. Inleiding

In deze monografie worden bouwstenen aangedragen met het oog op (beleids)beslissingen die worden genomen of moeten worden genomen ter zake van een ex ante of een ex post evaluatie. Bij een ex ante evaluatie gaat het om de vraag of ter voorbereiding van een wet en het ermee samenhangende beleid nog informatie nodig is. De beslissing een ex post evaluatie uit te voeren wordt in veel gevallen eveneens genomen bij de voorbereiding van een wet, maar dan is de vraag of het verstandig is na enige tijd te bezien hoe de wet in de praktijk werkt. In beide gevallen gaat het om een zo empirisch mogelijke onderbouwing van het beleid, voorgenomen (ex ante) en uitgevoerd (ex post).

Deze bouwstenen zijn gegroepeerd aan de hand van de vragen die bij besluitvorming altijd wel spelen: waarom, wat, wanneer, wie, waar en hoe. Er bestaan wat de behandeling van die vragen betreft, verschillen tussen ex ante evaluatie en ex post evaluatie. Immers, in het ene geval gaat het om een voorspelling van wat er zal gebeuren, in het andere geval juist om waarneming van wat er feitelijk is gebeurd. Beide vormen naderen elkaar weer in de situatie dat de ex post evaluatie materiaal moet bieden voor eventuele vernieuwing van wetgeving; dat feitelijk materiaal is prognostisch materiaal voor de vraag of de wet moet worden aangepast. 'Prognostisch' vanwege het onderscheid tussen causale en finale redeneringen. Een eenvoudig voorbeeld daarvan: falen van beleid en wetgeving kan het gevolg zijn van te weinig personeel, maar de oplossing 'meer personeel' (causaal) kan zodanige neveneffecten hebben dat deze aanpak toch onwenselijk is; dan moet dus worden bezien hoe de oplossing wél kan worden bereikt (finaal).

Ik noem nog enkele kwesties.

-De definitie van 'wetsevaluatie' is niet scherp. Evalueren betekent letterlijk: de waarde vaststellen. Dat doen mensen voortdurend. In geval van wetsevaluaties wordt gedoeld op min of meer wetenschappelijke vormen van het vaststellen van de waarde, dat wil zeggen: aan de hand van een expliciet aangeduide en herhaalbare onderzoeksmethode, neergelegd in een openbaar rapport. In dit verband kan het wijs zijn het al genoemde onderscheid aan te brengen tussen enerzijds het wetenschappelijk onderzoek dat in hoge mate de empirie betreft, en anderzijds de waardering van wat in de empirie aan verbanden en processen is gevonden. Dat laatste is meer beleidsmatig van karakter. De wereld van het onderzoek en die van het beleid verschillen in hoge mate: de mensen en hun belangstelling verschillen, hun type vragen verschillen, hun belangen en loyaliteiten verschillen, hun tijdsperspectieven verschillen. In een wetsevaluatie komen beide werelden bij elkaar, de spanning blijft. Ik maak daarom dat onderscheid tussen evaluatieonderzoek (systematische verzameling van en analyse van gegevens) en de waardering van de uitkomsten van dat onderzoek. Voor dit raamwerk is primair het evaluatieonderzoek zelf aan de orde.

Een tweede reden waarom de definitie niet scherp kan zijn, is dat in ons bestel wet en beleid onlosmakelijk met elkaar zijn verbonden. Een wet is het juridisch voertuig van beleid. Er zijn de

nodige beleidsevaluaties; in geval van een wetsevaluatie wordt expliciet gerefereerd aan de wet waarin het beleid is neergelegd.

-Vanwege het overzicht en de leesbaarheid zijn de zes onderwerpen onderscheiden, de genoemde 'waarom', 'wat', 'wanneer', 'wie', 'waar' en 'hoe'. Deze onderwerpen staan echter niet onafhankelijk van elkaar: het 'wat' heeft direct gevolgen voor het 'wanneer', het 'wie' en het 'hoe'. Omgekeerd, heeft het 'wanneer' direct gevolgen voor het 'wat', en dus ook voor het 'wie' en het 'hoe'. Het was niet altijd te vermijden om aan die samenhangen te refereren. Overigens bevat het 'hoe' geen handboek methodologie, maar geeft het enig inzicht in enkele methodologische kwesties die bij beslissingen rondom wetsevaluaties (mogelijkheden en beperkingen) een rol kunnen spelen.

-Vanwege alle gedoe in 2013 rond het verwarrende en eigenlijk onmogelijke begrip 'zelfplagiat': in deze bijdrage wordt voortgebouwd op eerder gepubliceerd werk van mijn hand ('Over wetgeving', 'Wetgeving met Beleid', 'Een empathische wetgever', 'Negen aanwijzingen voor wetsevaluatief onderzoek', 'Voorbij de horizon- over experimenteer- en horizonbepalingen', alle geschreven met medewerking van of zelfs in nauwe samenwerking met een of enkele anderen, hoewel het geen twijfel lijdt dat ik steeds de eerste auteur was), en in het bijzonder op enkele lezingen die ik hierover hield voor uiteenlopende internationale gezelschappen (Chinezen verbonden aan het Volkscongres, juristen uit de Maghreb en uit Indonesië), onder steeds dezelfde titel: 'Coping with Uncertainty – a Framework for the Evaluation of Legislation'). Deze bijdrage lijkt soms zelfs verrassend veel op een vertaling van die in het Engels gehouden lezingen. Het is maar dat u het weet.

2. Waarom wetsevaluatie?

2.1 Motieven op micro niveau

De wetgever maakt wetten, regels. Deze regels moeten het maatschappelijk leven, de samenleving, ordenen. Ordenen wil zeggen: in banen leiden. Daartoe scheidt de wetgever instellingen en instituties, biedt hij rechten en voorzieningen, legt hij ge- en verboden op aan burgers in hun diverse hoedanigheden al dan niet voorzien van een sanctie voor het geval van overtreding of legt hij procedures vast. Zo kan men weten men waar men aan toe is. Die veiligheid en die (rechts)zekerheid vormen een cruciale voorwaarde voor samen leven en een samenleving. De paradox is dat niet op voorhand is te zeggen of die regels ook zullen werken, het beoogde doel bereiken. Er zijn verschillende redenen waarom die onzekerheid bij de voorbereiding van wetten en regels bestaat. In de eerste plaats zal een wet in de toekomst werken, als de wettelijke procedure is afgerond en de wet is afgekondigd. De toekomst is relatief onzeker. Dat is op twee niveaus het geval. De toekomst van het veld waarop de wet van toepassing is, is onzeker; maar breder ook kan er van alles gebeuren dat van invloed blijkt te zijn op dat veld. Wetten worden gemaakt onder een 'ceteris paribus' conditie maar de omstandigheden blijven niet gelijk in de tijd. Daarbij speelt, in de tweede plaats, een rol dat mensen van vlees en bloed de wet tot leven moeten brengen en telkens is weer onzeker hoe zij zich zullen gedragen. We weten inmiddels dat in Nederland de kans redelijk is dat zij zich aan de regels zullen houden (Veerman 2013), maar dat is een kans, geen natuurwet. In de derde plaats brengt de aard van de regels enige onzekerheid mee. Regels zijn algemeen, zijn relatief digitaal, terwijl de werkelijkheid waarin zij worden toegepast uit allerlei specifieke contexten en situaties bestaat; zij is analoog. Er moeten in de praktijk bij de toepassing van regels twee vertaalslagen worden gemaakt: van regel naar context maar ook omgekeerd van context naar regel.

Deze onzekerheden zijn als zodanig niet weg te nemen. Er zijn wel manieren om ermee om te gaan. Een manier is om te doen of die onzekerheden niet bestaan. In gewone menselijke situaties zou dat een gebrek aan realiteitsbesef betekenen, zou die ontkenning van de onzekerheden neurotisch kunnen worden genoemd. Voor de wetgever ligt dat waarschijnlijk iets anders. De wetgever zal niet graag zeggen dat hij ook niet weet of de gekozen aanpak van het probleem zal werken, dat deze wet een slag in de ruimte is of een gok. Dat overtuigt het parlement niet en de bevolking niet, en het mechanisme van de 'selffulfilling prophecy' kan dan zijn, in dit geval mogelijk desastreuze, werk doen. Dus bij de 'verkoop' van de voorgestelde wettelijke regels kan het verstandig zijn niet te koop te lopen met alle onzekerheden. Maar er zijn middelen om tot een zekere reductie van de onzekerheden te komen: prognostisch onderzoek om betere voorspellingen te doen (de ex ante evaluatie¹) enerzijds en evaluatie van de werking van wetgeving (de ex post evaluatie) anderzijds.

Dat is het eerste motief: een evaluatieonderzoek wordt uitgevoerd als de wetgever meer wil weten. Een vraag kan zijn wat het probleem nu eigenlijk is en hoe dit het beste valt aan te pakken. Of de wetgever wil een beeld krijgen van het veld waarin de wet haar werk moet doen of doet, of het doel

¹ Klein Haarhuis (2014) gebruikt de term *ex ante analyse* en daar geef ik op zich de voorkeur aan boven de term 'ex ante evaluatie', omdat het hier primair om onderzoek en analyse gaat. Ik zal mij vanwege het gemak (regelmatig gaat het in deze tekst om ex ante analyses en ex post evaluaties beide) echter regelmatig houden aan het gangbare taalgebruik.

kan worden bereikt of is bereikt, van de (mogelijke) effecten van een (voorgenomen) wet. Of als er enkele alternatieven zijn om een probleem aan te pakken, kan een soort kosten-batenanalyse mogelijk wat meer houvast geven. In dit verband: uit een onderzoek naar het gebruik van prognostische kosten-batenanalyses (Schrijvershof c.s. 2013) bleek dat de betreffende ambtenaren wel begrepen dat de resultaten van die analyses niet erg hard waren en waarschijnlijk zelfs onjuist zouden blijken te zijn achteraf², maar dat deze toch behulpzaam waren omdat ze een houvast gaven. In een meta-evaluatie naar ex ante analyses (Klein Haarhuis 2014) komt naar voren, dat deze in meerderheid worden gehouden met het oog op de besluitvorming. Het moment van ex ante evaluaties wijst op een behoefte aan kennis en reductie van onzekerheid: 85% van de ex ante evaluaties werd gehouden voor de fase van besluitvorming, waarbij gemiddeld ruim drie alternatieven werden bestudeerd. In een aantal gevallen is een ex ante evaluatie verplicht.

Als een wet eenmaal van kracht is, kan het zin hebben die wet te evalueren als deze niet zo goed (b)lijkt te werken, of als er zich uitvoeringsproblemen voordoen. Een evaluatie kan dan materiaal bieden om zo nodig tot aanpassing over te gaan.

Een algemenere reden is dat met wetsevaluatie ex ante en ex post de kwaliteit van de wetgeving (en het ermee verbonden beleid) kan worden verbeterd. Er zijn terugkerende thema's (bijvoorbeeld: wanneer en hoe werken subsidies), er zijn in de loop der tijd diverse kwaliteitseisen geformuleerd waartussen een spanning kan bestaan en door onderzoek kan meer inzicht daarin worden verkregen. Dan wil men ook weten en leren, maar treedt men buiten het gezichtsveld van een individuele wet. Zoals evaluatie kan leiden tot beter beleid en tot goedkoper beleid (Bakker 2012), zo kan evaluatie tevens bijdragen aan betere wetgeving.

Dus: een belangrijke reden, de belangrijkste technisch gezien, om ex ante of ex post evaluaties te houden, is dat de instanties die de wetgever een gezicht geven, iets willen weten van (een) voorgenomen of van kracht zijnde wet(geving) of het reservoir van relevante kennis voor de voorbereiding van wetgeving willen vullen.

Een tweede motief voor wetsevaluatie, in het bijzonder ex post evaluatie, kan een strategische zijn.³ Indien er politiek gesproken aarzelingen zijn of een wet nodig is of dat een voorgenomen wet effectief zal zijn (haar doel zal bereiken), kan de afspraak die wet te eniger tijd te evalueren de aarzelaars over de brug halen. Iets dergelijks gebeurde, hoewel strikt genomen geen wetsevaluatie, in 2013 om het vijfpartijenakkoord over de woningmarkt mogelijk te maken door zo de cruciale stem van Eerste Kamerlid A. Duivesteijn te winnen (18/12/2013). Ook hier is overigens in feite onzekerheid over de werking de achterliggende reden. Een bijkomend strategisch voordeel, zowel voor een ex ante als een ex post evaluatie kan zijn dat het onderwerp een tijdje van tafel is, niet meer op de politieke agenda staat. Het nadeel kan zijn dat het terug komt op een ongewenst moment.

² Overigens vond Klein Haarhuis (2014) dat de relatie: voorspelling- conforme uitkomst een gemengd beeld gaf: soms wel, soms niet, deels wel, deels niet.

³ Een strategische inzet voor ex ante analyse kan nog zijn dat de opdrachtgever het veld of de uitvoering kan betrekken in de beleidsvorming om zo niet alleen kennis te krijgen maar ook draagvlak te creëren. De opdrachtgever dient wel alert te zijn op specifieke belangen die dan kunnen worden gearticuleerd.

2.2 Motieven op macro niveau

Leeuw heeft gewezen op een aantal algemene voordelen van evaluatie (waartegenover overigens ook de nodige kosten en risico's staan, zie hoofdstuk 3.1): meer transparantie, betere verantwoording, leren door organisaties, bijdrage aan een levende democratie (meer betrokkenen, scheppen van tegenspraak, discussie, dialoog, deliberatie), bijdrage aan vijvers van kennis (diverse evaluatiesystemen produceren kennis over evaluatie).

Het instrument 'wetsevaluatie' als instrument voor de omgang met onzekerheid is met name sinds de tachtiger jaren steeds vaker ingezet. Leeuw & de Jongste (2006) geven daarvoor enkele redenen:

- de scheiding tussen beleid en uitvoering ofwel het op afstand zetten van de uitvoering leidde tot een informatietekort bij de beleidsdirecties;
- in het heersende paradigma van het 'New Public Management' ging de aandacht uit naar de 'output' van het beleid;
- transparantie werd een belangrijk item;
- de doorgaande groei van regels leidde tot problemen;
- het mechanisme van 'spun-off' impliceert een 'autonome' groei (en in dat verband valt ook te wijzen op de toenemende beschikbaarheid van onderzoekers die zijn opgeleid in de toen nieuwe disciplines van bestuurskunde en rechtseconomie).

Met andere woorden: zowel aan de vraagkant als aan de aanbodkant van evaluatieonderzoek zijn ontwikkelingen aanwijsbaar die tot groei leidden. Of een keerpunt is bereikt, is moeilijk te zeggen. Enerzijds wordt de noodzaak van 'evidence based' (ook 'evidence based' beleid) benadrukt (zonder oog voor feiten en mechanismen bestaat het gevaar dat beleid 'blinde ideologie' is, aldus Slob & Staman (2012)), anderzijds hoorde ik onlangs de constatering van een hoge ambtenaar dat er inmiddels meer evaluaties zijn van de gezondheidswetgeving dan dat er wetten zijn op dat gebied, en dat die verhouding toch een tikje vreemd is: dat kan zo niet meer. Het is een uiting van wat in de literatuur wel 'death by evaluation' is genoemd, veroorzaakt door automatiseren in en routinisering van evaluatie (Leeuw 2010).

2.3 Enkele cijfers

Er zijn dus diverse redenen een wetsevaluatie te houden, maar voor de hand ligt het niet, en zeker niet bij de ex post evaluatie. Zoals gezegd, een wet en het daarmee verbonden beleid vormen de aanpak van een probleem en het valt dan niet makkelijk te zeggen dat deze aanpak een sprong in het duister is. De minister zal toch niet graag in het Parlement zeggen, dat de coalitie dit compromis wil maar dat hij slechts een vaag idee heeft van wat de effecten mogelijk zullen zijn. Het Parlement zal dat niet pikken, is mijn verwachting. Dit wordt ondersteund door de feiten: in ('slechts') zo'n 10 % van de jaarlijkse wetgevingsproductie wordt een evaluatiebepaling opgenomen. Daarnaast is er nog

in een aantal gevallen een toezegging een evaluatie te houden. Er zijn gemiddeld zo'n 18 wetsevaluaties per jaar.⁴ Het blijft een betrekkelijk gering percentage op het totaal van de jaarlijkse wetgevingsproductie.

Voor de ex ante evaluatie liggen de cijfers wat anders. Voor de voorbereiding van beleid worden, naar het zich laat aanzien, de nodige studies, kosten-batenanalyses, prognoses, 'pilots' en zo meer, uitgevoerd. Het zijn er vele; oudere schattingen komen op zo'n 1000 per jaar, maar dat was rijp en groen door elkaar. In genoemde meta evaluatie van ex ante analyses (Klein Haarhuis 2014) zijn voor de jaren 2005-2011 550 ex ante onderzoeken gevonden, waarvan er 306 voldeden aan haar definitie⁵, van 21 analyses in 2005 tot 69 in 2011. Waarschijnlijk is sprake van een onderrapportage vanwege problemen van vindbaarheid. Maar bij de stroom van beleid die dagelijks langs de kamerbankjes en burelen vliedt, is dit nog een overzichtelijk aantal. De onzekerheid bij de voorbereiding is weliswaar politiek meer aanvaardbaar dan in geval een kant en klare wet voorligt, maar andere mechanismen maken dat prognostische studies lang niet altijd worden gedaan. Dat kan te maken hebben met tijd en geld, maar ook met het idee van voorbereiders dat zij wel weten hoe het zit, dat in zogenoemde beleidsgemeenschappen tamelijk vaste opinies bestaan over wat de problemen zijn en welke aanpak het beste is.

⁴ Klein Haarhuis & Niemeijer 2008; Veerman 2013. Het is onmogelijk precies te zijn, bijvoorbeeld omdat de definitie van een 'wetsevaluatie' niet precies kan zijn (vanwege de verbondenheid van wet en beleid en de ruimte in het begrip 'evaluatie') en omdat de zoektocht naar wetsevaluaties een niet geheel zekere is.

⁵ Haar eisen waren: gevraagd door centrale overheid(sorganen), mbt consequenties van beleidsopties, gepubliceerd. Andere vormen zoals pilots, ramingen, probleemverkenningen of visiestukken bleven buiten beschouwing.

3. Wat evalueren?

Het antwoord op de 'wat' vraag is uiteindelijk: wat de opdrachtgever, in het algemeen een departement, wil weten. Dat kan met de onzekerheid over de werking te maken hebben (vooraf bij de voorbereiding of achteraf als er 'geluiden uit het veld' komen), maar ook met de verwachting dat anderen met die opdracht kunnen instemmen in het geval dat een wetsevaluatie onderdeel uitmaakt van een onderhandelingsproces over de inhoud van een voorgenomen wet. Want ook rond een evaluatie bestaan verschillende belangen en invalshoeken.

In geval van onzekerheid over de werking, zal het de opdrachtgever lang niet altijd meevallen te bepalen wat hij wil weten. Het is vaak moeilijk te weten wat je wilt weten. Ik heb diverse malen meegemaakt dat de opdrachtgever 'wel alles' wilde weten of juist iets vaags als 'iets met het oog op voorlichting of zo', en dan later niet wist wat met de opbrengst van het onderzoek te doen. En waarschijnlijk geldt de regel dat hoe minder je weet wat je wilt krijgen, des te minder je weet wat je zult krijgen. Ook om die reden zal ik specifieker zijn en een aantal onderwerpen noemen die bij een wetsevaluatie aan de orde kunnen komen. Ik maak daarbij wederom een onderscheid naar ex ante en ex post evaluatie. Bij ex post evaluatie kan men vervolgens een onderscheid maken naar onderwerpen die bij evaluatie van een concrete wet aan de orde kunnen komen enerzijds en anderzijds thema's die worden onderzocht.

3.1 Onderwerpen voor ex ante evaluatie

Het Integraal Afwegingskader, dat, zo heeft de Ministerraad besloten⁶, bij de ontwikkeling van alle beleid en wetgeving moet worden betrokken, bevat een zevental vragen. Deze moeten dus worden beantwoord met het oog op voorgenomen beleid en wetgeving. Deze zeven vragen zijn:

- Wat is de aanleiding?
- Wie zijn betrokken?
- Hoe wordt het probleem gedefinieerd?
- Wat is het doel van de overheidsinterventie?
- Wat rechtvaardigt de interventie?
- Welke instrumenten zijn geschikt (met aandacht voor uitvoerbaarheid, doelmatigheid en overgangssituaties)?
- Wat zijn de gevolgen (voor de samenleving, voor burgers en bedrijven, voor de overheid, voor het milieu)?

De vragen zijn op zich helder; ze zijn nader uitgewerkt in subvragen. Hoewel voor een deel van de vragen het antwoord primair normatief of politiek van aard is, voor de beantwoording van alle vragen geldt dat empirische onderbouwing behulpzaam of nodig kan zijn. En dat kan leiden tot ex ante evaluatie. Op een aantal van deze vragen ga ik in, vanuit een onderzoeksmatige invalshoek.

⁶ Brief van 14 april 2014, Kamerstukken II 2010/11, 29515, nr. 330. Er staat geen sanctie op overtreding van deze plicht. Als het Parlement stelselmatig gaat vragen naar de antwoorden op die zeven vragen, bestaat de kans dat de antwoorden ook stelselmatig in de memories van toelichting terechtkomen. Of de antwoorden aanvaardbaar zijn, moet dan nog wel worden besproken.

a. Doel en middelen: de beleidstheorie

Globaal gesproken staat de beleidsmaker en met deze de wetgevingsjurist, voor de opdracht een beleidstheorie te ontwerpen. Dat wil zeggen: een idee omtrent het wenselijke doel en de middelen die nodig zijn dat doel te bereiken. Het belang van een goede beleidstheorie is niet te onderschatten. Ik geef een eenvoudig, aan de praktijk ontleend, voorbeeld van de uitwerking van een foute beleidstheorie. In een Franse winkel stond een Nederlandse toerist die kaas wilde kopen en dit kenbaar maakte door het woord 'kaas' met een Franse intonatie uit te spreken. De winkelier keek hem niet begrijpend aan. De man zei hetzelfde op een wat luider toon en toen ook dat geen begrip opleverde, schreeuwde hij het. Hij dacht kennelijk dat iets dat op Frans lijkt ook Frans is en dat de winkelier hem niet verstond omdat deze doof zou zijn. Zijn probleemanalyse en de daarop gebaseerde beleidstheorie klopten niet; en daarom zou hij nimmer kaas hebben kunnen kopen als niet iemand die een andere analyse van het probleem had gemaakt en het Franse woord voor 'kaas' kende, hem had geholpen.

In dit voorbeeld is het doel erg helder en eenvoudig: de aanschaf van een stuk kaas. In de politieke werkelijkheid van beleid en wetgeving is het formuleren van een doel lastiger en daarom zal de neiging bestaan een doel globaler of vager te omschrijven. Uit onderzoek blijkt dat de kans dat een beleid doeltreffend is, toeneemt naarmate het doel preciezer is gedefinieerd (Veerman 2013). Maar dat is niet voldoende. Als het instrumentarium niet geschikt is, zal het doel hoogstwaarschijnlijk niet worden bereikt.

Ex ante onderzoek kan er mede op zijn gericht de beleidstheorie op haar aannemelijkheid te beoordelen. De juistheid ervan kan in zekere zin pas achteraf worden vastgesteld, maar vooraf kan wel een indruk worden verkregen op basis van beschikbare kennis. Een goede beleidstheorie hangt vanzelfsprekend samen met de probleemanalyse en het onderzoek naar het instrumentarium, maar ook kan op basis van algemene sociaalwetenschappelijke inzichten de waarschijnlijkheid van de theorie worden geschat. Klein Haarhuis kwam dit tegen in de vorm van een planevaluatie, hetgeen overigens zelden voorkwam (1 á 2%).

b. Probleemanalyse

Naar het mij voorkomt is de probleemanalyse een van de belangrijkste onderwerpen van een ex ante evaluatie, zonet het belangrijkste onderwerp. Daarvoor heb ik twee redenen. De eerste is dat naarmate er meer duidelijkheid bestaat over wat het probleem is, de kans op een behoorlijke aanpak ervan groter is. In veel gevallen is het probleem immers niet duidelijk of bestaan er tenminste meer opvattingen over wat het probleem is. Ik heb al vaker het voorbeeld van de prostitutie gegeven, waarbij men het probleem, als men prostitutie als probleem wil zien, op pakweg zes manieren kan beschouwen. De onderscheiden visies leiden ook tot een verschillende aanpak (gericht op bijvoorbeeld: de hoerenloper, de hoerenleverancier, de hoer zelf, de souteneur, de bordeelhouder, de lokale omgeving (verkeerscirculatie, bestemmingsplan)). De opvatting over het probleem kan een bestuurlijke, economische, criminologische of levensbeschouwelijke inslag hebben, bij de

probleemformulering kunnen diverse belangen zijn betrokken. Dus voordat de voorbereiders van beleid en wetgeving mogelijke effecten van beleidsmatige maatregelen laten uitzoeken, is het verstandig eerst te bezien wat het probleem is c.q. hoe het probleem wordt gepercipieerd door de verschillende betrokkenen. In dit verband wijs ik erop dat beleidsmakers makkelijk vast zitten in percepties die zij delen met die partijen uit het veld met wie zijn betrekkingen onderhouden en van wie zij voor hun informatie afhankelijk zijn. Dat kan een flinke bias opleveren.

Een tweede reden om het probleem goed te analyseren is dat daar in elk geval empirisch onderzoek naar is te doen. Het behoeft dus niet een postmodern relativisme te zijn, maar een feitelijk te onderzoeken vraag: wat is het probleem voor wie en welke belangen zijn er bij betrokken. Interessant is waarschijnlijk te achterhalen wie het probleem om welke reden op de beleidsagenda heeft weten te zetten.

Bij de probleemanalyse vallen nog enkele kanttekeningen te maken.

-De eerste is dat voor de doelbepaling van beleid en wetgeving een probleemanalyse nodig kan zijn, omdat de oplossing van een probleem het doel kan zijn. Echter, inzicht in de oorzaak van een probleem leidt, zoals gezegd, niet één op één tot een aanpak. Soms is de oorzaak niet aan te pakken en zal de beleidsmaker genoeg moeten nemen met middelen die de symptomen verzachten; soms zullen maatregelen die de oorzaak moeten wegnemen tot zodanige gevolgen leiden dat onwenselijke neveneffecten de overhand krijgen. In feite kan men makkelijker vanuit het doel terugredeneren naar welke maatregelen het gewenste doel zullen helpen bereiken. Het is het genoemde verschil tussen causaal denken en finaal denken.

-De tweede is dat de probleemformulering niet neutraal is. De manier waarop een probleem wordt gedefinieerd, kan beslissend zijn voor een succesvolle aanpak. Een voorbeeld uit mijn ambtelijk leven is de aanpak van de illegale naaiateliers. Toen dit probleem werd geduid in termen van een vreemdelingenbeleid, kwam de aanpak niet van de grond omdat grote politieke verdeeldheid bestond over de vraag of men arme vreemdelingen het brood uit de mond mocht stoten. Toen het werd geformuleerd als een fraudeprobleem en de aandacht werd verlegd naar de opdrachtgevers en hun ontduiking van belasting, arbeidsregels en energielasten, kwam de bestrijding op gang omdat fraudebestrijding moreel aanvaardbaar was, relatief technisch en omdat er een infrastructuur voor was. De probleemformulering was, met andere woorden, een kwestie van 'framing'.

-De derde is dat de omvang van het probleem een belangrijk element kan zijn. "De minimis non curat praetor", zeiden de oude Romeinen al. Het is dus zaak te bezien of sprake is van een of enkele incidenten dan wel van een meer structureel probleem. Incidenten kunnen niet 'niet aangepakt worden', maar de aanpak daarvan zal eerder een kwestie van 'damage control' zijn, terwijl structurele of omvangrijke problemen een daarop toegesneden beleid met de bijbehorende wetgeving zullen vergen.

c. Instrumentarium

Naast het doel maakt het instrumentarium deel uit van de beleidstheorie: welke middelen moeten worden ingezet om het beoogde doel te bereiken. Dat vergt inzicht in directe processen en

verbanden, en vervolgens in de financiële randvoorwaarden. Maar dat niet alleen. In geval van beleid en wetgeving spelen veelal diverse actoren een rol: uitvoerders, toezichthouders en handhavers, burgers. Zij moeten om een wet te kunnen uitvoeren, handhaven en naleven 'weten, willen en kunnen' zoals de mantra van de werkbaarheidsanalyse van het CVV het formuleert. Onderzocht kan daarom (ook) worden of en hoe de bekendheid met de wet kan worden bevorderd (wat is de doelgroep en hoe geïnformeerd is deze, welke voorlichting is nodig, welke intermediaire instanties geven voorlichting (zoals accountants, bedrijfsverenigingen, notarissen)) en welke materiële condities moeten worden vervuld om het 'willen' en het 'kunnen' te bevorderen. Bij 'materiële condities' gaat het niet alleen om geld, maar ook om de beschikbaarheid van voorzieningen, een infrastructuur, mensen, deskundigheden, tijd. Gebrek daaraan leveren knelpunten op, schreef Swanborn in 1999, 'tijd, geld, personeel zijn drie cruciale condities voor de uitvoering van beleid', zei de voorzitter van de Algemene Rekenkamer in 2007 in een voordracht, en in mijn onderzoek bleek eveneens dat verschaffing van materiële middelen bijdraagt aan de doelbereiking (Veerman 2013). In dat verband kan het raadzaam zijn veldonderzoek bij de doelgroep(en) te doen⁷, systematisch de processen en stappen die bij de werking van het beoogde beleid en de daarbij behorende wet aan de orde zijn, te beschrijven. Men kan dat statisch beschrijven, maar ook met oog voor de spanningen en dilemma's die een nieuwe wet oplevert bij degenen die de wet moeten uitvoeren, handhaven of naleven. De Tafel van Elf, de voorganger van de werkbaarheidsanalyse, onderscheidde een aantal dimensies om de mate van naleving bij de doelgroep voor welke de wettelijke regeling was bedoeld, te kunnen schatten: bekendheid met de regeling, kosten en baten van naleving (en van overtreding) waarop de mate van aanvaardbaarheid van de regeling van invloed kan zijn alsmede de gezagsgetrouwheid van de groep en de informele controle binnen die groep, de externe controle in termen van controlekans en pakkans alsmede van sanctie (kans op een sanctie, ernst van de sanctie). Een goede naleving is een voorwaarde voor doelbereiking. Niet de enige, de beleidstheorie moet bijvoorbeeld ook kloppen.

Klein Haarhuis zag in haar meta onderzoek flink wat uitvoeringsanalyses (11%) en in beperkte mate knelpuntenanalyses (1%).

d. Kosten en baten van het voorgenomen beleid.

Ook dit onderwerp kwam hierboven al ter sprake, als voorwaarde voor de effectiviteit van beleid en wetgeving. Er zijn de nodige vormen van kosten-batenanalyse, grofweg onder te verdelen in bedrijfseconomische en maatschappelijke kba's. In kba's worden effecten van maatregelen of projecten in geld uitgedrukt. Bedacht dient te worden a. welke kosten als kosten zullen worden meegeteld, en b. dat, zeker bij maatschappelijke kba's, de baten vaak moeilijker zijn vast te stellen dan de kosten. Men kan baten (zoals 'een veilige leefomgeving' of een 'adequaat levensjaar') in geld uitdrukken (€ 80.000,- voor een adequaat levensjaar), maar daarin zit de nodige willekeur. Ook voor sommige kosten (zoals vernietiging van sociale netwerken) geldt dat zij niet zomaar zijn te berekenen. Niettemin kan het prettig zijn die financiële waardering uit te voeren om een baken te

⁷ Overigens geldt dat meer algemeen dat het bij het maken van wetgeving en beleid verstandig is het veld waarin de wet later haar werk moet doen, te kennen en te begrijpen wat een wet voor de mensen van dat veld meebrengt. 'Een effectieve wetgever is een empathische wetgever'. (Veerman 2013)

hebben, zodat een koers valt uit te zetten. Ex ante kunnen kba's (of de verwante kosteneffectiviteitsanalyse) worden gebruikt om diverse beleids- en wetgevingsalternatieven te verhelderen, het aantal zo mogelijk te reduceren en te rangschikken. Klein Haarhuis vond dat minstens 16% van het ex ante onderzoek een (m)kba inhield. In een flink aantal gevallen waren analyses voor specifieke effecten uitgevoerd (14%), zoals milieueffecten. Daarnaast trof zij in zeer beperkte mate aan: risicoanalyses en kosteneffectiviteitsanalyses (vergelijking van alternatieven zonder nuloptie; effecten niet in geld uitgedrukt).

Globaal betrof de helft van de ex ante analyses de uitvoering en specifieke effecten van een voorgenomen wet, en de andere helft of de doelen zouden worden bereikt.

Intermezzo: het moment

Wat wordt onderzocht, hangt bij ex ante evaluatie mede af van het moment waarop wordt onderzocht. In een vroeg stadium kunnen alternatieven worden geformuleerd en gezien op kansen en kosten. Indien in een later stadium al een keuze in instrumentarium is gemaakt, kunnen nog slechts de mogelijke effecten van die keuze worden onderzocht; hoogstens is het dan zinvol om het nulalternatief nog uit te diepen, dat wil zeggen: voorspellen wat de effecten zijn als het voorgenomen instrumentarium niet wordt ingezet. In het algemeen is het raadzaam een dergelijke 'counterfactual' uit te voeren.

Het is helder dat onderzoek naar verschillende mogelijkheden tijdrovender en kostbaarder is dan onderzoek naar de mogelijke effecten van één alternatief. Een regelmatig gebruikt alternatief is om verkenningen of quick scans uit te voeren (Klein Haarhuis 2014).

e. Juridisch onderzoek

Ex ante kan vanzelfsprekend ook een juridisch onderzoek worden gedaan, bijvoorbeeld naar wat de juridische randvoorwaarden zijn voor een voorgenomen wettelijke regeling. Veelal bestaat op de departementen ruime expertise daarin. Maar toch. In het bijzonder kan het opsporen van Europese of internationale randvoorwaarden bijzondere expertise vergen. In sommige gevallen kan de onzekerheid worden gereduceerd door rechtsvergelijkend onderzoek te doen. Daarbij past de waarschuwing dat alleen rechtssystemen worden vergeleken en dat wordt vergeten dat regels betekenis krijgen in de praktijk.

Bij dit type onderzoek hoeft overigens niet alleen aan wettelijke maatregelen te worden gedacht maar kunnen ook alternatieven voor wetgeving worden overwogen, zoals convenanten. In het algemeen heeft de overheid wel een juridisch instrumentarium van node als legitieme basis van haar optreden, maar dat hoeft niet altijd in de vorm van een wet.

3.2 Onderwerpen voor ex post evaluatie

Globaal kan op twee momenten het besluit worden genomen een ex post evaluatie te houden: bij het voorbereiden van een wet (voor de zekerheid, als compromis) of wanneer de wet in werking is en

er kritische geluiden over de werking komen. Dat verschil in moment kan gevolgen hebben voor het wat en voor het hoe. Achteraf kan bijvoorbeeld geen nul-meting worden gehouden, hetgeen de mogelijkheid van het vinden van causale verbanden drastisch verkleint.

De diverse mogelijke onderwerpen passeren hieronder de revue.

a. Gelding (juridische evaluatie)

Men kan de evaluatie beperken tot de juridische aspecten, zoals de relatie tot andere wetgeving of de conformiteit met hogere regelgeving. Men kan de jurisprudentie daarin betrekken en bezien of er voldoende eenduidigheid is in de rechterlijke uitspraken (meerduidigheid hoeft overigens niet een indicatie te zijn van een onduidelijke tekst), of de rechter is ingeschakeld om normen die open zijn geformuleerd in de wettekst in concreto in te vullen. Men kan dergelijke onderzoek aanvullen met sociaalwetenschappelijk onderzoek door rechters of uitvoerders en handhavers te vragen of zij met de tekst uit de voeten kunnen, welke knelpunten zij tegenkomen, of bezien in hoeverre rechterlijke uitspraken doorwerken in de uitvoeringspraktijk. In de evaluaties die ik bestudeerde, kwam ik die juridische en sociaalwetenschappelijke aspecten van de rechterlijke bemoeienis weinig tegen.

Wel dient bedacht te worden dat de jurisprudentie een zeer bijzondere uitsnede is uit de werkelijkheid en op zich weinig tot niets zegt over de werking van een wet of over de knelpunten daarbij. Het begint er al mee dat alleen conflicten voor de rechter komen en vervolgens dat lang niet alle conflicten voor de rechter komen.

b. Loutere doelbereiking

De onderzoeksopdracht kan zijn te bezien of het beoogde doel is bereikt. Duidelijk is dat naarmate het doel vager is, het ook moeilijker zal zijn te bepalen of het doel is bereikt. De vaagheid van het doel kan een contra-indicatie zijn voor een dergelijke opdracht of er moet worden afgesproken welke *indicatoren* worden gebruikt om het bereiken van vage doelen aan op te hangen. Aan deze opdracht zitten voorts enkele problemen vast.

-Deze opdracht heeft het nadeel dat weliswaar het doel kan zijn bereikt, maar dat daar andere redenen voor waren dan de wettelijke interventie. Omdat vrijwel nooit het beoogde doel voor 100% wordt bereikt en de vraag kan zijn of voortzetting van het beleid of intensivering van het beleid nodig is, staat de opdrachtgever in feite met lege handen indien hij niet weet of de wet tot de doelbereiking leidde. Van Gestel gaf als voorbeeld de verplichting tot het dragen van een brommerhelm met het oog op minder en minder ernstige brommerongelukken. In absolute zin werd het doel in redelijke mate gehaald maar de reden was dat er minder brommer werd gereden vanwege de helm. Relatief waren er meer ongelukken omdat zowel automobilisten als brommerrijders meer risico namen. Op zich kan de opdrachtgever tevreden zijn, hij kan evenwel geen beleidsbeslissing baseren op het positieve resultaat (omdat dus de beleidstheorie niet was onderzocht).

-Een oplossing moet worden gevonden voor de vraag wat meetelt bij de doelbereiking. Indien een wet wordt gemaakt met het doel wachttijden te bekorten, en het resultaat is dat de wachttijden zijn bekort maar de behandeltijden langer zijn geworden (een verplaatsingseffect), is het de vraag of het doel is bereikt of toch niet zo.

-In het verlengde daarvan: de beperking tot het door de wetgever beoogde doel maakt dat andere effecten buiten beschouwing blijven. Indien het doel wordt bereikt maar de neveneffecten zijn negatief, is het de vraag of het beleid in die bestaande vorm moet blijven bestaan, zeker indien de kosten van de neveneffecten hoog zijn. Levitt & Dubner geven het voorbeeld van de strengere beveiliging van de luchthavens in de Verenigde Staten. Het neveneffect daarvan was onder meer dat het binnenlands vluchtverkeer afnam en het autoverkeer toe. Het gevolg was een toename van het aantal ongelukken, ook van ongelukken met dodelijke afloop. Die toename was groter dan de geschatte kans op doden bij een terroristische aanslag op een vliegtuig. Andere neveneffecten van die beveiliging waren: minder wetenschappelijke activiteiten en minder influenza.

-Van belang is het moment waarop de doelbereiking wordt gemeten. Indien dat kort na de inwerkingtreding is, is de kans groot dat het doel nog niet is bereikt. Dat kan beleidsmatig prettig zijn omdat falen als kinderziekte kan worden bestempeld (hetgeen juist kan zijn, maar ook een vorm van (zelf)bedrog). Na een lange periode bestaat de kans dat andere invloeden zo groot zijn geworden of negatieve effecten zo doordringend, dat de doelbereiking uit zicht is geraakt. Beleidsmatig gezien, geeft een hele lang termijn waarschijnlijk weinig relevante informatie.

-Bedenk dat wetgeving meestal betekent verdeling en herverdeling van middelen en dat er dus winnaars en verliezers zijn. De doelbereiking kan verschillen voor onderscheiden groepen.

-Niet altijd wordt in een memorie van toelichting het doel van een regeling geformuleerd. Deze zal achterhaald moeten worden, want bepaling van doelbereiking zonder dat het doel bekend is, is onmogelijk.

-Indien het doel is geformuleerd in termen van een vergelijking (sneller dan vroeger, bijvoorbeeld), dan is voor de bepaling van de doelbereiking dus ook een meting op die twee punten nodig.

c. Effecten

De opdrachtgever kan onderzoek naar bereiking van het doel uitbreiden naar alle effecten van de wet, de positieve en de negatieve. Deze opdracht wordt wel gepropageerd om onderzoekers de vrije ruimte te geven en de riskante reductie tot doelbereiking te vermijden. Zoals gezegd, kan die beperking een onjuist beeld geven.

Het probleem bij onderzoek naar effecten is tweërlei. Ten eerste moet de causale relatie tussen wet en effect kunnen worden gelegd. Dat valt meestal niet mee; dat is veelal een probleem bij het doelbereikingsonderzoek en daar weet de onderzoeker nog waar hij moet kijken. In het verlengde daarvan speelt ten tweede de vraag waar de onderzoeker moet kijken om mogelijke effecten te traceren? Het is als zoeken naar spelden in een hooiberg terwijl je niet precies weet wat een speld is. Dus een discussie over waar men welke effecten zou kunnen vermoeden, is bij zo'n opdracht zeer

nodig; de tijd tussen wet en effect speelt daarbij een rol. En ook bij de effecten kan het nuttig zijn naar verschillende groepen te kijken.

d. Doeltreffendheid

In geval van een doeltreffendheidsonderzoek wordt gezien of het al dan niet bereiken van het beoogde doel ook te danken is aan de wet en de daarbij ingezette instrumenten. Het is de gebruikelijke opdracht, geleid op de standaardformulering die de Aanwijzingen voor de regelgeving geven (Ar 164: “...een verslag over de doeltreffendheid en de effecten van deze wet in de praktijk...”). Het is ook een moeilijk realiseerbare opdracht. Alle problemen en aandachtspunten die aan de orde zijn bij doelbereikingsonderzoek bestaan hier in volle omvang en bovendien moet de onderzoeker ook aangeven of en hoe de wet en de daarbij ingezette instrumenten hun werk hebben gedaan. Deze kanttekening betekent dat men in zekere zin door de hond of door de kat wordt gebeten: met louter een doelbereikingsonderzoek weet de opdrachtgever niet wat hij moet doen als het doel niet of niet geheel is bereikt, maar een doeltreffendheidsonderzoek is wel extra ingewikkeld, levert zachte inzichten op.

Men kan het doeltreffendheidsonderzoek opzetten als een kosten-batenonderzoek. Dan geldt wat voor het kosten-batenonderzoek hierboven al is gezegd, dat moet worden bepaald welke kosten en welke baten in ogenschouw worden genomen, hoe die worden gewaardeerd. Bij ex post evaluatie bestaat het risico dat alleen wordt gekeken naar de kosten en baten die voortvloeien uit de wettelijke regels. Dat kan een blikvernauwing opleveren, bijvoorbeeld in de zin dat naar kosten en baten op microniveau wordt gezocht maar niet naar de effecten op macroniveau (welvaartseffecten), of juist omgekeerd dat in de analyse geabstraheerd wordt van kosten en baten op het microniveau van individuen (en individuen zijn potentiële stemmers).

Intermezzo: causaliteit

Volgens de methodoloog Swanborn is ‘het werkelijk verifiëren van een causaal proces nimmer mogelijk; ‘causaliteit’ blijft een werkhypothese van de onderzoeker.’ (1982:81) Hoogstens kan men correlaties vinden.

e. Uitvoeringsprocessen & mechanismen

In zo’n 80 % van de wetgeving wordt gebruik gemaakt van uitvoeringsorganen of is er sprake van toezichtsconstructies. Welke processen daar plaatsvinden, met welke dilemma’s uitvoerders en toezichthouders worden geconfronteerd en hoe zij daarmee omgaan, welke mechanismen erbij spelen, dit alles kan worden onderzocht. Daarbij kan men bijvoorbeeld letten op de verschillende boodschappen die de overheid afgeeft (wettelijke normen maar daarnaast financiële kaders en prestatie-eisen, informatie-eisen, prioriteitstelling bij verschillende taken, al dan niet voldoende tijd en middelen voor de invoering van de wet), de druk die sommige ‘klanten’ opleveren, de methoden die uitvoerders bedenken om te kunnen omgaan met druk van verschillende zijden of met werkdruk. Men kan die methoden als ‘coping mechanisms’ zien, zoals er ook andere menselijke mechanismen

bestaan, dat wil zeggen manieren waarop wordt omgegaan met impulsen zoals een wet. In de literatuur is veel te halen over hoe mensen omgaan met informatie, bijvoorbeeld, welke heuristieken zij toepassen om zo snel en gemakkelijk mogelijk een beslissing te nemen.

Op zich wordt met een onderzoek naar uitvoeringsprocessen en mechanismen de doelbereiking niet onderzocht, maar naar opvattingen daarover kan worden gevraagd in een procesonderzoek en aan de hand van de uitvoeringsprocessen valt mogelijk een indruk te krijgen van de mate van doelbereiking en van neveneffecten. Uitvoerders kunnen daar zicht op hebben. Onderzoek naar uitvoeringsprocessen en mechanismen kan vanzelfsprekend onderdeel uitmaken van een doeltreffendheidsonderzoek.

Dergelijk onderzoek kan eveneens worden uitgevoerd naar processen en mechanismen die plaatsvinden bij handhaving en naleving.

Ik noem in dit verband ook de implementatieprocessen als object van onderzoek. Als een wet het leven ziet en haar geboorte via het Staatsblad bekend is gemaakt, zal deze gepaard kunnen gaan met diverse invoerings- of implementatieprocessen. Het gaat dan om de eerder genoemde materiële condities die worden nodig geacht om wet en beleid mogelijk te maken, omdat het woord van de wet alleen veelal niet voldoende is. Er moeten organisaties voor worden ingericht of veranderd, gebouwen betrokken, informatiesystemen ontwikkeld, deskundigheden verkregen en personeel aangenomen, uitvoeringsprocessen geprotocolleerd en zo meer. Onderzocht kan worden of en hoe dat is gegaan, of, indien de evaluatie in de vorm van een monitor geschiedt, hoe dat gaat.

f. Knelpunten

Men kan, zeker indien er weinig ruimte is om van beleid te veranderen, kiezen voor een knelpuntenonderzoek. Dat zal in veel opzichten lijken op onderzoek naar uitvoeringsprocessen. Ook hier kunnen knelpunten verschillen voor onderscheiden groepen.

g. Vergelijking ex ante – ex post

Zoals gezegd, dient men voor een evaluatieonderzoek veelal de beleidstheorie te achterhalen: waarom dacht de wetgever met deze middelen het beoogde doel te bereiken. Indien bij de voorbereiding een ex ante evaluatie is gehouden, kan in een onderzoek worden nagegaan of en in welke mate de uitkomsten van het prognostisch onderzoek zijn bevestigd. Omdat op sommige beleidsterreinen beleid en wetgeving in verandering zijn, kan een dergelijke vergelijking een verrijking van inzicht betekenen. Het gebeurt, voorzover bekend, weinig. Klein Haarhuis (2014) vond dat in 85% van de ex ante analyses niet een ex post evaluatie aan de orde was. Redenen die zij vond waren onder meer dat het beleid van de baan was, dat het beleid nog in een ontwerpfase was, dat de tijd voor ex post evaluatie nog niet rijp was. Van de 15% waarin wel iets als een ex post evaluatie

was uitgevoerd, ging dat in meer dan de helft van de gevallen om een 'ex durante evaluatie', een vinger aan de pols; die evaluaties sloten in het algemeen niet aan op de ex ante analyse.

h. Thematische onderwerpen

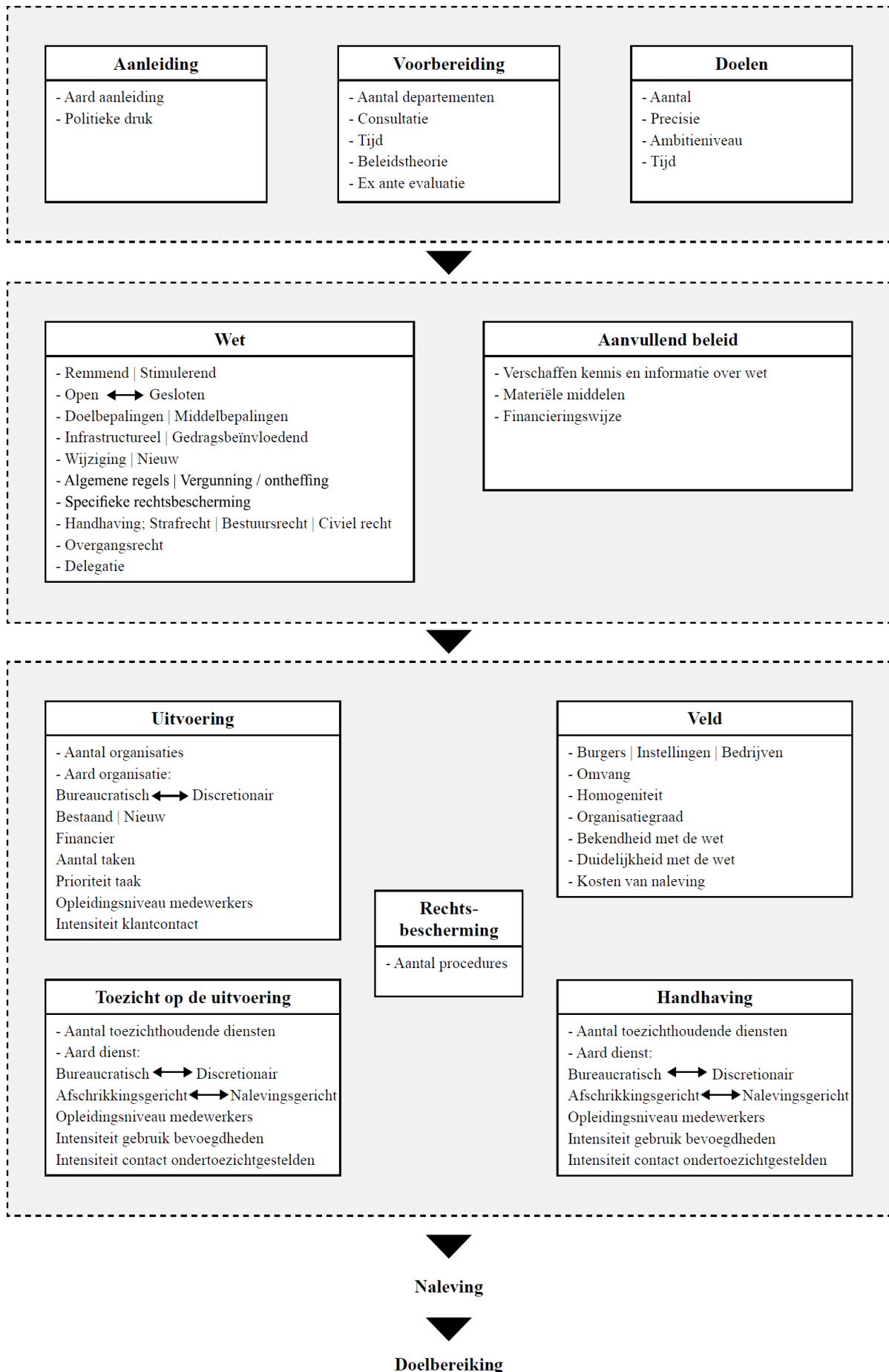
Het kan om diverse redenen (politieke gevoeligheid, beschikbare tijd, onderzoekbaarheid) wijs zijn evaluatie-onderzoek niet toe te spitsen op één wet, maar om thematisch een onderwerp te onderzoeken. Dat kan gebeuren door meer regelingen te onderzoeken of bij voorgenomen evaluaties een aantal vaste vragen mee te nemen. Onderwerpen kunnen bijvoorbeeld zijn: de effectiviteit van subsidies dan wel de effectiviteit van subsidies vergeleken met fiscale stimulansen, voorwaarden voor een goede invoering van wetten, omgang met en effecten van prestatie-eisen, gevolgen van verschillende financieringsregimes, gevolgen van verschillende handhavingsregimes, voor- en nadelen van systeemtoezicht, de rol van de rechter bij de uitvoering van wetten. In zekere zin gaat men in dergelijk onderzoek op zoek naar processen en mechanismen.

3.3 Ten slotte

Vanuit het departement of de politiek kunnen veel vragen over de werking van wetgeving worden gesteld. In het bijgevoegde schema waarbij de doelbereiking de afhankelijke variabele is, vindt men diverse factoren vermeld in de verschillende elementen waaruit het schema bestaat. Het schema is gebruikt over de meta-analyse van wetsevaluaties (Veerman & Mulder 2010, Veerman 2013). De elementen en factoren, de relaties tussen die elementen en die factoren kunnen empirisch worden onderzocht. Wat verstandig is in deze, hangt van de concrete situatie af en wordt mede bepaald door hierna nog te noemen overwegingen. Dat is een waarheid als een koe. Ik noem nog twee valkuilen.

Weten wat je wilt is moeilijk, een hele globale opdracht geven en de invulling verder overlaten aan de onderzoekers is een stuk makkelijker. Die handelwijze heeft een prijs. De teleurstelling achteraf kan groot zijn (want achteraf weet men beter wat men wilde en wil weten), daardoor zal het gebruik van de resultaten kleiner zijn en kan sprake zijn van weggegooid geld (tenzij de resultaten aan het reservoir van nuttige kennis kunnen worden toegevoegd). Het is belangrijk dat beleid en wetgeving enerzijds en de onderzoekers anderzijds overleggen over wat wordt onderzocht. Het betekent ook dat de opdrachtgever tijdens het onderzoek contact houdt met de onderzoekers over het wel en wee van het onderzoek.

Een specifieke onderzoeksopdracht biedt minder ruimte aan de onderzoekers. Vanuit die kant kan op uitbreiding worden aangestuurd omdat een te nauw perspectief niet een goed beeld van de werking van wetgeving oplevert. Dat is niet noodzakelijk vanuit een bedrijfsbelang van de onderzoeker gedacht, maar kan zeker een waardevolle aanvulling zijn. In het bovenstaande is daarop al gewezen.



4. Wanneer wetsevaluatie?

Een eenvoudige logica brengt mee dat als bij de voorbereiding van wetgeving de wetgever niet zeker weet of en hoe een wet zal werken, een ex ante en ex post evaluatie immer nuttig zijn. Dat is echter niet zo. Er kunnen diverse omstandigheden zijn die maken dat een evaluatie niet nuttig of nodig is. In dit hoofdstuk ga ik in op die condities (4.1).

De ‘wanneer’- vraag kan ook wat anders worden begrepen, namelijk of sommige wetgeving of sommige bepalingen beter zijn te evalueren dan andere. De aard van de wetgeving kan een indicator zijn voor de beslissing al dan niet te evalueren. Dat komt in 4.2 aan de orde.

4.1 Condities voor wetsevaluatie

Ik onderscheid vier elementen die een rol kunnen spelen bij de beslissing al dan niet een evaluatie te houden, in het bijzonder een ex post evaluatie. Die vier zijn:

- niveau van onzekerheid
- gebruik van de resultaten
- tijd en kosten
- onderzoeksmogelijkheden

Op deze vier elementen ga ik achtereenvolgens kort in. Daarnaast besteed ik aandacht aan de zogenoemde

- horizonbepaling.

ad onzekerheid

Zoals gezegd, zijn er diverse motieven om te besluiten een evaluatie te houden (overtuigen van tegenstanders, munitie voor een compromis), maar behoefte aan kennis en inzicht of reductie van onzekerheid is de meest voor de hand liggende. Je wilt ex ante een idee hebben of de beleidstheorie kan kloppen dan wel ex post of de wet haar doel bereikt (en de beleidstheorie klopte). In zekere zin kan men zeggen dat naarmate de onzekerheid groter is, de reden een evaluatie te houden navenant groter is. Een op een loopt die relatie niet. Er kunnen bijvoorbeeld politieke redenen zijn waarom evalueren juist meer of juist minder voor de hand ligt. Dat is bijvoorbeeld het geval als een evaluatie wordt ingezet in de onderhandelingen over de inhoud van een wetsvoorstel.

Soms zal de onzekerheid vooraf zodanig groot zijn dat het verstandig is een experiment te houden. In paragraaf 7.3 kom ik daarop terug.

ad gebruik

De opbrengst van een evaluatieonderzoek is kennis, kennis die bruikbaar moet zijn voor de te nemen beslissing, kennis die de onzekerheid kan reduceren. Het gebruik van kennis kan specifieke wetgeving

betreffen, maar ook meer algemeen zijn. Dat laatste is zeker het geval bij thematisch onderzoek. Daarmee wil men het reservoir van kennis vergroten, zodat daaruit geput kan worden als specifieke wetgeving daadwerkelijk wordt voorbereid. Het probleem zit meer bij het specifiek gebruik, gebruik dus van resultaten van onderzoek voor een specifiek wetsvoorstel, hetzij inhoudelijk in de fase van voorbereiding hetzij als men overweegt een evaluatiebepaling in het wetsvoorstel op te nemen (of toe te zeggen). In dat laatste geval moet de wetgever ex ante bedenken wat hij ex post met de resultaten van de evaluatie wil doen.

Een eerste onzekerheid omtrent het gebruik is dat de resultaten van dit type onderzoek zelden of nooit keihard zijn. Weliswaar hangt de aard en hoeveelheid van de kennisopbrengst onder meer af van de methode en de diepgang van het onderzoek, zowel bij ex ante evaluaties als bij ex post evaluaties zijn de onderzoeksresultaten immer (te) voorzien van voorbehouden. Dat geeft op zich niet, die kennis kan toch een baken zijn, maar enige terughoudendheid met het gebruik van onderzoeksresultaten is dan niet onbegrijpelijk. In dit verband is er op gewezen (bijvoorbeeld Leeuw 2010), dat het niet alleen mogelijk is dat onderzoekers gewoon fouten maken, maar ook dat zij opzettelijk fouten maken en hun bedrijfsbelang hoger achten dan hun professionele standaard: niet de opdrachtgever de harde waarheid (werkelijkheid) voorleggen, maar zorgen voor een volgende opdracht.

Een tweede onzekerheid omtrent het gebruik van de opbrengst is, althans in het geval van ex post onderzoek, dat als de onderzoeksresultaten worden geleverd het probleem mogelijk niet meer bestaat, het onderwerp van de politieke agenda kan zijn gevallen; er bestaat dan, met andere woorden, geen bestuurlijke behoefte meer aan de kennis die ter tafel komt. Een paradox van wetsevaluatie is dat het een middel kan zijn om de behandeling van een onderwerp uit te stellen (van tafel), maar dat het ongewenst een onderwerp weer ter tafel kan brengen.

De indruk bestaat, een indruk gebaseerd op onderzoek en ervaringen, dat er regelmatig gebruik wordt gemaakt van de resultaten van ex ante of ex post onderzoek. Klein Haarhuis (2014) constateerde dat van de resultaten van ex ante onderzoek veelvuldig kennis werd genomen. Wat betreft de beïnvloeding van standpunten en doorwerking (overnemen van resultaten) was het beeld wisselend. Uit mijn meta-evaluatie (Veerman 2013) van ex post evaluaties blijkt dat nieuw beleid regelmatig was gebaseerd op de bevindingen van onderzoek of op rapporten van deskundigen/commissies. Dat gebruik spreekt vanzelf indien die resultaten het nut van het ingezette beleid en de bijbehorende wetgeving bevestigt⁸; indien de uitkomst kritischer is, kan het om politieke redenen zo zijn dat geen gebruik wordt gemaakt van de bevindingen dan wel dat op ondergeschikte punten het beleid of de wetgeving wordt aangepast. Sprake is dan van selectief gebruik. Uit ouder onderzoek leid ik af dat de kwaliteit van het onderzoek en de prioriteit van het onderwerp (of de nieuws waarde)⁹ het belangrijkste zijn voor het gebruik (Dekker & Leeuw 1989) alsmede de

⁸ Ik maak deze opmerking ook omdat in het onderzoek naar ex ante evaluatie van Klein Haarhuis het gebruik van beleid ondersteunende resultaten enigszins negatief wordt gekwalificeerd als 'legitimerend gebruik'. Dat lijkt mij onjuist: als onderzoek uitwijst dat goed beleid gevoerd gaat worden dat mag dat vanzelfsprekend gebruikt worden en is dat niet negatief. Dat is alleen mis indien het departement druk heeft gezet op de onderzoekers voor het verkrijgen van ondersteunende uitkomsten of de onderzoekers uit eigen beweging de opdrachtgever ter wille zijn.

⁹ Klein Haarhuis noemt dit de besluitvormingscontext en de timing. De besluitvormingscontext vormde in haar vijf gevallen een zeer belangrijke factor voor het gebruik: naarmate het onderwerp minder omstreden was, was de doorwerking van de resultaten groter.

belangstelling voor het onderwerp bij beleid en wetgeving (wat verband kan houden met de prioriteit van het onderwerp).

Men kan het gebruik bevorderen door in het onderzoek de nadruk te leggen op manipuleerbare variabelen. Als men bijvoorbeeld alleen op 'de weersomstandigheden' als verklarende factor let, is de kans op gebruik van dit resultaat op voorhand wel op nul te schatten. Meer algemeen verdient het aanbeveling met de onderzoekers te bespreken welke variabelen van invloed zouden kunnen zijn.

ad tijd en kosten

Een belangrijke vraag is wanneer het departement de resultaten nodig heeft. Dat is vaak snel. Dat geeft beperkingen. Onderzoek uitvoeren kost tijd: denktijd, studietijd, ontwikkeling van een onderzoeksontwerp, tijd voor de praktische organisatie en tijd om het verslag te schrijven. Ik vermoed dat opdrachtgevers de neiging hebben de benodigde tijd te onderschatten. Op een half jaar voor eenvoudiger onderzoek mag men al snel rekenen. Ingewikkelder onderzoek kan veel meer tijd vergen. Naarmate er minder tijd beschikbaar voor wordt gesteld, zal de opbrengst waarschijnlijk ook geringer en minder betrouwbaar zijn.

Niet alleen de tijd die onderzoek kost, speelt een rol, ook de tijd die nodig is om de wetgeving de kans te geven zich te nestelen. In geval van institutionele wetgeving zal eerst de organisatie moeten worden vormgegeven en vervolgens moeten degenen die de instituties bevolken aan het werk gaan en hun plek vinden. Dat zal veelal tijd kosten. Dat betekent dat voor de kortere termijn bijvoorbeeld wel een knelpuntenanalyse kan worden uitgevoerd maar niet een onderzoek naar de effectiviteit. Na hoeveel jaar na de inwerkingtreding een wet het best kan worden geëvalueerd, hangt af van de inhoud van die wetgeving en van het doel ervan. In een wet wordt vaak een termijn van vijf jaar genoemd; dan moet het 'verslag over de doeltreffendheid van de wet en de effecten in de praktijk' aan het Parlement worden toegezonden; soms wordt een kortere termijn genoemd. Die termijn kan een compromis zijn, soms is de termijn gegeven door politieke druk. Evaluatie binnen korte termijn betekent veelal dat er de nodige manco's worden gevonden die dan wel, zoals hierboven gezegd, door het beleid als 'kinderziekten' kunnen worden bestempeld, hetgeen veelal een voorbarige conclusie is. Een redelijke termijn is ergens tussen de 3 en 5 jaar. Wil de wetgever de vinger aan de pols houden, dan zal een monitor als methode moeten worden gebruikt, niet een doeltreffendheidsonderzoek.

In termen van tijd dient ook bedacht te worden of een beschrijving van de te onderzoeken situatie op een moment vóór de inwerkingtreding van de wet nodig is (t_0) naast een moment ná de inwerkingtreding (t_1). Zeker indien men de doeltreffendheid wil vaststellen zal een vergelijking in de tijd het minste zijn, qua onderzoeksontwerp. Bij de keuze van een T_0 kan nog een rol spelen of de wetgever anticiperend gedrag wil uitsluiten of niet. Mensen en instanties kunnen immers handelingen verrichten om zoveel mogelijk profijt of zo min mogelijk hinder van een voorgenomen nieuwe wet te hebben. Als men de werking van de wet als zodanig wil vaststellen, zal het nodig zijn dergelijk anticiperend gedrag buiten te sluiten. Bij wetgeving en beleid gaat het veelal om ordening van de samenleving en betreft de evaluatie de vraag of die ordening tot stand komt. In dat geval is

een zuivere nulmeting minder cruciaal. De volgende vraag daarbij is dan op welk moment de situatie opnieuw onder de loep moet worden genomen (T1 dus) om eventuele verschillen waar te kunnen nemen.

Vervolgens neemt de vertaling van de onderzoeksresultaten in beleidsopvattingen tijd. Ik maakte hierboven onderscheid tussen enerzijds onderzoek en anderzijds de waardering van de onderzoeksresultaten; die waardebeoordeling kan, het lijkt vanzelfsprekend, pas plaatsvinden als de resultaten beschikbaar zijn. Het is een probleem dat wel in experimentele wetgeving voorkomt: wat gebeurt er in de tijd dat het experiment onderzoeksmatig is afgerond, maar evaluatief nog niet voltooid dan wel nog niet in regelgeving omgezet; dat moet dus worden geregeld.

Ook de kosten kunnen het besluit al dan niet een wet te evalueren beïnvloeden alsmede het type onderzoek dat wordt uitgevoerd. Het is duidelijk dat het ene type onderzoek (gekoppeld aan bepaalde vragen) duurder is dan het andere type (mogelijk gekoppeld aan andere vragen). Verder geldt ook hier dat de opdrachtgever moet bepalen welke kosten hij als kosten wil beschouwen: alleen de directe kosten van het onderzoek of ook bijvoorbeeld overhead, verzonken kosten of alternatieve kosten ('opportunity costs').

Ad onderzoeksmogelijkheden

Naast tijd en kosten bepalen de onderzoeksmogelijkheden de beslissing een evaluatieonderzoek uit te voeren. Als het niet mogelijk is de noodzakelijke gegevens te vinden, als het niet mogelijk blijkt de doelen van de wet te operationaliseren, als er voortdurende reorganisaties zijn op het te onderzoeken veld of wetswijzigingen ter zake al zijn voorgenomen, houdt het op. Men kan dan hoogstens op zoek gaan naar alternatieven als 'monitoring' of knelpuntenanalyse om de vinger aan de pols te houden.

En er moeten vanzelfsprekend betrouwbare onderzoekers beschikbaar zijn om het onderzoek uit te voeren.

Het idee is wel geuit om alle wetten te evalueren. Op het eerste gezicht heeft dit idee iets redelijks. Maar de condities van tijd, geld, de beschikbaarheid van onderzoekers en de mogelijkheden van onderzoek bepalen in hoge mate de praktijk. De Comptabiliteitswet kent de bepaling (art. 20) dat alle beleidsvelden regelmatig worden geëvalueerd. Deze bepaling wordt, zo is mijn indruk, niet hartgrondig en intensief nageleefd.

De conclusie is dat de beslissing een wetsevaluatie te houden is gebaseerd op een kosten-batenafweging waarbij politieke, financiële, cognitieve (welke kennis is nodig) en onderzoeksmatige factoren een rol spelen.¹⁰ Leeuw heeft in een essay (2010) een overzicht gegeven van baten en kosten

¹⁰ Zie M. Bamberger c.s. voor methodologische mogelijkheden bij diverse beperkingen, in: Shoestring Evaluation: Designing Impact Evaluation under Budget, Time and Dataconstraints, Am.Jn. of Evaluation, Vol.25, nr.1, 2004, p. 5-37.

van evaluatie, die vanzelfsprekend evenzeer van toepassing zijn op wetsevaluatie. De baten zijn in het tweede hoofdstuk genoemd, hier volgen de kosten: directe financiële kosten, kosten vanwege fouten (met reputatiekosten als gevolg maar ook kosten van verkeerd beleid), kosten vanwege de 'performance paradox' en administratieve kosten (vanwege bijvoorbeeld protocollering van evaluatie, routinisering, naleving evaluatieregels), risico van verlies van kritisch vermogen van de organisatie, risico van legitimering van beleid.

Overigens zijn er ook andere manieren om wetten te evalueren, met methoden die wat minder wetenschappelijk zijn: jaarverslagen, periodieke uitvoeringsrapporten, gebruik van kengetallen en statistische analyses, meer journalistieke essays. (Winter 2002).

ad horizonbepaling

Als element van het 'wanneer' besteed ik nog aandacht aan wat een 'horizonbepaling' wordt genoemd. Alle wetgeving is tijdelijk; de gemiddelde levensduur van een wet is vijf jaar (de Jong & Herweijer 2004). Sommige wetgeving is met opzet tijdelijk gemaakt, bijvoorbeeld als overgangsregeling of om een tijdelijk probleem aan te pakken. Een bijzondere vorm hiervan is de horizonbepaling. De definitie ervan staat niet vast. Dogmatisch gezien is het een bepaling dat deze wet na x jaar ophoudt te bestaan, tenzij wordt geoordeeld (bijvoorbeeld op grond van een evaluatie) dat de wet moet blijven voortbestaan. Dat is dus eigenlijk niets bijzonders, dat kan ook zonder zo'n bepaling. Deze bepaling wordt echter gezien als een expliciete valbijconstructie, als een manier om transparantie te bevorderen of minder regelgeving te bewerkstelligen. Los van het antwoord op de vraag of die veronderstellingen kloppen, de horizonbepaling heeft daarmee politiek betekenis gekregen. Zij kan worden ingezet in de onderhandelingen over een wetsvoorstel. Het komt er dan op aan het moment te bepalen waarop de wet zal expireren ('tenzij'), in aanmerking nemende dat vooraf onderzoek nodig is. Vragen over wat en hoe zullen daarbij een rol spelen.

4.2 Kenmerken van de wetgeving als indicatoren?

Uiteindelijk is beslissend wat de wetgever wil weten en dat is een politieke beslissing. Indien de beslissing al dan niet te evalueren niet gegeven is, kunnen naast bovengenoemde condities mogelijk ook kenmerken van wetten een indicatie geven voor die beslissing. Op het eerste gezicht lijkt dat een redelijke veronderstelling. Met enig gezond verstand kan men denken dat een minuscule, taalkundige verbetering minder reden geeft tot een evaluatie dan een wet die een nieuw zorgstelsel in het leven roept; of dat een wet waar de belangen gering zijn niet wordt geëvalueerd, terwijl daar wel reden voor kan zijn als er veel geld mee gemoeid is. Wat betreft ex post evaluaties kan men denken aan

- type wetten
- type normen
- functies van wetten
- motieven voor wetten.

Ik loop deze indicatoren langs maar kan op voorhand al aangeven dat deze indicatoren niet echt harde hulpmiddelen zijn.

ad type wetten

Een globale typologie van wetgeving is die van strafrecht, bestuursrecht en civiel recht en daarbinnen tussen materieel en formeel (procedureel) recht, met de erkenning dat sprake is van onderlinge toenadering. De meeste wetsevaluaties komen voor op het terrein van het bestuursrecht en, in de vorm van criminologisch onderzoek, het strafrecht en het strafprocesrecht. Dat het materiële civiele recht qua wetsevaluaties wat buiten de boot valt, heeft waarschijnlijk niet te maken met de materie zelf, die immers ook ordenend is en dus vatbaar voor sociaalwetenschappelijk onderzoek (Veerman 2012).

Andere typeringen zijn die in

- a. institutionele wetgeving of wetgeving gericht op gedrag, en
- b. meer codificerende of meer modifierende wetgeving.

Met institutionele wetgeving worden algemene voorzieningen in het leven geroepen, zoals een rechterlijke macht of procedures; codificerend betekent dat bestaand gedrag of bestaande informele regels in een wet worden vastgelegd; modifierend is dat instituties of gedragsregels worden gewijzigd dan wel nieuwe worden voorgeschreven. De beide typologieën zijn niet strikt.

Men zou mogelijk denken dat er eerder reden is modifierende regelgeving te evalueren omdat men kan verwachten dat wijzigingen tot onzekerheid leiden en dat moeilijker voorspelbaar is hoe mensen daarmee omgaan. Maar de geschiedenis leert dat codificaties nooit zonder meer codificeren en dat de formalisering die een codificatie betekent, tot ander gedrag kan leiden. Bekend is het empirisch waargenomen mechanisme van 'crowding out'; dat is het mechanisme dat mensen hun spontane gedrag veranderen als het een wettelijke verplichting wordt. Dus codificaties zijn ook modificaties.

Institutionele voorzieningen komen tot leven als zij bevolkt worden of als mensen er gebruik van (kunnen) gaan maken. Dat kan sociaalwetenschappelijk worden onderzocht.

Een stevige indicator levert deze typologie niet op.

ad type normen

Een eerste onderscheid is dat tussen open normen en gesloten normen, zo men wil tussen vage en precieze normen. In geval van open normen zijn doeltreffendheid en effecten mogelijk moeilijker vast te stellen dan in geval van strikte normen; er zal waarschijnlijk meer variatie (kunnen) bestaan. Ook is het behalen van precieze doelen makkelijker vast te stellen dan van vage doelen. Dat zouden redenen kunnen zijn niet snel tot evaluatie van open regelingen over te gaan. Echter, in termen van processen en mechanismen zijn open normen weer interessant en kan worden bezien in hoeverre de processen overeenkomen met wat de wetgever voor ogen stond.

Een ander onderscheid is dat tussen rechten en plichten. In geval van rechten kan het gebruik ervan worden onderzocht (wie, wanneer, waarom wel/niet) en of dat gebruik overeenstemt met de doelstelling ervan dan wel of de materiële condities voor de uitoefening van rechten wel zijn vervuld. Waar rechten kunnen botsen (denk aan grondrechten), kan worden onderzocht welke belangen daarbij aan de orde zijn (de opvattingen worden wel geventileerd). In geval van plichten kan de naleving worden onderzocht (wie, wanneer, waarom wel/niet) en in dat verband de condities waaronder dat gebeurt, en de beleidstheorie (plicht –naleving -doelbereiking).

Ook het gebruik van bevoegdheden is goed te onderzoeken. Het spreekt niet vanzelf dat instanties of autoriteiten hun bevoegdheden gebruiken, en onderzocht kan worden wanneer dat wel en niet gebeurt, welke belangen en mechanismen daarbij een rol spelen. Een belangrijk aspect in het gebruik kan zijn de aanwezigheid van materiële middelen als voorwaarde voor het gebruik van bevoegdheden.

In geval van procedurele voorschriften kunnen voor hun werking, alweer, de wijze van implementatie en de beschikbaarheid van materiële voorzieningen in het onderzoek worden betrokken, hoe de procedures worden ervaren door verschillende betrokkenen, wat hun effecten zijn. Ook de wisselwerking met materiële normen kan worden gezien. Immers, bij wetgeving is vaak sprake van een mix van inhoudelijke normen, procedurele normen en handhavingsregels.

ad c functie

De functie van de betreffende wetgeving kan een rol spelen bij de beslissing al dan niet een wetsevaluatie te laten houden. De primaire functie van wetgeving is de ordenende, en die functie stimuleert op zich onderzoek naar de werking. Maar er kunnen belemmeringen zijn. Indien de wet een moeizaam bereikt compromis behelst, zal de neiging de discussie na vier of vijf jaar weer te heropenen vanwege de dan verschenen onderzoeksresultaten niet groot zijn (tenzij een evaluatiebepaling onderdeel uitmaakt van het compromis). Onderzoeksmatig kan het wel omdat het compromis ook een ordenende werking kan hebben. Andere functies van wetgeving, zoals de consoliderende en de symbolische kunnen weer wel heel goed voorwerp van evaluatie zijn; ook daar is de wet tevens ordenend.

ad d motief

Indien het motief ordening is, is er weinig bijzonders aan de hand en gelden de hiervoor genoemde overwegingen. De wetgever wil weten of de gekozen oplossing voor het geïdentificeerde probleem werkt. Dat is ook het geval als de wetgeving het gevolg is van een incident. Ex ante zal bezien kunnen worden of sprake is van een structureel probleem en of op grond daarvan wetgeving nodig is.

Het komt voor dat nieuw beleid wordt ingezet om reden van bezuinigingen. Dat kan openlijk worden aangegeven, het kan ook verborgen zijn als in de officiële documenten wordt benadrukt dat de

systemwijziging nodig was, dat er knelpunten waren en dat nu twee vliegen in een klap worden geslagen. Dat kan voor een departement een reden zijn geen evaluatie te houden dan wel een evaluatie waarin wordt gezien of het doel, bezuiniging, ook werkelijk wordt gehaald. Hier geldt dat onderzoek naar neveneffecten, in casu de systeemverbetering, het beeld zal completeren.

De conclusie moet zijn dat de bovengenoemde indicatoren mogelijk aanleiding zijn voor specifieke vragen bij een wetsevaluatie, maar dat zij voor de beslissing al dan niet te evalueren geen rol spelen. Het gaat er inderdaad steeds weer om: wat wil de opdrachtgever weten.

5. Wie evalueren?

5.1 Beleidsmakers of onderzoekers?

Het waarderende deel van de wetsevaluatie is een beleidszaak en dat doen dus politici/bewindslieden en degenen die zij daartoe opdracht hebben gegeven, hun beleidsambtenaren veelal. Voor het onderzoeksmatige deel van de wetsevaluatie is dat minder vanzelfsprekend. Op zich lijkt het aantrekkelijk de eigen staf het onderzoek te laten doen: de verzonken kosten en de alternatieve kosten zijn niet zo zichtbaar, het oogt goedkoper. Bovendien zal de leiding de eigen mensen meestal wel vertrouwen in de zin dat zij niet al te lastige uitkomsten genereren. Voor het gewone beleidsonderhoud kan dat misschien nog wel, hoewel? Zodra sprake is van onderzoek, kan dat zeker niet. Ambtenaren zijn geen wetenschappelijke onderzoekers, wetenschappelijk onderzoek is een vak apart. Daarbij komt dat beleidsambtenaren en wetgevingsjuristen zichzelf gaan evalueren. En zij hebben een zekere bias; naarmate zij langer met hun onderwerp bezig zijn, gaan zij geloven in wat zij beweerden (moesten beweren). En als zij kritisch kunnen zijn ten opzichte van het voorgenomen of uitgevoerde beleid, komen zij in een dilemma terecht. Hun superieuren houden meestal niet van negatieve berichten, houden niet van problemen.

In een democratie is een faire discussie van belang, waarin argumenten ter tafel kunnen komen. Uit onderzoek van Pröpper en van Winter blijkt dat de inzet van wetenschappelijk onderzoek in het debat tot een betere discussie in het parlement leidt, zij het niet noodzakelijk tot betere wetgeving. Dat wijst erop dat de politieke rationaliteit de doorslag geeft; daar kan men in een democratie vrede mee hebben. Er is in dit verband een ander argument om de evaluatie niet door de betreffende ambtenaren zelf uit te laten voeren. De geloofwaardigheid van de onderzoeksbevindingen zal gering zijn indien dat gebeurt; de evaluatoren kunnen worden beschouwd als slaggers die hun eigen vlees keurden.

Onderzoek kan worden uitgevoerd door individuele onderzoekers, universitaire onderzoeksgroepen, professionele onderzoeksbureaus of departementale onderzoeksinstituten.

Intermezzo: aanbevelingen

In het algemeen stellen opdrachtgevers het op prijs, soms zelfs als eis, dat de onderzoekers aanbevelingen doen naar aanleiding van hun onderzoek. Dat is begrijpelijk: de uitkomsten zijn niet altijd hard, de onderzoekers hebben een goed beeld van wat er speelt, dus hun deskundig oordeel wordt op prijs gesteld. Ik heb twee reserves bij deze opvatting en praktijk, een wetenschappelijke en een bestuurlijke. De wetenschappelijke reserve is dat de aanbevelingen op zich niet zijn onderzocht. Het onderscheid tussen causale en finale redeneringen speelt hier een doorslaggevende rol. Mijn bestuurlijke reden is dat de verantwoordelijkheid voor het beleid rust bij de bestuurders. Zij moeten hun conclusies trekken uit het onderzoek en daarover verantwoording afleggen.

5.2 Organisatie van het onderzoekwerk

Een ander aspect van de 'wie'-vraag is hoe de opdrachtgever aan onderzoekers komt, en als hij onderzoek uit wil laten voeren, hoe hij dat organiseert.

Het helpt opdrachtgever en onderzoekers erg als een evaluatieonderzoek wordt gebaseerd op een onderzoeksopdracht, waarin de hoofdvraag en eventuele subvragen worden geformuleerd, alsmede de tijd die beschikbaar is voor het onderzoek. De opdracht kan ook elementen van de methode bevatten (wat of wie worden in elk geval in het onderzoek 'meegenomen', het minimum aantal respondenten, en zo meer). De opdracht kan worden vooraf gegaan door een tender waarbij onderzoekers/onderzoeksbureaus op basis van een startdocument worden uitgenodigd een offerte te doen. De gewoonte is niet voor de voorstellen te betalen. Omdat in een goed voorstel al veel onderzoekswerk is uitgevoerd, waar de opdrachtgever baat bij heeft, zou ik geneigd zijn wel een geldelijke tegenprestatie te leveren, als regel of bijvoorbeeld indien een voorstel niet wordt gegund maar wel veel nuttige gedachten bevat.

Ik herhaal mijn waarschuwing dat het belangrijk is dat de opdrachtgever niet te veel ruimte biedt aan de onderzoekers (want op welke vragen krijgt deze dan antwoord) maar ook niet te weinig vanwege de bias in de vraagstelling (denk aan de loutere doelbereiking). Een goed gesprek is nodig over de mogelijkheden en de beperkingen van vraagstelling en methode. Voor elk onderzoek geldt –mede daarom– wel dat het verstandig is een begeleidingscommissie in te stellen. Deze kan bestaan uit departementale vertegenwoordigers (beleidsmedewerkers, wetgevingsjuristen), betrokkenen en wetenschappers; van het concrete onderzoek hangt af welke geledingen en hoeveel personen deel uitmaken van de begeleidingscommissie. Het verdient de voorkeur een wetenschapper als voorzitter te kiezen, op voorwaarde dat deze ook technisch kan voorzitten. De begeleidingscommissie begeleidt het onderzoek, opent deuren, bewaakt de voortgang en de uitvoering conform de onderzoeksopdracht, bespreekt kritisch de werkwijze en de bevindingen; zij komt bijeen op enkele strategische momenten. Als veelvuldig betrokkene bij begeleidingscommissies voeg ik hier aan toe dat zowel het beleid als de onderzoekers soms graag het onderzoek aanvullen met nieuwe vragen die ook zo interessant zijn. Soms is het nodig omdat vooraf niet alles te bedenken was, soms is het de pure belangstelling; in beide gevallen dient gesproken te worden over de consequenties ervan in tijd en geld.

In sommige gevallen, bijvoorbeeld bij monitoring en zeker indien de mogelijkheid van bijstelling en aanpassing in het implementatieproces bestaat, is het verstandig een (beleidsgeoriënteerde) stuurgroep in te stellen die de verzamelde data interpreteert en de minister kan adviseren tot diepte-onderzoek of aanvullende maatregelen. Een budget is dan immer nodig.

a. Marktpartijen

Thans zijn de nodige onderzoeksbureaus werkzaam, is er een markt met de nodige aanbieders. Ook universitaire onderzoekers of onderzoeksgroepen horen tot die markt. Daarbinnen zijn kwaliteitsverschillen, en bureaus verschillen in inhoudelijke expertise. Er zijn bureaus die gespecialiseerde methodes hanteren, die gebruik maken van bijvoorbeeld panels welke kunnen worden ingezet. Dat zal vaak goedkoper zijn dan wanneer een bureau dergelijke panels zelf moet

gaan organiseren. Let altijd op wie in welke mate het uitvoerende werk doen en wie begeleiden; de prijzen voor de verschillende betrokkenen verschillen aanmerkelijk.

Er zal een neiging zijn om te blijven werken met bureaus met wie de samenwerking goed beviel. Dat heeft voordelen, bijvoorbeeld dat het bureau het veld kent en dat open kan worden gediscussieerd over welke vragen worden gesteld en wat kan worden onderzocht. Een nadeel kan zijn dat de kritische afstand die nodig is tussen opdrachtgever en onderzoeker kleiner wordt. Zoals gezegd, verschillen de werelden van onderzoek en beleid wezenlijk. Er dient weliswaar een brug te worden geslagen tussen beide, maar beide partijen dienen ook te weten waar zij vandaan komen. In dit verband zij de algemene opmerking gemaakt dat de opdrachtgever er op bedacht moet zijn dat onderzoekers als ondernemers graag een volgende opdracht wensen en dat het gevaar bestaat dat hun uitkomsten te rooskleurig zijn (of te rooskleurig worden gepresenteerd), of dat te veel is gekeken naar het door de wetgever beoogde doel en te weinig naar de neveneffecten.

b. Een departementaal onderzoeksinstituut

Het komt voor dat departementen een eigen onderzoeksinstituut hebben. Het WODC, al in 1946 opgericht, is daarvan het voorbeeld.¹¹ De spanning tussen onderzoek en beleid speelt ook daar. Deze spanning wordt gereguleerd door een aantal procedures die enerzijds de onafhankelijkheid van de onderzoekers moeten waarborgen en anderzijds de politieke en bestuurlijke verantwoordelijkheid de ruimte geven. Die procedures bestaan. Belangrijk daarin is dat in geval van een onderzoeksopdracht uitvoerig overleg plaatsvindt tussen onderzoekers en het beleid (over de beleidsvraag, over de termijn), dat elk onderzoek wordt gepubliceerd maar dat het beleid een periode de tijd krijgt een reactie voor te bereiden op de uitkomsten van het onderzoek, dat er een gemengde begeleidingscommissie is. Ook al is de onderzoeksmethode vrij en worden geen instructies gegeven voor de uitkomst (dat kan ook niet, de geloofwaardigheid van het WODC zou direct in gevaar komen en daarmee het nut ervan), er blijft een zeker risico (althans: de geur van een risico) vanwege de nabijheid. Mijn stelling is dat de onderzoekers van het WODC geen ondernemers zijn en zich geen zorgen hoeven te maken voor hun volgende opdracht; in die zin zijn zij onafhankelijker dan onderzoeksbureaus die in de markt opereren. Als onderzoekers staan zij bovendien met minstens anderhalf been in de wereld van wetenschap en onderzoek, ligt hun primaire loyaliteit daar.¹²

Sommige andere landen kennen eveneens onderzoeksinstituten die zijn verbonden aan een departement (zoals het VK en Australië).

¹¹ Het Kennisinstituut voor Mobiliteitsbeleid (2006) bij het Ministerie van I en M is eveneens ingebed in de departementale organisatie. Het SCP, het CPB en het RIVM zijn nogal onafhankelijke instituten die als Directoraat of Directoraat-Generaal zijn georganiseerd. Zij spelen als planbureaus meer een rol bij ex ante onderzoek dan bij ex post evaluatie.

¹² Ik heb zelf acht jaar bij het WODC gewerkt. De spanning tussen onderzoek en beleid was soms merkbaar (lang niet altijd overigens, de mensen van beleid en wetgeving zijn ook gewoon nieuwsgierig) maar er bestond oog voor en respect voor elkaars positie. Nimmer is mij een gewenste uitkomst opgelegd; voor het beleid kritische resultaten werden niet blijmoedig in ontvangst genomen, maar leidden niet tot niet-publicatie.

c. Intermediairs

Een derde vorm is dat een departement onderzoek uitzet via een onderzoeksbeleidsafdeling. De medewerkers van zo'n afdeling kennen de wereld van het onderzoek én zij zijn collega's van de beleidsmedewerkers. De brug tussen onderzoek en beleid wordt daar geslagen. Het ministerie van VWS heeft deze intermediaire functie ondergebracht in een afzonderlijk instituut, ZonMW.

Voor thematisch onderzoek kunnen dergelijke departementale constructies als onder b en c genoemd ook worden ingezet. Of omgekeerd, thematisch onderzoek vergt een organisatie. In geval thematisch onderzoek breder is en bijvoorbeeld de wetgeving van meer departementen aangaat, zal een gezaghebbend interdepartementaal gremium daarin leiding moeten geven.

6. Waar evalueren?

a. Meestal is een evaluatieonderzoek een mix van studie (bureauonderzoek) en veldwerk. Waar het veldwerk plaatsvindt, hangt vanzelfsprekend af van waar de normadressaten zich bevinden. Daarbij is het van belang voor ogen te houden dat wetten in veel gevallen middelen verdelen en herverdelen. Er zijn dan ten minste twee groepen die in het onderzoek worden betrokken. Daarbij komt dat de samenleving uit vele groepen bestaat, dat mensen tot verschillende groepen behoren, en dat in het veldwerk met die variëteit wordt rekening gehouden. Bedrijven als Motivaction gebruiken typologieën die wel uit acht of meer typen bestaan.

Binnen groepen hebben mensen weer verschillende posities, en die posities kunnen betekenisvol zijn wat betreft de werking van een wet. Hierboven gaf ik al aan dat in een gemeente de burgemeester andere kennis, belangen en opvattingen heeft dan zijn ambtenaren in de binnen- of de buitendienst. Tussen die twee laatste bestaan ook weer verschillen, zoals uit empirisch onderzoek bleek. Vooraf dient daarom te worden bezien waar het empirisch onderzoek zinvol zal zijn.

b. Een tweede kwestie in dit verband is of internationaal vergelijkend onderzoek moet worden gedaan. Ik zou daar 'een genuanceerd standpunt' in willen innemen. De betekenis van internationaal vergelijkend onderzoek (dat met name bij ex ante evaluatie aan de orde is), is naar mijn mening vooral een inspirerende; het brengt op ideeën. Of wat in andere landen werkt ook hier zal werken, blijft een grote vraag en het meest waarschijnlijke antwoord is: nee. De condities zijn daarvoor te verschillend. Dat geldt zeker voor rechtsvergelijkend onderzoek. In aanmerking nemende dat rechtsregels betekenis krijgen in de context, zal het zeer ongewis zijn dat regels uit andere rechtsstelsels hier zullen werken. Daar is ook empirisch materiaal voor (bijvoorbeeld Jonkers 2003). Vergelijking van de stelsels op zich is niet voldoende.

In het bureauonderzoek kan het verstandig zijn resultaten van buitenlands onderzoek te betrekken, omdat die aanwijzingen kunnen geven omtrent wat bij wie en waar te kijken. Dat geldt voor zowel ex ante als ex post evaluatie.

Een andere functie kan zijn dat beleidstheorieën ermee worden onderzocht. Zo bleek uit een rechtsvergelijkende studie naar echtscheiding dat in landen met een uniform echtscheidingsregime grote verschillen tussen regio's bestonden in aantallen echtscheidingen, terwijl in gebieden met verschillende wetgeving geen verschillen daarin te zien gaven. (Lenters 1993) Echtscheidingen zijn dus niet alleen het effect van de echtscheidingswetgeving. In dergelijk onderzoek kan dan tevens worden bezien of bijvoorbeeld de sociale wetgeving een belangrijker oorzaak voor echtscheiding is dan de echtscheidingswetgeving zelf.

Dergelijk onderzoek om een beleidstheorie te toetsen kan ook in ex post evaluatie worden toegepast. In ex post evaluaties kan rechtsvergelijking bovendien de uitkomsten relativeren, bijvoorbeeld in de zin dat sommige problemen zich ook elders voordoen.

Vergelijkend onderzoek is vaak kostbaar en vereist een flinke organisatie; het hebben van een internationaal netwerk of gebruik maken van instituten met een dergelijk netwerk, zoals het Max Planckinstituut, kan erg praktisch zijn.

7. Hoe evalueren?

7.1 Waarom deze methodologische opmerkingen?

Een belangrijke reden om de 'hoe'-vraag aan de orde te stellen is dat 'wat' en 'hoe' (en 'wanneer') nauw samenhangen. De opdrachtgever kan wel de doeltreffendheid willen laten onderzoeken maar als hij geen flauw benul heeft van onderzoeksontwerpen, kan hij de waarde van het onderzoek en de waarde van de uitkomst niet bepalen. Methodologie gaat over de manier waarop antwoorden kunnen worden gevonden op de gestelde vragen. Maar misschien nog meer gaat methodologie over het soort uitspraken dat kan worden gedaan. Willem Frederik Hermans, ook een wétenschapper, zei ooit: als iemand in het buitenland is, heeft hij in het algemeen aan twee waarnemingen genoeg om algemene uitspraken over dat land te doen. Wetenschappelijk gesproken moet men zeggen: met zo'n kleine onderzoeksgroep en zonder iets te weten over de representativiteit ervan, zijn dergelijke uitspraken over een land onbetrouwbaar en ongeldig. De 'betrouwbaarheid' betreft de juistheid van de meting (in casu: twee waarnemingen), de 'geldigheid' de juiste operationalisering (in casu: het object van de waarneming als indicator voor het hele land). Wil de opdrachtgever iets kunnen met de bevindingen van de onderzoekers, dan moeten de bevindingen geldig en betrouwbaar zijn.

Een andere reden er op in te gaan is dat in wetsevaluaties de causaliteit zo'n belangrijke rol speelt. In een intermezzo werd al gesteld dat causaliteit een hypothese is in het hoofd van de onderzoeker; dat kan nu worden uitgelegd. Ik ga er daarbij vanuit dat vaak juristen betrokken zijn bij wetsevaluatie en dat zij weinig methodologische kennis hebben; de juridische opleiding voorziet daar in elk geval niet in.

In het onderstaande komen enkele basale elementen van methodologie aan de orde. Daarbij wordt onderscheid gemaakt tussen onderzoeksontwerpen (globale onderzoeksstrategieën) en onderzoeksmethoden (technieken om data te verzamelen).

7.2 Twee onderzoekstradities

Wetten zijn de juridische voertuigen van beleid. Zij kunnen worden geëvalueerd als juridische documenten, de gelding is dan aan de orde. Zij kunnen ook worden geëvalueerd met het oog op de beleidsmatige inzet. De werking of doelbereiking zijn dan aan de orde; sociaalwetenschappelijk onderzoek kan daarover gegevens genereren.

Globaal kan men in het sociaalwetenschappelijk beleidsonderzoek twee onderzoekstradities onderscheiden. In de ene traditie poogt men een causale relatie te leggen tussen de interventie en het resultaat van die interventie door zoveel mogelijk externe invloeden uit te schakelen. Men kan dat de experimentele school noemen, omdat het laboratoriumexperiment de meest krachtige setting is waarin dat kan gebeuren. De experimentele traditie komt voort uit de beta-wetenschappen en is overgezet naar maatschappelijke situaties. In de andere traditie wordt juist wel ingegaan op die externe invloeden, wordt in de context(en) bekeken welke processen plaatsvinden na een interventie; het wordt wel het C-M-O ontwerp genoemd: context-mechanisme-outcome.

Ik denk dat onderzoek in de geest van beide tradities mogelijk is bij zowel ex ante evaluatie als bij ex post evaluatie; let wel: 'in de geest'.

ad experimentele ontwerpen

Een echt experimenteel ontwerp is niet mogelijk in geval van wetsevaluatie ex post. Wetten worden gelanceerd in de werkelijke wereld en externe invloeden zijn niet of nauwelijks onder controle te houden. Het hoogst haalbare is het quasi-experimenteel ontwerp dat inhoudt dat niet alleen de situatie voor (t 0)¹³ en de situatie na de wettelijke interventie (T1) worden vergeleken, maar ook de verschillen voor twee groepen: een willekeurig samengestelde groep die is blootgesteld aan de wettelijke interventie en de eveneens willekeurig samengestelde controlegroep die er niet aan is blootgesteld. Door beide, tijd en groep, te vergelijken, kunnen verschillen een aanduiding geven van het effect van de wettelijke interventie. In de praktijk komt ook dit ontwerp nauwelijks tot niet voor. Het probleem zit veelal in de samenstelling van de groepen. Het is voor veel wetgeving niet mogelijk, althans het wordt niet wenselijk geacht, om regels voor willekeurig bepaalde groepen wel én niet te laten gelden. Dat zou overigens wel kunnen indien men het experiment wettelijk vastlegt. Soms zijn alternatieven mogelijk zoals onderzoek naar gemeenten of andere uitvoerende organen die verschillen in hun uitvoering of tussen personen die wel en die niet gebruik maken van voorzieningen. Maar dan zijn het waarschijnlijk niet willekeurig samengestelde groepen die zomaar vergelijkbaar zijn. Harde causale uitspraken zijn dus niet mogelijk, de invloed van de wet op de doelbereiking is hoogstens aannemelijk te maken.

Het next best ontwerp, maar in deze traditie niet indrukwekkend, is wat betreft de wetsevaluatie de meting van verschillen tussen t0 en t1. Omdat er veel andere invloeden kunnen zijn geweest zijn causale uitspraken niet te doen. Het ontwerp leidt tot een indruk.

Intermezzo: de heuristische cirkel

Een groot probleem dat zich bij dergelijke designs voordoet is: wat wordt beschreven op tijdstip 0. Eigenlijk moet men voor de beschrijving van de situatie op t0 al weten wat belangrijk is voor de vergelijking, welke factoren er toe doen, gelet op het wetsvoorstel. Dat is vaak niet zo duidelijk; dat zou men pas kunnen weten als het onderzoek is voltooid. Vandaar de benaming 'heuristische cirkel'. Hoogstens kan op grond van de literatuur en mogelijk praktijkervaringen een idee worden gekregen van wat wezenlijk is; die veronderstellingen kunnen worden gebruikt om als het ware door de cirkel te doorbreken. Dit maakt ook duidelijk waarom onderzoek tijdrovend kan zijn: er moet veel gestudeerd worden.

Het meest voorkomende ontwerp voor wetsevaluaties is een meting slechts op t1. Bijvoorbeeld wordt aan respondenten gevraagd of zij invloeden zien van de wettelijke interventie en zo ja welke, eventueel aangevuld met data uit deskresearch waarbij bijvoorbeeld gegevensverzamelingen over belastingopbrengsten, zorggebruik, verhuisgedrag etcetera, in de tijd worden vergeleken. Het is duidelijk dat de causale uitspraken met zo'n onderzoeksontwerp boterzacht zijn geworden. Dat wil niet zeggen dat er geen nuttig inzicht kan worden gegeven, maar de causaliteit tussen wet en gedrag

¹³ De situatie op T0 kan op diverse manieren worden beschreven, voluit of aan de hand van bijvoorbeeld een aantal indicatoren. Voor een beschrijving dient men een idee te hebben van wat het beoogde doel is en vandaaruit gezien welke data op T0 moeten worden verzameld om een goede vergelijking te kunnen maken.

is niet zomaar te onderbouwen. De respondenten hebben hun bias, hun belangen, hun slechte geheugen, hun denkfouten; de dataverzamelingen zijn globaal en veelal niet opgesteld met het oog op juist deze wetgeving.

Bij dergelijke zachte ontwerpen als deze kan het nuttig zijn een 'counterfactual' op te stellen: wat zou er zijn gebeurd als deze wet niet in werking was getreden (wat dus de 'nulhypothese' is), en dat te vergelijken met de uitkomst van t1. Het blijft dan zacht, maar het is wel beter beredeneerd.

ad contextuele ontwerpen

In de geest van de andere traditie wordt gepoogd te begrijpen wat een wettelijke interventie teweegbrengt in de context waarin de wet moet werken. Beleidsinterventies werken verschillend voor de diverse betrokkenen. Wat werkt voor wie onder welke omstandigheden?, zijn de vragen die worden gesteld. Mensen worden geleid door verschillende mentale en sociale mechanismen¹⁴; het hangt van de context af welke een rol spelen. 'De context' is een abstract begrip dat op vele wijzen kan worden ingevuld. Mensen leven in diverse contexten. Een context omvat sociale, economische, politieke, organisatorische, geografische, historische bijzonderheden. Wettelijke interventies komen in contexten terecht en kunnen leiden tot bepaald gedrag, kunnen sommige mechanismen activeren of andere juist bevriezen. Er is een interactie tussen interventie, context en mechanismen; het resultaat van de interactie is de outcome. Met inzicht in wat voor wie werkt onder welke omstandigheden kan men ook een idee krijgen of het doel van de wetgeving wordt bereikt, zeker indien sprake is van een uitgewerkte beleidstheorie.

In deze traditie wordt dus de context(en) gezien waarin een interventie haar werk moet doen. Omdat er veel contexten zijn, is de representativiteit mogelijk gering en zijn geen algemene causale uitspraken te doen. Deze manier van onderzoeken leidt tot inzicht in mechanismen die een wet in werking zet, tot een 'Verstehen'.

Een voordeel van inzicht in processen en mechanismen is, dat zij kunnen bijdragen aan inzicht in de werking van ook andere wetten. Immers, dezelfde menselijke mechanismen doen zich in meer situaties voor.

7.3 Andere ontwerpen

a. Monitoring

Zeker indien het beleid de vinger aan de pols wil houden voor de voortgang van de uitvoering of wil weten hoe de implementatie van de wet verloopt, kan een monitor een behulpzaam instrument zijn. Op gezette tijden wordt informatie gegenereerd over een beperkt aantal sleutelindicatoren. Indicatoren zijn te zien als thermometers: ze geven niet de werkelijkheid weer, zij hebben een signaalfunctie; zij geven bij bepaalde waarden aanleiding tot een gesprek, tot nader onderzoek.

¹⁴ Zie voor een beschrijving Veerman, 2013.

Cruciaal voor een nuttige opbrengst zijn zowel de keuze van de indicatoren¹⁵ als de wijze waarop de gegevens worden verzameld. Niet uit te sluiten valt dat in geval van monitoring de aandacht voor de gekozen indicatoren de uitvoerders van de wet in een richting stuurt. De werking van de wet als zodanig is dan minder goed vast te stellen. Dat is op zich geen probleem, daar hoeft het bij wetgeving niet om te gaan. Het doel van monitoring kan zelfs zijn beleidsmatig te interveniëren als de ontwikkeling in verkeerde richting dreigt te gaan. Cruciaal voor het gebruik van de resultaten is dat de opdrachtgever zelf de belangstelling voor de uitkomsten behoudt, die tot zich neemt en er zonodig wat mee doet. Ik heb meermalen meegemaakt dat dat niet het geval was. Dat moet dus worden georganiseerd, bijvoorbeeld met behulp van een beleidgeoriënteerde stuurgroep.

b. Gevalstudies

In de C-M-O traditie zijn onder meer gevalstudies geschikt om inzicht te krijgen in processen en mechanismen. Een geval kan een individu zijn, een groep, een gebeurtenis. De onderzoeker duikt diep in het geval en beschrijft systematisch wat er gebeurt.¹⁶ Het is een empirisch onderzoek dat een fenomeen bestudeert in zijn 'real life' context. Vanzelfsprekend zijn in een gevalstudie meer gevallen te bestuderen. Ook bij gevalstudies kunnen kwantitatieve data en diverse bronnen worden benut. Hoewel over de generaliseerbaarheid van de bevindingen verschil van mening bestaat, generalisering kan niet zonder meer plaatsvinden. Bij meer gevallen kan dat weer wat beter.

Gevalstudies zijn goed te gebruiken als explorierend onderzoek.

c. Experimenten

Leeuw heeft als voordeel van evaluaties genoemd dat zij bijdragen aan een geest van experimenteren. Dat is inderdaad belangrijk omdat beleid en wetgeving een sprong in het duister betekenen, wat niet altijd zo wordt ervaren en zeker niet altijd zo naar buiten gebracht. Een meer experimentele houding zou de inhoud van het beleid en de gewenste zekerheid omtrent de werking kunnen bevorderen. De Aanwijzingen voor de regelgeving kennen de mogelijkheid van experimenteerbepalingen in de vorm dat in een wet kan worden opgenomen dat tijdelijk afgeweken kan worden van de geldende regels (Ar 10a en 10b). Ook tijdelijke wetten die als experiment worden gekwalificeerd komen voor. Toch bestaat de nodige terughoudendheid. Wetten werken algemeen, moeten rechtszekerheid en rechtsgelijkheid waarborgen. Zeker in geval van een quasi-experimenteel ontwerp (vergelijking tussen twee tijdstippen én tussen twee groepen) lijkt dat moeilijk. Toch zou het moeten kunnen, naar mijn oordeel. Beperking van een wettelijke regeling in de tijd is heel gebruikelijk. Eveneens is gebruikelijk dat bij wet groepen worden aangeduid op wie de wet van

¹⁵ Het vaststellen van relevante indicatoren is lastig. Een probleem kan zijn dat het te controleren verschijnsel leidt tot een indicator maar dat die indicator ook een signaal kan zijn van een ander verschijnsel: een indicator voor de werking van de echtscheidingswetgeving kan zijn het aantal echtscheidingen, maar het aantal echtscheidingen kan ook een indicator van bijvoorbeeld economische ontwikkelingen zijn.

¹⁶ Yin (1984) heeft een belangrijke bijdrage geleverd aan de onderzoeksmethode van de gevalstudie. Een recente loot aan deze tak is het zogenoemde N=1 ontwerp.

toepassing is, wat betekent dat zij op anderen juist niet van toepassing is. In dit geval van een experiment is het bijzondere dat een groep op wie de wet naar inhoudelijke maatstaven van toepassing is, nu van toepassing wordt uitgesloten om een vergelijking mogelijk te maken met de groep waarop die regeling van toepassing is. De rechtvaardiging hiervan is dat het doel van het experiment is de best mogelijke regeling te vinden en die wordt na enige tijd bereikt. Dan zal achteraf blijken dat sommigen het beste af waren, maar anderen helaas (maar tijdelijk) niet. Dat zonder experiment de verkeerde oplossing met de bijbehorende wetgeving zou hebben kunnen zijn gekozen en dat een wetswijziging nodig zou zijn, rechtvaardigt mijns inziens het experiment. Anders gezegd: nu heeft een kleiner deel van de doelgroep tijdelijk onder een nadelig regime geleefd in plaats van de gehele doelgroep.

Bij experimenten moet vanzelfsprekend goed worden uitgezocht wat de inhoud van de regeling zal zijn (of zo nodig van meer regelingen), naast de controlegroep (voor wie het nul-alternatief geldt). In de wet kan mijns inziens beter niet het onderzoeksontwerp worden uitgesponnen, dat moet in overleg met onderzoekers. De wet moet alleen de legitimering ervoor geven.

7.4 Methoden van dataverzameling

a. Algemeen

- Binnen de hiervoor genoemde ontwerpen moeten steeds data worden verzameld. Daarvoor bestaan ook weer verschillende methoden, kwantitatieve en kwalitatieve. Kwantitatieve methoden vertalen de sociale werkelijkheid in numerieke gegevens. Het voordeel daarvan is dat het mogelijk is diverse variabelen aan elkaar te koppelen en de aard en sterkte van hun onderlinge relaties te bepalen. Het nadeel is dat er minder inzicht bestaat in de diverse contexten waarin de variabelen een rol spelen.

Kwalitatieve methoden zijn gericht op het begrijpen van sociale verschijnselen. Recht wordt gedaan aan de complexiteit daarvan, de analoge werkelijkheid wordt niet gedigitaliseerd. Het nadeel is dat generalisaties minder goed mogelijk zijn, dat niet makkelijk algemene uitspraken kunnen worden gedaan.

Het risico van kwantitatieve methoden is 'decontextualisering', het risico van kwalitatieve methoden is 'over-contextualisering'.¹⁷ Beide methoden hebben dus hun sterke en een zwakke kanten. Het ligt voor de hand om waar mogelijk (indachtig tijd, kosten en onderzoekbaarheid) voor een methodenmix te kiezen, ook wel triangulatie genoemd.

- Het eerste dat moet gebeuren bij het verzamelen van data is vaststellen welke data nodig zijn. Dat betekent veelal een operationalisering van de gebruikte begrippen en variabelen. Een geldige

¹⁷ Het is tevens duidelijk dat kwantitatieve methoden meer voor de hand liggen in de experimentele traditie, en kwalitatieve meer in de contextuele.

operationalisering en dat valt niet mee. Het is al lastig indicatoren voor die begrippen en factoren te vinden, een probleem is vaak dat die indicatoren ook indicatoren van iets anders kunnen zijn. Een voorbeeld. Een indicator voor welstand kan zijn het aantal ramen in een huis (hoe meer ramen, des te rijker; het is ooit gebruikt bij de belastingheffing), maar het aantal ramen kan ook een indicator zijn voor een bouwperiode of voor de behoefte aan openheid dan wel beslotenheid.

Als een beslissing is genomen over de indicatoren, dienen ze te worden verzameld. Daarvoor gelden vier eisen:

- Validiteit: meet je wat je wilt meten;
- Betrouwbaarheid: de consistentie van de meting (meet je steeds hetzelfde, de herhaalbaarheid van de meting);
- Precisie: in data en bronnen, beschrijving ervan;
- Werkbaarheid: zo gering mogelijke belasting van betrokkenen.

b. Enkele kwantitatieve methoden

-Men kan veel mensen zien te bereiken door schriftelijke vragenlijsten rond te sturen, per brief of per e-mail, dan wel telefonisch een aantal vragen te stellen. De inhoud zal beperkt zijn, wil men de respondenten niet te zeer belasten en de respons redelijk te laten zijn. De respons lijkt minder goed te zijn dan vroeger. Om te bepalen welke waarde men aan de respons mag hechten, zal een analyse van de non-respons nodig zijn. Belangrijk is ook dat vragenlijsten naar de juiste mensen worden gestuurd. Een steekproef van personen uit het neutraal ogende telefoonboek¹⁸ of uit het bevolkingsregister op de neutraal ogende geboortedatum¹⁹ kan niet a-select zijn en dus vertekeningen opleveren

Bij dergelijke questionnaires is de vraagstelling is van groot belang, want mensen zijn geneigd de vrager ter wille te zijn. Ik heb meegemaakt dat een onderzoek waar werd gevraagd naar de tevredenheid over de politie een even hoge uitkomst gaf als een vrijwel gelijktijdig uitgevoerd onderzoek naar de ontevredenheid over de politie. De formulering van de vragen bepaalde waarschijnlijk de uitkomst.

Voor een indruk van het gedrag van mensen is een questionnaire, zeker een schriftelijke, minder geschikt. Mensen veranderen de waarheid makkelijk, men vergeet gemakkelijk, de kans bestaat dat mensen zeggen wat ze zouden (moeten) doen, niet wat ze doen; een risico bij dergelijke vragenlijsten is dat men foute of sociaal wenselijke antwoorden krijgt.

-Kosten-batenanalyses zullen vooral worden gebruikt voor ex ante analyses (in geval van ex post evaluatie heet het al snel een doeltreffendheidsonderzoek) en zijn veelal kwantitatief van aard. Gebruik wordt vaak gemaakt van historische gegevens, al dan niet in combinatie met simulatiemodellen. Kba's zijn er in vele soorten en maten. Het voordeel is dat onvergelykbare grootheden via deze vertaling met elkaar kunnen worden vergeleken, hetgeen tevens het nadeel is. Immers, in de vertaling van vage kosten of baten in financiële eenheden ligt de nodige normatieve

¹⁸ Specifieke groepen bleken geen telefoon of een geheim nummer te hebben.

¹⁹ In kringen van migranten worden soms geboortedata gekozen: 1 januari, 1 februari, 1 augustus.

waardering besloten. Een probleem blijft altijd welke kosten en baten worden meegenomen. In feite wordt een beleidstheorie ontworpen, terwijl niet altijd duidelijk zal zijn waarop die theorie is gebaseerd. De harde cijfers zijn, met andere woorden, zo hard nog niet.

c. Enkele kwalitatieve methoden

- Een specifieke en regelmatig gebruikte methode is die van de Delphi rondes, waarbij in verschillende rondes met experts wordt gesproken om zo tot accumulatie van kennis en inzicht te komen. Het is geen empirisch onderzoek al is het mogelijk en wenselijk dat de experts empirische kennis inbrengen. Deze methode ligt wat meer voor de hand bij ex ante evaluaties, maar kan op zich ook worden toegepast bij ex post evaluaties.

- Bij ex ante evaluaties kan gebruik worden gemaakt van scenariostudies. Voor de toekomst worden enkele mogelijke scenario's uitgewerkt, waarbij zogenoemde robuuste factoren kunnen blijken te bestaan, factoren die in meer of alle scenario's een rol spelen. Het is zinvol daaraan beleidsmatig aandacht te besteden. De beleidsmakers kunnen opteren voor het verder werken met het meest wenselijke scenario. Hard zijn de uitkomsten niet, maar als men het adagium wil omarmen 'de toekomst, die maak jezelf, dan biedt een scenario zeker steun. Mocht dat cynisch klinken, dan klopt dat; maar dat neemt niet weg dat het werken vanuit een oogmerk duidelijkheid en zekerheid verschaffen, hetgeen van waarde is.

Ook het gebruik van spelsimulaties zal vooral van nut zijn bij ex ante evaluaties.

- Interviewen van betrokken personen geeft eveneens inzicht in de mogelijke werking of de feitelijke werking van wetgeving. Belangrijk daarbij is in de eerste plaats de keuze van de te interviewen personen: de burgemeester van een gemeente heeft andere belangen, kennis, opinies dan zijn medewerkers op de werkvloer. Om het hard te zeggen: de burgemeester heeft meestal geen idee van wat zich daar afspeelt. Gesprekspartners hebben belangen bij bepaalde standpunten (en uitkomsten) en zijn deels afhankelijk van hun eigen waarnemingen en herinneringen; daar kan de nodige bias en daar kunnen de nodige hiaten in zitten. Mede daarom is een goed interview afnemen erg moeilijk, en zijn veelal meer interviewers per gesprek nodig. In geval van diepte-interviews is het kostbaar en soms lastig veel personen te interviewen; het aantal bepaalt de hardheid van de inzichten en uitspraken die op grond daarvan kunnen worden gedaan.

Klein Haarhuis (2014) constateerde dat in ex ante onderzoek het meest gebruik wordt gemaakt van kwalitatieve methoden, in het bijzonder interviews en deskresearch.

Geraadpleegde literatuur

Aeken, K. van, Proeven van wetsevaluatie, diss. Antwerpen, Antwerpen 2002.

Bakker, R., Evaluatie als fundering voor het beleid, Den Haag (Ministerie van BZK) 2012.

Dekker, P.J., F.L. Leeuw, Beleids- en programma-evaluaties, Leiden (DSWP) 1989.

Evaluatiemethoden- een introductie, Rapport van de Afd. Beleidsanalyse, Den Haag (Ministerie van Financiën) 1986-3.

Jong, de, P.O., M. Herweijer, Alle regels tellen, Den Haag (WODC rapport 225) 2004.

Jonkers, P., Diskwalificatie van wetgeving, Amsterdam (Aksant) 2003.

Klein Haarhuis, C.M., m.m.v. S.H.C. Keulemans, Ex ante onderzoek in meta perspectief, Den Haag (WODC O&B 311, Boom) 2014.

Leeuw, F.L. & W.M. de Jongste, Beleids- en wetsevaluaties: trends en topics, RegelMaat 2006/3, p. 83-95.

Leeuw, F.L., Benefits and Costs of Evaluation, in: Zs. F. Evaluation, jg.9, Heft 2, 2010, p. 211-227.

Lenters, H., De rol van de rechter in de echtscheidingsprocedure, Arnhem 1993.

Levitt, S., S.J. Dubner, Superfreakonomics, Amsterdam (De Bezige Bij) 2010.

Pawson, R., N. Tilley, Realistic Evaluation, London etc. (Sage) 2008-11.

Schrijvershof, C., K. Douma, L. Aarts, Uitvoeringsanalyses in de justitiële ketens, Den Haag (Ape/WODC) 2013.

Slob, M., J. Staman, Beleid en het bewijsbeest, Den Haag (Rathenau Instituut) 2012.

Swanborn, P.G., Methoden van sociaal-wetenschappelijk onderzoek, Meppel/Amsterdam (Boom) 1982-2.

Swanborn, P.G., Evalueren, Amsterdam (Boom) 1999.

Veerman, G.J. & C.M. Klein Haarhuis, Negen aanwijzingen voor wetsevaluatief onderzoek, RegelMaat 2009/4, p. 215-229.

Veerman, G.J., Evidence en privaatrechtelijke wetgeving, in: *Civilologie: opstellen over empirie en privaatrecht*, Den Haag (Boom) 2012, p. 53-70.

Veerman, G.J., m.m.v. D. de Kok en B.J. Clement, *Over wetgeving- principes, paradoxen en praktische beschouwingen*, Den Haag (Sdu) 2012-3, ihb h.12.

Veerman, G.J., m.m.v. R.J.Mulder & E.S.M. Meijsing, *Een empathische wetgever*, Den Haag (Sdu) 2013.

Winter, H.B., *Evaluatie van Wetgeving, Preadvies Vereniging voor de vergelijkende studie van het recht van België en Nederland*, Deventer (Kluwer) 2002.

Een risico bij dergelijke vragenlijsten is dat men de sociaal wenselijke antwoorden krijgt.

Over de schrijver

Gert Jan Veerman (1947) studeerde rechten aan de Vrije Universiteit te Amsterdam en promoveerde daar in 1977 bij prof.mr. P.H. Kooijmans en prof.dr. A.J. Hoekema op het proefschrift 'Het zelfbeschikkingsrecht der naties en de rechten van de mens'. Hij werkte op de ministeries van Binnenlandse Zaken en Justitie bij beleids- en wetgevingsdirecties: Constitutionele Zaken, Minderhedenbeleid, Staats- en Strafrecht, Politie, Wetgeving, en werkte van 1983-1991 bij het WODC als raadadviseur voor het wetgevings- en rechtsplegingsonderzoek. Bij de Directie Wetgeving van het Ministerie van (Veiligheid en) Justitie was hij betrokken bij het wetgevingskwaliteitsbeleid en was hij eerst hoofd van het Kenniscentrum Wetgeving (2001-2008) en vervolgens van het Clearing House voor Wetsevaluatie (2009-2012). In 2003 werd hij tevens hoogleraar aan de Universiteit Maastricht, met als leeropdracht 'Wetgeving en de kwaliteit van wetgeving'. Hij publiceerde diverse boeken en artikelen, onder meer over wetgeving en evaluatie van wetgeving:

De lengte van de wetgevingsprocedure, WODC rapport 70, Den Haag (Sdu) 1986.

Wetsevaluatief onderzoek- op zoek naar een beleid, RegelMaat 1988/3.

Ik zal eens even vragen naar zijn naam - voor- en nadelen van een legitimatieplicht, m.m.v. G. Paulides, WODC rapport 89, Arnhem (Gouda Quint) 1989.

Over regels en appèl – de conflictopwekkendheid van art. 57 A.A.W., i.s.m. H.R. Schimmel, WODC rapport 102, Arnhem 1990.

De paradox van de wetsevaluatie, JV 1991/5 .

Over jaarverslagen en wetsevaluatie, WODC, Den Haag 1992.

O tempora! o mores! Over wet en normbesef, in: Om de norm, Lelystad (Vermande) 1993, p. 3-12.

Het meesterschap van de wetgevingsjurist, Den Haag (Sdu) 2001.

De wet als zinsbegoochelingstoestel, oratie UM, Maastricht (UM) 2004.

Over wetgeving - principes, paradoxen en praktische beschouwingen, m.m.v. S. de Lange (1^{ste} en 2^{de} druk) en D. de Kok en B. Clement (3^{de} druk), Den Haag (Sdu) 2007; 2009-2 en 2012-3.

Negen aanwijzingen voor wetsevaluatief onderzoek, i.s.m. C.M. Klein Haarhuis, RegelMaat 2009/4, p. 215-229.

Een Grondwet met beleid, in: Brieven aan de staatscommissie grondwetsherziening, Nijmegen (Wolf) 2009, p. 17-23.

Wetgeving met Beleid, i.s.m. R. Mulder, Den Haag (Boom) 2010.

Belle en het Beest – over wetgeving en beleid, in: La Heukers, Den Haag 2011.

Vorbij de horizon - over experimenteer- en horizonbepalingen, ism S. Bulut, Ts voor Constitutioneel Recht 2011/3.

‘Om den algemeenen gang des bestuurs en der wetgeving met kalmte en nauwkeurigheid gade te slaan’ of: de toetsing van wetgeving door de Raad van State, in: De Raad van State in perspectief, Den Haag (Boom) 2011, p. 191-210.

Evidence en privaatrechtelijke wetgeving, in: Civilologie: opstellen over empirie en privaatrecht, Den Haag (Boom) 2012, p. 53-70.

Parlement en wetgeving- een ‘redelijk revolutionair’ voorstel, in: Kringen in de Hofvijver, liber amicorum voor Joop van den Berg, 2012, p. 227-232.

Een empathische wetgever, m.m.v. R.J. Mulder & E.S.M. Meijsing, Den Haag (Sdu) 2013.