Wetenschappelijk Onderzoek- en
Documentatiecentrum
Ministerie van Veiligheid en Justitie

Evaluatievermogen bij beleidsdepartementen

Praktijken rond uitvoering en gebruik van ex post beleids- en wetsevaluaties

C.M. Klein Haarhuis

m.m.v. A. Parapuf

Inhoud

Afkortingen
Samenvatting

Hoofdstuk 1 Inleiding en opzet

1.1 Achtergrond en doel van dit onderzoek
1.2 Onderzoeksvragen
1.3 Conceptueel kader: evaluatievermogen in context
1.4 Aanpak
1.5 Opbouw van dit rapport

Hoofdstuk 2 Internationale ervaringen met evaluatievermogen

2.1 Structuur
2.1.1 Evaluatie-instituties (organisatieonderdelen)
2.2 Programmering en evaluatieverplichtingen
2.3 Middelen
2.4 Evaluatiekader
2.4.1 Standaarden en programma's
2.4.2 Evaluatiedesigns
2.5 Borging van evaluatiekennis en –ervaring
2.6 Gebruik
2.7 Werkt evaluatievermogen?
2.8 Samenvatting

Hoofdstuk 3 Praktijkervaringen met evaluatievermogen bij beleidsdepartementen

3.1 Evaluatiebeleid en wetgevingskwaliteitsbeleid
3.1.1 Verantwoording binnen het rijksbrede evaluatiestelsel
3.1.2 Wetgevingskwaliteitsbeleid en wetsevaluaties
3.2 Evaluatiestructuur
3.2.1 Doelen van evaluatieonderzoek
3.2.2 Organisatieonderdelen met evaluatietaken
3.2.3 Samenvattend over structuur
3.3 Evaluatieprogrammering
3.3.1 Beleidsdoorlichtingen
3.3.2 (Evaluatie)onderzoek gericht op kennisvergroting
3.3.3 Evaluatiebepalingen in wetten
3.3.4 Politieke context en evaluatieprogrammering
3.3.5 Samenvattend over programmering
3.4 Middelen
3.5 Evaluatiekader
3.5.1 Evaluatieprogramma's specifiek gericht op wetgeving
3.5.2 Evaluatieprogramma's algemeen
3.5.3 Handreikingen, richtlijnen, standaarden
3.5.4 Samenvattend over evaluatiekader
3.6 Borging van evaluatiekennis en –ervaring
3.6.1 Communicatie van het evaluatiekader en cursussen
3.6.2 Selectie van onderzoekers
3.6.3 Begeleidingscommissies
3.6.4 Delen en verankeren van ervaringen met evaluatie (aanvragers)

3.6.5 Samenvattend over borging
3.7 Bruikbaarheid en gebruik van (wets)evaluaties
3.7.1 Communicatie tussen aanvrager en onderzoeker
3.7.2 Wijze van presentatie, aanbevelingen
3.7.3 Gebruik
3.7.4 Leeropbrengsten
3.7.5 Samenvattend over bruikbaarheid en gebruik

Hoofdstuk 4 Slotbeschouwing

4.1 Inleiding
4.2 Internationale ervaringen met evaluatievermogen (onderzoeksvraag 1)
4.3 Praktijken en ervaringen bij Nederlandse beleidsdepartementen (onderzoeksvragen 2 en 3)
4.3.1 Structuur: organisatieonderdelen voor evaluatie
4.3.2 Evaluatieprogrammering
4.3.3 Middelen
4.3.4 Evaluatiekader: inhoud en proces
4.3.5 Borgen en verankeren van evaluatiekennis en –ervaring
4.3.6 Bruikbaarheid en gebruik
4.4 Evaluatievermogen in context
4.4.1 Politiek en verantwoording
4.4.2 Evaluatiebepalingen in wetten
4.4.3 Organisatiecontext
4.5 Wetsevaluaties.. een vak apart?
4.6 Slotconclusies (onderzoeksvraag 4)

Summary
Literatuur

Bijlagen
1. Begeleidingscommissie
2. Respondenten en interviews
3. Internationale ijkpunten
4. Tijdlijn met ontwikkelingen in en om het centrale evaluatiestelsel & Praktijken/leidraden binnen beleidsdepartementen

Afkortingen

AEA	*American Evaluation Association*
AEP	Algemene Economische Politiek (Directie bij Economische Zaken)
Anw	Algemene nabestaandenwet
AR	Algemene Rekenkamer
Awb	Algemene wet bestuursrecht
BAT	Beleidsanalistenteam (opvolger van de Regiegroep M&E bij Economische Zaken)
BC	Begeleidingscommissie
BEC	Beleidskwaliteit- en Evaluatiecommissie (Economische Zaken)
Bopz	Wet bijzondere opnemingen in psychiatrische ziekenhuizen
BVO	Bureau Verkenning en Onderzoek (vallend onder DG Bestuur en Koninkrijksrelaties bij het ministerie van BZK
BZ	Ministerie van Buitenlandse Zaken
BZK	Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
CCV	Centrum voor Criminaliteitspreventie en Veiligheid
CER	Commissie Evaluatie Regelgeving
CES	*Canadian Evaluation Society*
CIE	*Counterfactual Impact Evaluations*. Om onwtikkelingen op de doelvariabele te kunnen attribueren aan beleid, wordt een *counterfactual* (bijv. controlegroep) betrokken: wat zou de situatie zijn geweest *zonder* het beleid?
CPB	Centraal Planbureau
CROW	Kenniscentrum voor verkeer, vervoer en infrastructuur
CSO	*Chief Science Officer* (SZW)
Cw	Comptabiliteitswet
DFEZ	Directie Financieel-Economische Zaken
DG	Directoraat-Generaal
DG B&I	Directoraat-Generaal Bedrijfsleven en Innovatie (Economische Zaken)
DGIS	Directoraat-Generaal Internationale Samenwerking
DWJZ	Directie Wetgeving en/of Juridische Zaken (bij de meeste departementen zo geheten)
EC	Europese Commissie
ECB	*Evaluation Capacity Building*
EES	*European Evaluation Society*
EFRO	Europees Fonds voor Regionale Ontwikkeling
EPTF	*Evaluation Policy Task Force* van de *American Evaluation Association*
ESF	Europees Sociaal Fonds
EWB	Extern-wetenschappelijke Betrekkingen (afdeling van WODC bij VenJ)
EZ	Ministerie van Economische Zaken
FEZ	(Directie) Financieel-Economische Zaken
FMC	Financiën, Management en Control ('FEZ' IenM)
GAO	General Audit Office
HBJZ	Hoofddirectie Bestuurlijke en Juridische Zaken (IenM)
IAK	Integraal afwegingskader voor beleid en wetgevin g
IBP	Interdepartementale Begeleidingsgroep Prestatiegegevens en Beleidsevaluatie
IEG	Independent Evaluation Group (Wereldbank)
IenM	Ministerie van Infrastructuur en Milieu
IOB	Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (IOB)
IOFEZ	Interdepartementaal overlegorgaan Financieel-economische Zaken
IOWJZ	Interdepartementaal overlegorgaan Wetgeving en Juridische Zaken
IUC	Inkoopuitvoeringscentra (per 2014)

KiM	Kennisinstituut voor Mobiliteitsbeleid (Ministerie van Infrastructuur en Milieu)
KCWJ	Kenniscentrum Wetgeving en Juridische Zaken
M&E	Monitoring en Effectmeting (o.a. voorm. Regiegroep bij EZ) of: *Monitoring & Evaluation*
MKBA	Maatschappelijke Kostenbatenanalyse
MvT	Memorie van Toelichting
MT	Managementteam
NAO	National Audit Office (Rekenkamer Verenigd Koninkrijk)
NLR	Netherlands Aerospace Centre
NRO	Nationaal Regieorgaan Onderwijsonderzoek
NWO	Nederlandse Organisatie voor Wetenschappelijk Onderzoek
OBPR	Office of best practice Regulation (Australië)
OCW	Ministerie van Onderwijs, Cultuur en Wetenschappen
OER	Overzicht Doeltreffendheidsonderzoek Rijksoverheid (Algemene Rekenkamer)
OESO	Organisatie voor Europese Samenwerking en Ontwikkeling
PBL	Planbureau voor de Leefomgeving
PER	Programma Evaluatie Regelgeving (ZonMw) – op het terrein van gezondheidsrecht
PIR	*Post Implementation Review* (Verenigd Koninkrijk en Australië)
RCT	*Randomised controlled trials* (gerandomiseerd experiment)
REFIT	*Regulatory Fitness* programma (EU)
Regiegroep M&E	(voorm.) Regiegroep Monitoring & Effectmeting (EZ). Vanaf 2016 Beleidsanalistenteam (BAT)
RIA / RIS	*Regulatory Impact Assessment / Statement*
RIVM	Rijksinstituut voor Volksgezondheid en Milieu
RVO	Rijksdienst voor Ondernemend Nederland
RPE	Regeling Periodiek Evaluatieonderzoek
RWS-WVL	Rijkswaterstaat – landelijke dienst Water, Verkeer en Leefomgeving
SC	*Steering Committee* (EU), adviserende stuurgroep of begeleidingscommissie
SCP	Sociaal en Cultureel Planbureau
SKIA	Strategische Kennis- en Innovatie-agenda (SZW)
SKO	Strategie Stelsel en Onderzoek (directie SZI bij SZW)
SUWI	Wet Structuur Uitvoeringsorganisatie Werk en Inkomen
SWOV	Stichting Wetenschappelijk Onderzoek Verkeersveiligheid
SZI	Sociale Zekerheid en Integratie (directie bij SZW)
SZW	Ministerie van Sociale Zaken en Werkgelegenheid
TK	Tweede Kamer
UKES	*United Kingdom Evaluation Society* (evaluatiegemeenschap Verenigd Koninkrijk)
UNEG	*United Nations Evaluation Group*
VBO	Vereniging voor Beleidsonderzoek (Nederland)
VenJ	Ministerie van Veiligheid en Justitie
VU	Vrije Universiteit (Amsterdam)
VWS	Ministerie van Volksgezondheid, Welzijn en Sport
WAO	Wet arbeidsongeschiktheid
WKB	Directie Wetgevingskwaliteitsbeleid van het Kenniscentrum Wetgeving
WODC	Wetenschappelijk Onderzoek- en Documentatiecentrum, VenJ
Wro	Wet ruimtelijke ordening

Samenvatting

Inleiding en vraagstelling

In de borging van wetgevingskwaliteit spelen evaluaties een belangrijke rol. Willen deze daadwerkelijk een bijdrage leveren aan de kwaliteit van wetgeving, dan is het zaak om: (1) te zorgen dat ze worden verricht; (2) ze zó op te zetten dat ze bruikbare, valide en betrouwbare inzichten genereren; en (3) dat ze worden gebruikt door beleidsmakers en wetgevingsjuristen in de terugkoppeling naar hun producten.
Dit is in de praktijk echter niet vanzelfsprekend. Wetten en wetswijzigingen zijn doorgaans complexe evaluatieobjecten: een mix van regels, normenkaders en beleidsinterventies. Behalve empirische en beleidsevaluatie-vaardigheden vereisen wetsevaluaties dus ook een juridische blik. Veel individuele aanvragers zeggen niet goed te weten wat te doen als zij een wetsevaluatie (proces- of ex post) moeten gaan opleveren. Deze factoren kunnen eraan hebben bijgedragen dat wetsvaluaties in de praktijk heel verschillend worden aangepakt: zowel qua inrichting van het aansturingsproces als methodologisch-inhoudelijk. Een centraal wetsevaluatiebeleid is er niet. Dit laatste hoeft niet problematisch te zijn, ware het niet dat eerder in meta-onderzoek ook aanwijzingen zijn gevonden voor verschillen in de validiteit en betrouwbaarheid van wetsevaluaties.
De gevarieerde praktijk van wetsevaluatie houdt mogelijk verband met verschillen in *evaluatievermogen*: het vermogen om wetsevaluaties te verrichten en ze bovendien te gebruiken. In hoeverre zijn binnen beleidsdepartementen de voorwaarden aanwezig om het doen en gebruiken van evaluaties mogelijk te maken? Die voorwaarden kunnen zowel het evaluatieproces als de inhoud van evaluaties betreffen.
Doel van deze studie is om een beeld te krijgen van bestaande departementale praktijken rond het verrichten en gebruiken van evaluaties – met name wetsevaluaties. Aan de basis ligt een streven bij de aanvrager van dit onderzoek, de directie Wetgevingskwaliteitsbeleid (WKB) van het Kenniscentrum Wetgeving en Juridische Zaken (KCWJ), naar meer vergelijkbaarheid tussen wetsevaluaties. Dit om te zorgen dat ervaringen *met* en kennis *uit* wetsevaluaties worden geaccumuleerd om bredere lessen te kunnen trekken.

Onderzoeksvragen en aanpak

1. Welke inzichten bieden de internationale (a) literatuur en (b) normen en handreikingen met betrekking tot *evaluatievermogen*? (hoofdstuk 1 en 2)
2. Welke praktijkervaringen met evaluatievermogen zijn opgedaan door beleidsmedewerkers en wetgevingsjuristen, betrokken bij (wets)evaluaties? (hoofdstuk 3)
3. Welke inhoudelijke en methodologische eigenschappen van *ex post*-evaluaties en -wetsevaluaties acht men in de praktijk van belang? (hoofdstuk 3)
4. Welke lessen bieden de antwoorden op voorgaande onderzoeksvragen voor verdere gedachtevorming over de ontwikkeling van evaluatievermogen – vooral met betrekking tot wetgeving? (hoofdstuk 4)

Ruggengraat van dit onderzoek vormden de volgende aspecten van evaluatievermogen en hun kenmerken – gebaseerd op literatuur over evaluatievermogen:

- Structuur: doelstellingen en organisatieonderdelen voor evaluatie
- Programmering en evaluatieverplichtingen
- Middelen voor evaluatie: financiële en andere

- Evaluatiekader: evaluatieprogramma's en richtlijnen bij de aanpak (proces en inhoud)
- Borging van evaluatiekennis en –ervaring
- Bevordering van gebruik

In deze structuur, waarmee een deel van onderzoeksvraag 1 is beantwoord, zijn eerst een aantal aanpakken en ervaringen van internationale organisaties en grotere landen op een rij gezet. Vervolgens zijn, aan de hand van *desk study* en een twintigtal gesprekken met 35 sleutelinformanten, praktijkgegevens verzameld bij negen Nederlandse beleidsdepartementen.
Evaluaties en wetsevaluaties vinden niet in een vacuüm plaats; ze worden verricht vanuit verschillende motieven. Soms zijn ze verplicht uit oogpunt van verantwoording richting de Tweede Kamer (bijv. zijn begrotingsgelden goed besteed?) of een wettelijke evaluatiebepaling, een andere keer komen ze van binnenuit, bijvoorbeeld om te leren met het oog op verdere beleidsontwikkeling. Ook kunnen evaluaties strategisch worden ingezet. In de literatuur worden twee ideaaltypische evaluatiemotieven onderscheiden: (a) verantwoorden en (b) op 'leren' gerichte motieven zoals kennisvergroting of *enlightenment* en eventuele beleidsaanpassing of *embetterment*. Bij het eerste motief passen eindevaluaties gericht op de vraag 'heeft het gewerkt', bij het tweede passen tussentijdse evaluaties, gericht op het begrijpen van de uitwerking en eventuele bijsturing. De politieke context bepaalt dus mede de motieven voor evalueren, die kunnen doorwerken in hoe evaluaties worden opgezet, aangestuurd en gebruikt. Uiteindelijk werkt ze ook door in evaluatievermogen.

Internationale ervaringen met evaluatievermogen (Onderzoeksvraag 1)

Uit internationale handboeken van bijvoorbeeld de VN, VS, VK, Wereldbank en de Europese Commissie (EC) blijkt dat beleidsmedewerkers een steeds belangrijker rol krijgen toebedeeld in het managen van evaluaties. De kwaliteit en het gebruik van evaluatieonderzoek zijn niet alleen afhankelijk van hoe onderzoekers werken en hun resultaten presenteren, maar juist ook van de aanvragers, zo stelt men.
In de eerste plaats nemen die aanvragers bepalende beslissingen: over evaluatieprogrammering, budget, reikwijdte en onderzoeksvragen. Ook hebben ze vaak een rol in de verdere aansturing of begeleiding van de evaluatie. Er circuleren uiteenlopende normen en leidraden ten aanzien van opdrachtformulering, aanbesteding, begeleiding, eindbeoordeling en het gebruik van evaluaties, waarvan enkele ter illustratie zijn opgenomen in Bijlage 3 bij dit rapport.
Niet alleen het evaluatietraject zelf, maar ook de inhoud en opzet van te evalueren beleid of wetgeving en de onderbouwing daarvan zijn medebepalend voor evaluatiekwaliteit. Evalueren wordt steeds vaker beschouwd als onderdeel van de beleidscyclus, en minder als een activiteit die daar los van staat. Ontbreekt bijvoorbeeld de argumentatie voor beleidskeuzes (beleidstheorie), dan wordt het lastiger om een evaluatie te focussen. Dit zou volgens de EC het evalueren zelfs duurder kunnen maken. Ook in het monitoren van beleid vervult de (latere) evaluatie-aanvrager een cruciale rol. Meerdere instituties vinden het een gemiste kans voor hoogwaardige evaluaties als niet, of niet op tijd, wordt begonnen met nulmetingen en de verdere monitoring van belangrijke proces- en uitkomstindicatoren. Ook doelgroepen, zoals gemeenten, burgers en bedrijven kunnen in de monitoring een rol vervullen.

Ook gidsen voor het design van evaluaties, de inhoudelijke evaluatie-aanpak, zijn vaak niet alleen voor onderzoekers, maar ook voor aanvragers bedoeld. Hierin zijn verschillende stromingen te onderscheiden. Illustratief is het onderscheid tussen de attributieve (black box) en de 'contributieve' (clear box) stroming. In de eerste

stroming staat het kunnen toeschrijven van resultaten aan beleid centraal: heeft het beleid gewerkt en in welke mate? Met de designs in deze stroming wordt gepoogd de effecten van het beleid of de wet in kwestie zo goed mogelijk te isoleren van andere, om tot een gerichte, netto-uitspraak over doeltreffendheid te komen. In de praktijk komt het dan vaak neer op experimentele designs. De aanpakken in de tweede stroming zijn gericht op het begrijpen van hoe en waarom – onder welke condities – veranderingen zijn opgetreden en resultaten zijn geboekt. De stromingen sluiten elkaar niet uit, ze kunnen wederzijds aanvullend worden gebruikt. Terugkerend uitgangspunt bij de keuze voor een benadering is *fit for purpose*: keuzes in het evaluatiedesign worden afhankelijk gesteld van de situatie, o.a. soort beleid, beleidsfase, evaluatieonderwerp, hoofdvragen en soort doelgroep. In dit licht ontwikkelde onder meer de Amerikaanse federale Rekenkamer handgrepen voor de evaluatie van complexe federale programma's. Net als Nederlandse wetten zijn ook deze meestal geen 'cleane' interventies waarvan de werking eenvoudig gemeten en bovendien geïsoleerd kan worden van de context, zoals externe invloeden en parallel beleid. Met wat voor soort evaluatiedesigns valt dan toch het maximale te zeggen?

Naast het evalueren als zodanig beschouwen grote instituties en landen ook het evaluatiegebruik steeds meer als onderdeel van het beleidsproces en niet slechts als resultante van een eindrapport. Zo voorziet het *Smart Regulation*-programma van de EC in terugkoppeling van ex post evaluatiebevindingen naar de regelgevingscyclus, via de verplichte ex ante evaluaties waarin die bevindingen een plek dienen te krijgen. Ook bij de VN wordt evaluatiegebruik voorgeschreven en gemonitord; in dat verband stimuleert men ook het aanleggen van metastudies en kennisbanken om het overzicht over evaluatiekennis te behouden. Uit recent internationaal onderzoek blijkt dat op landenniveau het systematisch gebruiken van evaluaties lastig te bewerkstelligen. Wel spelen evaluatiegemeenschappen en parlementen zeer waarschijnlijk een stimulerende rol in de totstandkoming en het gebruik van evaluaties.
Behalve de aanvrager speelt de evaluatieonderzoeker nog steeds een belangrijke rol in de bevordering van gebruik. Dit blijkt bijvoorbeeld uit leidraden om bevindingen toegankelijk en bruikbaar te presenteren. Ook blijkt het uit competentiebeschrijvingen van onderzoekers: niet alleen methodologisch-inhoudelijke vaardigheden worden belangrijk gevonden, maar ook interpersoonlijke en communicatieve vaardigheden, flexibiliteit en een gevoelige antenne voor de (politieke en beleids-)context.
Hoewel internationaal veel is verschenen over het bevorderen van uitvoering en gebruik van evaluaties, is nog niet aangetoond dat dit een positieve uitwerking heeft op de doeltreffendheid van beleid of wetgeving. Ook is gewaarschuwd voor ongewenste effecten van evaluatievermogen. Zo blijkt dat het ontwikkelen en handhaven van standaarden opportunistisch gedrag of ritualisering in de hand kan werken.

Praktijkervaringen met evaluatievermogen bij beleidsdepartementen (onderzoeksvragen 2 en 3)

Beleidsdepartementen hebben heel verschillend vorm gegeven aan evaluatievermogen, zowel in termen van de betrokken organisatieonderdelen als in termen van procesmatige en inhoudelijke evaluatiekaders.
Het verantwoorden van begrotingsuitgaven blijkt een belangrijk doel van de evaluaties bij (rijks)beleidsdepartementen. Veruit het belangrijkste verantwoordingsinstrument is de beleidsdoorlichting op begrotingsartikelniveau, die volgens de Regeling Prestatiegegevens en evaluatieonderzoek (RPE) elke vier tot maximaal zeven jaar dient plaats te vinden en aan de Kamer moet worden

gestuurd. Voor de inhoud van de beleidsdoorlichtingen, een synthesedocument, zijn beleidsdirecties verantwoordelijk, de directie financieel-economische zaken (FEZ) heeft bij elk departement een coördinerende en bewakende rol. De beleidsdoorlichting werkt door in de ambtelijke evaluatiepraktijken waar beleidsevaluaties plaatsvinden: idealiter stapelt zich hier in de loop van elke 4 tot 7 jaar het benodigde evaluatiemateriaal op.

Organisatieonderdelen gericht op evalueren
Evaluaties worden voor het overgrote deel uitbesteed aan externe partijen, vanwege interne capaciteitsrestricties en vermeende objectiviteit. Binnen beleidsdepartementen vervullen verschillende organisatieonderdelen een rol in het managen van evaluaties. Doorgaans ligt het initiatief voor ex post evaluaties van wetgeving en beleid bij de verantwoordelijke onderzoekscoördinatoren van beleidsdirecties of DG's. Geen van de onderzochte departementen heeft een aparte evaluatieafdeling – afgezien van onderzoeksinstituten die ook evaluaties doen, zoals het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) bij VenJ. Wel zijn er steeds een of meerdere stafafdelingen, units of directies die de evaluatiefunctie ondersteunen en stimuleren. Een gemeenschappelijke noemer is de rol in deze van de directie Financiële en Economische Zaken (FEZ). Wat evalueren betreft, vullen directies FEZ hun rol per departement verschillend in – hoewel verantwoording uiteraard wel steeds het hoofddoel is. Zo heeft FEZ bij SZW een stimulerende en aanjagende rol in planning, kwaliteitsborging en gebruik van evaluaties en beleidsdoorlichtingen, terwijl andere directies FEZ dichter op hun formele controletaken blijven. Binnen EZ is er sinds 2012-2013 een netwerk van organisatieonderdelen inclusief FEZ, met een Beleidskwaliteit- en Evaluatiecommissie (BEC) en een Regiegroep monitoring & evaluatie (M&E).[1] Dit netwerk richt zich specifiek op de kwaliteit van doeltreffendheidsevaluaties.
Anders dan FEZ dragen kennisdirecties en -organen zorg voor meer strategisch en toekomstgericht onderzoek. In de praktijk maakt evaluatieonderzoek in zeer beperkte mate deel van uit van het kennisgeoriënteerde werk. Dit onderscheid tussen directies FEZ en kennisdirecties is echter niet zwart-wit. Neem bijvoorbeeld de rol die het Centraal Planbureau (CPB) en een enkele kennisdirectie hebben gespeeld in de bevordering van quasi-experimentele doeltreffendheidsstudies; en omgekeerd de rol die de *Chief Science Officer* (CSO) bij SZW vervult in relatie tot evaluatie- en onderzoeksinitiatieven.
Een aantal van de organisatieonderdelen en -structuren zijn nog vers en vereisen een wenperiode: het is zoeken naar een rolverdeling die niet alleen op papier, maar ook in de praktijk werkt. Een uitzondering is de Commissie Evaluatie Regelgeving (CER) van ZonMw, die al langer actief is voor een deel van de *wets*evaluaties in het domein van met name VWS.

Programmering en middelen
In elke jaarlijkse departementale begroting wordt een evaluatieplanning opgenomen. Eventuele extra of parallelle onderzoeksprogrammeringen, met name strategische kennisagenda's, bevatten nauwelijks ex post evaluatieonderzoek en zijn vooral toekomstgericht. Programmering vindt ook plaats via (periodieke) evaluatieverplichtingen. 10-20% van de wetten kent een dergelijke verplichting. De resulterende beleids- en wetsevaluaties worden door directies gebruikt als basismateriaal voor beleidsdoorlichtingen.
Een adviespanel bij BZ oordeelde dat de nadruk in de evaluatieprogrammering sterk op verantwoorden ligt – het afdekken van begrotingsartikelen – en minder op 'leren'. Dit lijkt relevant voor meer bestudeerde departementen. De politieke context werkt door in de evaluatieprogrammering. Sommige respondenten relativeren het

[1] De Regiegroep M&E is in 2016 opgegaan in het Beleidsanalistenteam, de BAT.

feitelijke belang van programmering: zij stellen dat evaluaties vooral samenhangen met veranderingen in kabinetssteun voor een bepaald beleid of met beleidsaanpassingen en -ontwikkelingen.
De meeste departementen stellen eerst een integraal budget voor onderzoek vast. Uitgevoerde evaluaties moeten binnen dat financiële kader passen. In de praktijk gaat men vaak uit van de bestedingsplannen. Onderzoeksinstituten beheren vaak hun eigen budget.
Wat de niet-financiële middelen betreft, melden respondenten vaak dat beleidsdirecties weinig tijd hebben voor evaluaties en onderzoek, onder meer blijkend uit beperkte capaciteit voor onderzoekscoördinatie. Dit zet ook de programmering onder druk. Bovendien werkt tijdgebrek in de hand dat men te laat begint aan geplande evaluaties, en belangrijke meetkansen 'mist' – zoals het aanleggen van een monitoringsysteem of het vormen van een controlegroep. Het laatste is een veelvoorkomend euvel.
In de context van verantwoording zijn de afgelopen paar jaar door FEZ en sommige andere organisatieonderdelen initiatieven genomen om de evaluatieprogrammering proactiever te maken, vooral met het oog op aankomende beleidsdoorlichtingen en het meer doordacht onderzoeken van doeltreffendheid/doeltreffendheid en doelmatigheid. We hebben niet overkoepelend kunnen vaststellen wat dit heeft opgeleverd.

Evaluatiekader
Met evaluatiekader doelen we voornamelijk op evaluatie(bevorderende) programma's – die in de praktijk vaak zowel het proces als de inhoud van evaluaties betreffen (onderzoeksvragen 2 en 3).
In het derde hoofdstuk zijn praktijkervaringen met drie speciaal op wetgeving gerichte evaluatieprogramma's de revue gepasseerd, te weten het Programma Evaluatie Regelgeving (PER), de Awb-evaluaties en het programma Structurele Evaluatie Milieuwetgeving (STEM). Deze bevatten lessen over het evaluatieproces (programmering, afbakening hoofdvragen door de aanvrager), het evaluatiedesign en evaluatiecriteria. Een overkoepelende les die uit de praktijk van elk van deze programma's is getrokken, betreft de meerwaarde van een multidisciplinaire aanpak van *wets*evaluaties. Het voldoen aan de juridische randvoorwaarden voor doeltreffendheid betekent nog niet dat de wet in de werkelijkheid ook doel zal treffen – en omgekeerd. In het PER is een gecombineerd juridisch-sociaalwetenschappelijke aanpak de eerste van zeven basisvereisten voor een wetsevaluatie.
Daarnaast zijn een aantal algemene beleidsevaluatieprogramma's besproken, die van toepassing zijn op zowel beleid als wetgeving. Bij OCW is beschreven, hoe de directie Kennis combinaties van *black box*- en *clear-box* (experimentele en verklarende) evaluatiebenaderingen poogt te bevorderen. Bij EZ worden evaluatie-handvatten – in lijn met wat internationaal wordt voorgestaan – geïntegreerd aangeboden in een handreiking voor beleidskwaliteit. Het overkoepelende doel van die handreiking is effectiever beleid bewerkstelligen en één van de middelen is evalueren. De handreiking omvat alle fasen van de beleidscyclus, van probleemanalyse, instrumentkeuze en ex ante gevolgenbeoordeling, tot en met politieke beslissing, uitvoering en – uiteindelijk – ex post evaluatie. Ze bevat ook aanwijzingen en tips bij het opstellen van (a) vroege beleidsexperimenten en (b) het opzetten van een monitoringsysteem.
In de Nederlandse praktijk verschilt de inhoudelijke evaluatiefocus enigszins tussen departementen, vaak passend bij de rol van de meest actieve organisatieonderdelen. Is FEZ relatief actief, dan weerspiegelt zich dat in een meer summatieve, op verantwoording gerichte aanpak, zoals bij EZ en SZW. Bij EZ staat men aanvragers bijvoorbeeld bij met uitleg en checklists voor de opzet van doeltreffendheidsevaluaties (procesmatig en inhoudelijk) en voor de beoordeling van

offertes en rapporten. De ervaring met de Wet Bevordering Speur- en Ontwikkelingswerk (WBSO) laat zien dat een empirische wetsevaluatie volledig kwantitatief kan zijn; het fiscale instrumentarium keert in meerdere wiskundige formules terug. Nauwe betrokkenheid van een directie Kennis of onderzoeksinstituut uit zich vaak in verschillende soorten voorgestane aanpakken, zowel experimenteel (black box) als verklarend. Het onderscheid is dus niet zwart-wit; zo wijzen respondenten van meerdere directies FEZ en ook het ministerie van Financiën ook vaak op toetsing van veronderstellingen in de beleidstheorie.
Het IAK wordt nauwelijks genoemd in het verband van ex post evaluatie; bij IenM probeert men bij ex post beleidsdoorlichtingen terug te grijpen op de resultaten van het ex ante te gebruiken IAK voor beleid en wetgeving.

Borging van evaluatiekennis en –ervaring
De aanvrager wordt ook in ons land steeds vaker gezien als een belangrijke schakel in procesverloop en evaluatiekwaliteit. Hoe is bij beleidsdepartementen geborgd dat aanvragers beschikken over de kennis en ervaring om evaluaties in goede banen te leiden? Bij de meeste DG's en directies vervullen de onderzoekscoördinatoren hierin een spilfunctie en gezien hun relevante expertise en ervaring fungeren zij tevens als vraagbaak voor collega's. Verder verschilt de ondersteuning sterk tussen departementen, afhankelijk van de betrokken organisatieonderdelen. Zo heeft bij SZW de *Chief Science Officer* (CSO) inhoudelijk een ondersteunende rol; bij OCW is dat de directie Kennis en bij EZ de vroegere Regiegroep M&E (inmiddels BAT). Bij VenJ wordt bij een belangrijk deel van de (wets)evaluaties de aanvragersrol voor een deel overgenomen door projectbegeleiders binnen het WODC – de aanvrager maakt wel deel uit van de begeleidingscommissie.
Meerdere departementen en rijksacademies bieden bovendien cursussen of trainingen aan voor beleidsmedewerkers, die deels of geheel over evaluatie gaan. Bij de selectie van onderzoekers wordt vooral gebruik gemaakt van de competitie tussen onderzoeksgroepen in het aanbestedingsproces. Daarnaast zijn een beperkt aantal initiatieven genoemd, zoals het gerichter werken met selectiecriteria en het bijhouden van databases (WODC) met capaciteiten van externe onderzoeksgroepen. Door verloop is het een uitdaging gebleken om dergelijke gegevens up-to-date te houden. Juist bij wetsevaluaties is, zoals aangegeven, een multidisciplinair onderzoeksteam van belang. In de praktijk is die combinatie lang niet altijd waargemaakt – maar er zijn ook tegenvoorbeelden, zoals de vaste procedure bij ZonMw-wetsevaluaties en de latere Awb-evaluatie van versnellingsinstrumenten.
Ook begeleidingscommissies (BC's) worden vaak ingezet ter bewaking van voortgang, inhoud of kwaliteit. Bij WODC en CER-ZonMw is het werken met een BC standaardpraktijk; in geval van WODC-evaluaties wordt altijd een wetenschapper als voorzitter aangewezen, bij ZonMw is dat een CER-lid.
Het uitwisselen van ervaringen op het vlak van evaluatie gebeurt binnen departementen vaak via informele overdracht of in het afdelingsoverleg. Betwijfeld kan worden of dit volstaat in het borgen van evaluatie-ervaring.
Interdepartementaal vindt uitwisseling vooral plaats tussen FEZ-directies en Financiën, in de begeleidingsgroep IBP, wat positief wordt gewaardeerd.
Initiatieven ter versterking van het evaluatievermogen lijken kwetsbaar want dreigen soms in de knel te komen: beleidsmedewerkers zijn er niet klaar voor, of er zijn reorganisaties (zoals bij inkoop), of snel personeelsverloop. Dit bevestigt een observatie uit de internationale praktijk dat evaluatievermogen een kwestie van een lange adem is.

Bruikbaarheid en gebruik
Evaluatiegebruik geeft zin aan alle eerder behandelde aspecten van evaluatievermogen: zonder gebruik vindt geen terugkoppeling naar de beleids- of wetgevingscyclus plaats en rijst de vraag wat het nut van de evaluatie is geweest.

Contact en overleg tussen aanvrager en onderzoekers gedurende het evaluatietraject – al dan niet in een BC – is volgens de literatuur en ook volgens dit onderzoek een cruciale manier om de kans op gebruik te vergroten. Ook timing en tijdigheid in relatie tot de beleids- en besluitvorming zijn dat.
Over het algemeen hanteert elk onderzocht departement een vergelijkbare formele wijze van communicatie en (korte-termijn) kennisname en gebruik van evaluatieonderzoek. In de regel worden rapporten openbaar gemaakt – en vaak ook naar de Tweede Kamer gestuurd – met een beleidsreactie. Door onderzoekers vindt daarnaast vaak nog verspreiding plaats in de vorm van academische en vakpublicaties en via toelichtingen op de departementale werkvloer. Bij BZ is voor 'leren' een adviespanel ingesteld om de bruikbaarheid en terugkoppeling van evaluaties naar beleid te vergroten.
Vervolgens is het aan beleidsmedewerkers en politiek wat er met bevindingen gebeurt. Er zijn bij departementen geen 'verplichte' constructies gevonden voor terugkoppeling naar de beleidscyclus, los van de horizonbepaling in financiële regelingen die stelt dat de regeling na vijf jaar vervalt. Eventuele voorzetting dient te worden gemotiveerd, bijvoorbeeld met onderzoek naar doeltreffendheid en doelmatigheid. Niettemin percipiëren enkele respondenten dat de Tweede Kamer toenemende belangstelling laat zien in ex post evaluaties en in wat de minister daarmee wil gaan doen. Het betreft wel met name de verantwoordingsfunctie en financiële lasten.
Om de leerfunctie van evaluatie een belangrijker plaats te geven, zouden ex post-evaluaties volgens sommige gebruikers meer actie- en toekomstgericht moeten zijn. Ook worden evaluaties leerzaam gevonden als ze de impliciete aannames waarop wetgeving of beleid rust, boven water halen en toetsen aan de empirie. Een aanvullende weg is om – in lijn met wat bijvoorbeeld in de VN en bij BZ wordt gestimuleerd – meer werk te maken van de systematische stapeling van kennis uit evaluaties, bijvoorbeeld in metastudies en kennisbanken. Zo verzet de aanvrager van dit onderzoek (WKB) op dit punt al geruime tijd werk.

Slotbeschouwing (Onderzoeksvraag 4)

Veel van de beschreven constructies, procedures en aanpakken zijn nog vers, waardoor niet kan worden nagegaan in hoeverre ze structurele verbeteringen hebben opgeleverd in termen van leren en verantwoorden of zelfs tot effectiever beleid hebben geleid. Uit ervaringen elders is opgemaakt dat evaluatievermogen ook een kwestie van doorzettingsvermogen is. Dit maakt het interessant om de verdere ontwikkelingen te volgen.
Een randvoorwaarde voor 'lerend evaluatievermogen' lijkt dat initiatieven ter vermeerdering van beleidsrelevante kennis niet worden afgestraft, maar ondersteund door de politiek en door de instrumenten in het centrale evaluatiestelsel. Sommige respondenten betreuren de nadruk die met het centrale evaluatiestelsel en de beleidsdoorlichtingen op (eind)doeltreffendheid en verantwoording is komen te liggen. Van plan- en procesevaluatie – en uiteraard vele andere soorten beleidsonderzoek – kan bijvoorbeeld veel geleerd worden, ondanks dat de 'hamvraag' van doeltreffendheid en doelmatigheid niet of niet volledig wordt beantwoord.

Wat betekenen bovenstaande bevindingen nu voor wetgevingsdirecties en de aanvrager, WKB? Directies Wetgeving en/of juridische zaken lijken op grond van dit onderzoek nauwelijks aangehaakt bij ex post beleidsevaluaties, hoewel zij soms wel een wetsevaluatie begeleiden, zeker als er meerdere directies (bijv. Spoorwetgeving) of departementen (bijv. Awb) bij betrokken zijn. Over het algemeen zijn beleidsdirecties leidend, terwijl de wetgevingsjurist bijna los staat van het centrale evaluatiestelsel met de beleidsdoorlichtingen. Aan de voorkant zijn

wetgevingsdirecties nadrukkelijker betrokken, in het kader van ex ante onderzoek en het Integraal Afwegingskader (IAK). Wetgevingsjuristen en andere specialisten kunnen echter een waardevolle rol spelen in *ex post* wetsevaluaties; zuiver sociaalwetenschappelijke aanpakken zijn immers te arm gebleken. Zij zouden vooralsnog eerst meer betrokken kunnen raken bij het evalueren, voordat wordt ingezet op een aparte lijn van 'wetsevaluatievermogen'.
Ook kunnen wetgevingsjuristen een aantal andere dingen betekenen. Zo vallen evaluatiebepalingen in wetten zelf meer in de richting van 'lerend evalueren' te formuleren; standaard vragen om onderzoek naar 'doeltreffendheid en doelmatigheid' is daarvoor niet altijd genoeg. Er kan bijvoorbeeld tevens om inzicht gevraagd worden in de uitvoeringspraktijk en in het hoe en waarom van de werking. Daarbij zijn realistische evaluatietermijnen in de bepaling nodig om recht te doen aan de ontwikkeling van het beleid dan wel het uitkristalliseren van doeltreffendheid en neveneffecten. Ten slotte draagt een adequate toelichting op de wet in de Memorie van Toelichting (MvT) bij aan de werking van de evaluatiebepaling: wat is de bedoeling van de wetgever; wat zijn de aannames, welke resultaten van de wet worden op basis hiervan verwacht? Op welke punten kan een wetsevaluatie hier over x jaar aan bijdragen? WKB heeft inmiddels een schrijfwijzer voor MvT's voltooid, die kan bijdragen aan de onderbouwing en verheldering van wetgeving en beleid (beleidstheorie) en zo aan beter gefocuste evaluaties.
Het conceptrapport heeft de inbreng gevormd voor een expertmeeting met als doel de discussie en gedachtenvorming over de (eventuele) verdere ontwikkeling van evaluatievermogen – in het bijzonder rond wetgeving – aan te jagen.

1. Inleiding en opzet

1.1 Achtergrond en doel van dit onderzoek

Meer dan twintig jaar geleden constateerde de Algemene Rekenkamer op basis van een rijksbreed onderzoek naar de inrichting en uitvoering van de wetgevingsfunctie dat 'structurele voorzieningen voor de evaluatie van wetgeving veelal ontbreken'.[2] Volgens de Rekenkamer betwijfelen de wetgevingsafdelingen het nut van evaluaties. Voor hen zijn de contacten met uitvoerende organisaties en maatschappelijke groeperingen meestal voldoende, wetsevaluatie is in hun visie een taak voor de beleidsafdelingen (vgl. Olsthoorn-Heim, 2003: 14).
In de tussentijd is heel wat water door de Rijn gegaan. Allereerst is de theorievorming op het vlak van wetsevaluatieonderzoek verder voortgeschreden. Winter (1996) deed daartoe een belangrijke aanzet, door met zijn dissertatie inzicht te verschaffen in de mogelijkheden en beperkingen van wetsevaluatie. Analoog aan evaluatie in het algemeen zijn verschillende benaderingen van wetsevaluatie onderscheiden (doelbereiking, doeltreffendheid, effecten). Ook is erop gewezen dat wetsevaluatie in feite draait om het toetsen van veronderstellingen van de wetgever aan de werkelijkheid. Goed evaluatieonderzoek zou moeten voldoen aan theoretische, methodische en onderzoekstechnische eisen. Het ontwikkelen van kwaliteitsnormen werd wenselijk geacht, evenals een goed doordacht evaluatiebeleid en een goede evaluatie-infrastructuur bij de opdrachtgevers (Olsthoorn-Heim, 2003). Over hoe wetsevaluaties methodologisch-inhoudelijk aan te pakken zijn in latere jaren verschillende publicaties verschenen (o.a. Nelen, 2000). Daaronder bevinden zich meta-studies ter inventarisatie en systematisering van gevolgde aanpakken en uitkomsten van wetsevaluaties (Klein Haarhuis en Niemeijer 2008; Veerman, 2013).

Waarom zijn wetsevaluaties belangrijk?
Elk departement is verantwoordelijk voor de kwaliteit van wetgeving op zijn eigen terrein. Het ministerie van VenJ draagt zorg voor het overkoepelende kwaliteits- en evaluatiebeleid. Het Kenniscentrum Wetgeving en Juridische zaken (KCWJ) van dat ministerie ondersteunt de directies Wetgeving en/of Juridische Zaken van de afzonderlijke departementen (KCWJ, 2015).
Wetsevaluaties – ex post – vervullen een belangrijke functie in de waarborging van wetgevingskwaliteit: door terug te blikken op de in- en uitvoering en de werking van wetten, kan geleerd worden voor toekomstige wetten of wetswijzigingen. Als wetgevingsjuristen en beleidsmakers die lessen vervolgens ook inzetten, dan dragen wetsevaluaties bij aan de kwaliteit van wet- en regelgeving en/of aan de kwaliteit van de uitvoering ervan. Op deze manier wordt de regelgevingscyclus, de *regulatory cycle,* gesloten (EC, 2013; OESO, 2015; Smismans, 2015; Mastenbroek et al., 2015). Dit draagt bij aan de professionalisering van het beleidsproces.

Wetsevaluaties zijn dus een essentieel onderdeel van het wetgevingskwaliteitsbeleid. Willen ze daadwerkelijk een bijdrage kunnen leveren aan de kwaliteit van wetgeving, dan is het vanuit het perspectief van de aanvrager van belang om: (1) ze tot stand te brengen; (2) ze zó op te zetten dat ze bruikbare en liefst ook valide en betrouwbare kennis en inzichten genereren; (3) dat beleidsmakers/ wetgevingsjuristen kennis nemen van wetsevaluaties en deze gebruiken in de terugkoppeling naar hun producten.

[2] TK 1993-1994, 23710, nrs. 1-2.

Probleemschets
In de praktijk werkt het echter niet zo systematisch. Uit de twee genoemde metastudies is naar voren gekomen dat wetsevaluaties sterk verschillen in termen van hun focus en opzet. Verder bestaan verschillen in de inrichting van het evaluatieproces, zoals bij de keuze van uitvoerders of het al of niet gebruiken van BC's voor het bewaren van voortgang, focus of kwaliteit. Dit hoeft op zichzelf niet problematisch te zijn – zeker als we de variëteit van te evalueren wetten in ogenschouw nemen.[3] Maar wetsevaluaties bleken eerder ook te verschillen in termen van de *methodologische kwaliteit* (o.a. Klein Haarhuis en Niemeijer, 2008).[4] Sommige bevatten bijvoorbeeld geen duidelijke verantwoording van aanpak en beperkingen, andere wel.
De vraag is waar de verschillen in proces en kwaliteit van wetsevaluaties vandaan komen. Veel individuele aanvragers – zowel van 'beleid' als 'wetgeving' - zeggen niet goed te weten wat te doen als een wetsevaluatie op handen is (interne communicatie Kenniscentrum). Het centrale beleid ten aanzien van wetsevaluaties valt sinds het begin van deze eeuw als volgt te typeren 'Het regelmatig evalueren van regelgeving is nuttig en belangrijk, maar het ontwikkelen van een algemeen evaluatiebeleid heeft geen prioriteit en is niet nodig. De beslissing om wel of geen evaluatiebepaling[5] in een wet op te nemen wordt aan de wetgever overgelaten, de keuze voor het al dan niet systematisch volgen van de werking van de wet aan de departementen. In criteria voor wetsevaluatie wordt van overheidswege niet voorzien' (Nota Wetgevingskwaliteitsbeleid en wetgevingsvisitatie, 2000).
Hoewel de aandacht voor wetsevaluaties is toegenomen zijn deze uitgangspunten ten aanzien van wetsevaluatie zijn anno 2015 niet fundamenteel gewijzigd. Voor de ex ante fase van wetgeving en beleid is het IAK als hulpmiddel ontwikkeld, voor latere evaluatiefasen bestaat zo'n integraal raamwerk niet.[6]
Niet alleen het feit dat er geen overkoepelend evaluatiekader is, maar ook de complexiteit van wetten als evaluatieobject (zie Box 1) kan eraan hebben bijgedragen dat wetsevaluaties in de praktijk heel verschillend worden aangepakt, zowel qua inrichting van het aansturingsproces, als qua inhoud.

Box 1 Wetgeving: een complex evaluatieonderwerp

Als evaluatieobject is wetgeving over het algemeen complexer dan ander beleid (Bussmann, 2010). Veerman (2007) bespreekt de verschillende functies van wetten uitgebreid. Veel wetten verenigen een ordenende of 'infrastructurele' en een sturende of 'instrumentele' functie in zich. De ordeningsfunctie is doorgaans gericht op de langere termijn. Denk aan het regelen van rechtsbescherming en rechtpositie, zoals van burgers in relatie tot de overheid. De instrumentele functie heeft vaker betrekking op de kortere termijn en is in de loop van de tijd steeds meer centraal komen te staan (Winter en Klein Haarhuis, 2007). Mede vanwege de

---

[3] Bijna niets zo rijk geschakeerd en divers als het wetgevingsbestand. Uit de genoemde meta-onderzoeken bleek van een scala van verschillende soorten wetten – alleen al in termen van domein, topic, ambitieniveau van doelstellingen – met daarbinnen heel verschillende onderliggende 'interventies'. Vergelijk bijvoorbeeld de omvangrijke Awb met de wet die specifiek is gericht op particuliere beveiligingsorganisaties.
[4] Kwaliteit is in de metastudies gedefinieerd als een in opzet en uitvoering valide en betrouwbare methodologie. Met valide wordt bedoeld dat de methodologie vrij is van systematische meetfouten, met betrouwbaar dat deze vrij is van toevallige meetfouten. In de eerstgenoemde metastudie (Klein Haarhuis en Niemeijer, 2008) zijn 16 van 74 wetsevaluaties vanwege significante gebreken in één of (meestal) beide opzichten buiten de verdere synthese gelaten; in de tweede studie (Veerman, 2013) zijn 67 van 99 gevonden wetsevaluaties betrokken: louter juridische evaluaties en interne studies zonder methodologische verantwoording zijn niet meegenomen.
[5] Een artikel in de betreffende wet, dat aangeeft dat de wet binnen x jaar al dan niet periodiek wordt geëvalueerd (zie hoofdstuk 3).
[6] Hier verbinden we nadrukkelijk geen kwalificatie aan.

steeds belangrijker wordende instrumentele functie van wetgeving is het verschil tussen wets- en beleidsevaluatie op voorhand niet altijd duidelijk. Een wet bestaat uit recht, maar is een 'voertuig van beleid'; zij maakt deel uit van een breed beleid. Wet en beleid zijn hierdoor in veel gevallen niet goed meer te scheiden (Veerman, 2007: 228).
Het meerledige karakter van wetgeving betekent dat naast onderzoek naar de werking van het in de wet neergelegde beleidsinstrumentarium – een empirische activiteit vergelijkbaar met beleidsevaluatie – óók juridisch georiënteerde onderzoeksactiviteiten worden ondernomen. Denk aan wetstechnische analyse (hoe duidelijk is de wettekst, past deze in het rechtssysteem, hoe is de relatie met andere wet- en regelgeving?) en verder aan jurisprudentie-analyse naar de gelding van de wet en aan internationaal rechtsvergelijkend onderzoek.[7] Een wetsevaluatie is [dus] idealiter een combinatie van juridisch en empirisch onderzoek, 'beleidsevaluatie met een juridische plus' (Winter & Klein Haarhuis, 2007; Van Aeken, 2007; ZonMw, 2013).
In veel empirische wetsevaluaties – vooral door externe onderzoeksbureaus – bleek echter een juridische invalshoek te ontbreken (o.a. Klein Haarhuis & Niemeijer 2008). Elders ontbrak juist weer de empirische component en stonden juridische aandachtspunten, zoals gelding en de consistentie met binnenlandse en EU-wet- en regelgeving centraal.

Evaluatievermogen
De zojuist beschreven gevarieerde praktijk van wetsevaluatie houdt verband met verschillen in wat in de literatuur als '*evaluation capacity'* is aangeduid, en wat wij verder zullen noemen: evaluatievermogen. Daarbij gaat het om het vermogen om (a) wetsevaluaties te verrichten en (b) ze te gebruiken (Bourgeois & Cousins, 2013). Dit vermogen kan verschillen tussen departementen maar ook daarbinnen. Concreet kan daarbij gedacht worden aan: kenmerken van de beleidsorganisatie (bijv. evaluatie-units), het hebben van een evaluatieprogrammering en aan leidraden voor aanvragers ten aanzien van evaluatieproces en –inhoud.
Het doel van dit onderzoek is om een beeld te krijgen van bestaande departementale praktijken rondom de uitvoering en het gebruik van evaluaties en met name, *wets*evaluaties. Hoe worden evaluaties 'gemanaged', welke organisatieonderdelen en procedures zijn ervoor ingericht?
Hier aan ten grondslag ligt een streven bij de aanvrager van voorliggend onderzoek, de directie wetgevingskwaliteitsbeleid (WKB) van het Kenniscentrum Wetgeving en Juridische Zaken (KCWJ), naar meer vergelijkbaarheid tussen wetsevaluaties, niet alleen in termen van de inhoud, maar ook procesmatig. Vergelijkbaarheid in deze termen bevordert dat ervaringen *met* wetsevaluaties en kennis *uit* wetsevaluaties kunnen worden geaccumuleerd en men beter in staat is om lessen te trekken voor de toekomst.
In het verleden is via het zogenoemde Clearing House Wetsevaluaties al getracht om wetsevaluaties meer vergelijkbaar te maken.[8] In vervolg op dit meerjarige

[7] Vgl. Oosting (1987) die in 'De veelzijdigheid van wetsevaluatie' meerdere niveaus onderscheidt: analyse & beoordeling van de tekst van de wet (formele gelding, soorten regels, doelen en instrumenten en wetstechnische aspecten); de nadere verkenning van de materiële bepalingen (normrelevante situaties, normadressaten en de beoogde rechtsgevolgen van de wettelijke bepalingen), het onderzoek naar de werking van de wet (effecten in de praktijk, als zodanig en als gevolg van de uitvoering) en de beoordeling en advisering (op basis van voorgaande, en met betrekking tot de geldings- en werkingsaspecten van de wettelijke regeling)**.** Zie ook het stappenplan van Nelen (2000): de inrichting van de evaluatie, de analyse van de voorgeschiedenis van de wet, de analyse van de wijze van uitvoering, de juridische ijkpunten, het vaststellen en beoordelen van effecten en, ten slotte, het uitdragen van onderzoeksresultaten.
[8] Dit gebeurde door aandacht te besteden aan een aantal vaste, door wetgevingsdirecties gedragen punten (Veerman, 2013):

1. de aard van de norm (open of gesloten, doel of middel, algemene regels of vergunningstelsel);
2. de aard van de handhaving en sanctionering (strafrechtelijk, bestuursrechtelijk, civielrechtelijk);

initiatief heeft de directie WKB investeringen gedaan om inhoud en resultaten van wetsevaluaties beter toegankelijk te maken en bovendien te onderzoeken hoe met behulp van wetsevaluaties de werking van wetgeving verder kan worden bevorderd.[9] Dit is mede relevant in het licht van de nota Toekomstbestendige wetgeving van medio 2015.[10]
Aan de bevindingen van de onderhavige studie, die iets zeggen over het evaluatievermogen, kunnen allereerst bouwstenen worden ontleend voor het op gang brengen van de discussie tussen diegenen die het aangaat, zoals WKB, beleidsmedewerkers, onderzoekscoördinatoren en medewerkers van de departementale directies Wetgeving e/o Juridische Zaken. Vervolgens is het denkbaar dat de bevindingen worden ingezet bij eventuele verdere ontwikkeling van het evaluatievermogen rondom wetgeving.
Omdat beleids- en wetsevaluaties zo verstrengeld zijn, richten we ons in de beschrijving van evaluatievermogen op *beleidsevaluaties in het algemeen* – en beschouwen wetsevaluaties als een bijzondere verschijningsvorm daarvan.
De intentie van dit onderzoek is niet om aan departementen en evaluatieonderzoekers vóór te schrijven, hoe een wetsevaluatie aangepakt behoort te worden. Alleen wetten zelf verschillen onderling dusdanig in termen van o.a. domein, reikwijdte, ambitieniveau en aantal en aard van onderliggende interventies, dat vele wegen naar Rome leiden. Wel zal speciale aandacht uitgaan naar welke handreikingen en richtlijnen er al zijn voor de aanpak van wetsevaluaties (zowel procesmatig als inhoudelijk), welke eisen in de praktijk aan (wets)evaluaties gesteld worden en wat de ervaringen met zulke handreikingen en eisen zijn.
Het lijkt dus raadzaam om te leren van de praktijk, door langs de lijnen van evaluatievermogen ervaringen met de aanvraag en aansturing van (wets)evaluaties te beschrijven. Daarbij komen ook internationale voorzieningen en ervaringen aan de orde.
Tegen de achtergrond van dit onderzoek spelen ontwikkelingen in de internationale arena. Zo ziet de OESO (2015a,b) wetgevingskwaliteitsbeleid als een speerpunt in de bevordering van het economisch leven en welzijn. Daartoe behoort ook het (ex ante en) ex post evaluatiebeleid.[11]

1.2 Onderzoeksvragen

Het doel van dit onderzoek is, zoals gesteld, een beeld krijgen van bestaande departementale praktijken rondom uitvoering en gebruik van evaluaties en met name wetsevaluaties. Beantwoording van de volgende onderzoeksvragen moet dit doel dichterbij brengen:

---

3. de aard van het toezicht (afstand tot het departement, nalevingsgericht of afschrikkingsgericht);
4. de aard van het uitvoeringsorgaan (bestaand of nieuw);
5. de aard van de rechtsbescherming (Awb of bijzonder);
6. of sprake is van delegatie;
7. of overgangsrecht is opgenomen.

Uit de praktijk is niet gebleken dat de zeven punten bij wetsevaluaties systematisch zijn ingezet. Wel heeft een meta-onderzoek door het Clearing House nader inzicht geboden in de elementen die de werking van wetten en de effectiviteit van wetgeving beïnvloeden (Veerman, 2013, 2014).

[9] Zie: https://www.kcwj.nl/gereedschapskist/onderwerppagina/evaluatiebeleid-wetgeving
[10] TK 2014-2015, 33 009, nr. 10.
[11] In 2015 heeft de OESO voor het eerst institutionele voorzieningen ten aanzien van ex post evaluatie en horizonbepalingen voor de OESO-landen in kaart gebracht (OESO, 2015a,b). Hoofdstuk 2 bevat een korte bespreking.

1. Welke inzichten bieden de internationale (a) literatuur en (b) normen en handreikingen met betrekking tot *evaluatievermogen*? (hoofdstuk 1 en 2)
2. Welke praktijkervaringen met evaluatievermogen zijn opgedaan door beleidsmedewerkers en wetgevingsjuristen, betrokken bij (wets)evaluaties? (hoofdstuk 3)
3. Welke inhoudelijke en methodologische eigenschappen van ex post-evaluaties en -wetsevaluaties acht men daarbij van belang? (hoofdstuk 3)
4. Welke lessen bieden de antwoorden op voorgaande onderzoeksvragen voor verdere gedachtevorming over de ontwikkeling van evaluatievermogen – vooral met betrekking tot wetgeving? (hoofdstuk 4)

Ex ante evaluaties worden in dit onderzoek buiten beschouwing gelaten. We richten ons op ex post evaluaties: onderzoek naar doelbereiking, doeltreffendheid en neveneffecten. Procesevaluaties staan niet centraal, maar kunnen voor het begrip van de doeltreffendheid wel belangrijke inzichten bieden.[12] Om die reden sluiten we procesevaluaties niet uit.
Met onderzoeksvraag 3 doelen we vooral op wat beleidsdepartementen adviseren of voorschrijven rondom het design, het methodologische ontwerp van een evaluatieonderzoek. Deze onderzoeksvraag naar evaluatie-aanpakken komt voort uit behoefte bij de aanvrager (KCWJ-WKB) aan nadere kennis en informatie op dit punt. De afbakening en designkeuzes in een evaluatie bepalen in belangrijke mate de inhoud en ook de kwaliteit. Bijvoorbeeld, of op basis van de onderzoeksgegevens en de analyse een gefundeerd oordeel kan worden gegeven over de in- en uitvoering van een wet, over de mate van doeltreffendheid of doelbereiking en over neveneffecten, zoals in termen van milieu en administratieve lasten. Die kwaliteit is afhankelijk van het toepassen van geschikte methoden, het op tijd verzamelen van data en nauwkeurigheid in de interpretatie van data (o.a. Cooksy, 2012).
Er bestaat niet één ideaal evaluatiedesign, dit is situatieafhankelijk.
In dit rapport stellen we zelf geen inhoudelijke of methodologische normen of criteria op, maar geven we een beschrijving van bestaande leidraden hierover en zo mogelijk ook van wat hun toepassing heeft opgeleverd. Daarbij houden we oog verschillen tussen domeinen, typen wetgeving en evaluatievragen. Na het voorlopig beantwoorden van onderzoeksvraag 4 in het conclusiehoofdstuk van dit rapport, zal een expertmeeting worden georganiseerd ten behoeve van verdere discussie en vooruit kijken.[13]
De vraag, of de aanwezigheid en inzet van evaluatievermogen ook echt doorwerkt in de inhoud en kwaliteit van (wets)evaluatieonderzoek, blijft in deze studie buiten beschouwing. Dit is overigens wel een veel gemaakte aanname in de literatuur – eentje die goeddeels ongetoetst is gebleven (zie hoofdstuk 2).

1.3 Conceptueel kader : evaluatievermogen in context

Wat is evaluatievermogen?
Sleutelwoord in de grote hoeveelheid internationale wetenschappelijke literatuur over dit onderwerp is *evaluation capacity (building)* of *ECB*. Dit is een veelomvattend begrip, dat door de tijd heen verschillend – doch steeds breder – is gedefinieerd. Omdat 'capaciteit' in ons taalgebied al gauw vernauwd wordt tot middelen en mankracht, gebruiken we vanaf hier het begrip 'evaluatievermogen', omdat dit meer omvat dan mensen en middelen alleen.

[12] Zo laat de mate van programma-integriteit zien in hoeverre een wet of beleid is geïmplementeerd (Lösel, 1995). Blijkt de implementatie niet of maar deels volbracht, dan kan van effectiviteit geen sprake zijn.
[13] De resultaten van die expertmeeting zijn niet meer in dit rapport verwerkt.

Voor de eerste grove definiëring van evaluatievermogen sluiten we aan bij het tweeledige construct van evaluatievermogen van Bourgeois & Cousins (2013): het vermogen om (a) evaluaties te verrichten en (b) evaluatiebevindingen te gebruiken. Het door deze onderzoekers ontwikkelde en gevalideerde construct geeft volgens hen een compleet beeld van de mate waarin een organisatie de voorwaarden in huis heeft om evaluatieonderzoek uit te voeren en de resultaten hiervan te verspreiden en te gebruiken.

Politieke context

Evalueren vindt niet in een politiek vacuüm plaats.[14] De politieke context bepaalt in belangrijke mate het vermogen om evaluaties te verrichten en te gebruiken (Cooksy, 2012; Bovens et al., 2008; Van Aeken, 2007). Enigszins gechargeerd is beleidsevaluatie wel geduid als de voortzetting van politiek met andere middelen (Bovens et al., 2008, p. 319). Als evaluaties vanwege politieke druk binnen een korte tijdspanne moeten worden afgerond of binnen een krap budget, lijdt de kwaliteit daar mogelijk onder. Er is dan weinig ruimte om voldoende gegevens te verzamelen om conclusies te onderbouwen (ibid). In verschillende fasen van het evaluatieproces kan volgens genoemde auteurs manipulatie plaatsvinden – de mate waarin dit gebeurt, verschilt met bijv. de politieke gevoeligheid van een evaluatieonderwerp. Dit kan ten eerste gebeuren in de fase van het initiëren van de evaluatie: tijdstip, wie doet het, wie zit er in de BC, welke bronnen en middelen worden beschikbaar gesteld? Ten tweede kan politieke 'sturing' aan de orde zijn in de fase van het openbaren en gebruiken van de evaluatiebevindingen: in de la stoppen of in beslotenheid bespreken, bevindingen niet terugkoppelen naar besluitvorming (Bovens et al., 2008, p. 320).[15]

De politieke context kan echter ook een andersoortige invloed hebben op hoe wordt omgegaan met evaluaties; zo zijn parlementaire instituties rond evaluaties (Jacob et al., 2015) een criterium voor de mate waarin evaluatie in het democratische systeem is ingebed. Waakzaamheid van het parlement op dit vlak kan de aandacht van bewindslieden en beleidsmedewerkers voor evaluaties verscherpen.[16]

Het bouwen aan evaluatievermogen is op te vatten als tegenwicht bieden tegen eroderende invloeden van buitenaf. Door bijvoorbeeld te werken met een (meerjarige) evaluatieprogrammering en standaarden voor inhoud, kwaliteit en gebruik, wordt gepoogd een schil om het evaluatieproces aan te leggen en weerstand te bieden aan – onder meer – zaken als tijdsdruk en de politieke 'waan van de dag'.[17] Maar tegelijkertijd – vooral op langere termijn – beïnvloedt het politieke klimaat de ruimte die wordt geboden aan het (verder) ontwikkelen van evaluatievermogen.

Evaluatiegemeenschappen

Behalve de politiek oefenen ook andere factoren van buitenaf invloed uit op evaluatievermogen. Zo ondersteunen volgens Cooksy (2012) landelijke evaluatiegemeenschappen (in Nederland is dit VIDE met daaronder het Evaluatorennetwerk) het evalueren via bijvoorbeeld trainingen, cursussen, conferenties, presentaties en lezingen. Dit bevordert de kennisuitwisseling en de reflectie op manieren om evaluaties aan te pakken.[18]

---

[14] In de internationale literatuur over evaluatievermogen komt de rol van de politiek echter enigszins bedekt aan bod – deels omdat het hier doorgaans over organisaties in hun algemeenheid gaat.

[15] *'The list of tactics used by parties to influence the course and outcomes of evaluation efforts is long (...)'.*

[16] Even los van of het oppositie- of regeringspartijen betreft (zie daarvoor Van Aeken, 2007).

[17] Zie ook Van Aeken (2007) die de term evaluatie-infrastructuur gebruikt.

[18] In o.a. de VS en Canada stellen de evaluatiegemeenschappen óók algemene richtlijnen, standaarden en checklists voor evaluatieonderzoek op. In NW-Europa is dat vooralsnog in

Evaluatievermogen: kunnen uitvoeren en kunnen gebruiken

In wat volgt werken we met behulp van de literatuur over evaluatievermogen verder uit, wat evaluatievermogen inhoudt. Voor zover wij hebben kunnen nagaan, is de internationale academische belangstelling voor evaluatievermogen *niet of nauwelijks toegespitst op de specifieke kenmerken van wet- en regelgeving* en doen we het dus met bredere inzichten.

Het vermogen om evaluaties *te verrichten* gaat volgens Bourgeois en Cousins (2013) over het hele evaluatietraject vanaf het aanvragen, opstellen van startnotitie, uitvoeren en begeleiden van evaluatieonderzoek. Dit vereist organisatorische en individuele capaciteit, een evaluatieprogrammering en -activiteiten. Deze aspecten werken we verderop uit, onderscheid makend tussen de vraag- en aanbodzijde van evaluaties.

De tweede dimensie van evaluatievermogen die we van Bourgeois & Cousins overnemen is *het vermogen om evaluatiebevindingen te gebruiken*. Deze dimensie geeft de mate aan waarin men in staat is kennis en inzichten uit (wets)evaluaties terug te koppelen naar het proces van beleid en wetgeving. Gebruik is volgens de auteurs een belangrijke dimensie, want als een departement een meerjarige evaluatieprogrammering hanteert zonder vervolgens de bevindingen in te zetten om beleid en wetgeving te verbeteren, is evalueren weinig zinvol.

Het vermogen om evaluaties te gebruiken wordt volgens dezelfde auteurs bepaald door: de *betrokkenheid van de aanvrager en andere stakeholders* en de mate van *integratie van evaluatie* in het proces van besluitvorming (vgl. Smismans, 2015).[19] Bourgeois & Cousins (2013) noemen in het kader van gebruik ook de leeropbrengsten van evalueren; verwant aan wat elders wel conceptueel of indirect gebruik is genoemd (o.a. Winter, 1996). Daarmee worden bedoeld de attitude- en gedragsveranderingen voortvloeiend uit evaluaties, de mate waarin leerervaringen formeel en informeel worden gedeeld en het 'denken in evaluatietermen' door organisatieleden. Dit laatste is bijvoorbeeld het geval als beleidsmedewerkers al bij de ontwikkeling van beleidsdoelen en instrumenten rekening houden met evalueerbaarheid.

Evaluatievermogen: vraag- en aanbodzijde

Het model van Nielsen et al. (2011) overlapt in veel opzichten met dat van Bourgeois & Cousins (2013). Een verschil is echter dat Nielsen et al. nadrukkelijk onderscheid maken tussen de vraag- en de aanbodzijde van evaluatievermogen.[20] De vraagzijde heeft betrekking op de aanvragers van evaluaties, de aanbodzijde op de uitvoerders. Omdat juist in Nederland veel beleidsonderzoek extern wordt aanbesteed, nemen we dit onderscheid over in de verdere uitwerking van evaluatievermogen.

Belangrijke aspecten van evaluatievermogen aan de *vraagzijde* (hier: de aanvragers, beleidsmedewerkers) zijn het hebben van *duidelijke evaluatiedoelstellingen en structuren en processen voor evaluatie.*

---

mindere mate het geval (zie ook Hoofdstuk 2) en hebben evaluatiegemeenschappen een minder prominente rol. Wel is er de European Evaluation Society (EES).

[19] Ook Johnson et al. (2009) lieten op basis van een internationale metastudie (met slagen om de arm) zien dat het betrekken van belangrijke stakeholders in het evaluatietraject het meest cruciaal is voor gebruik; waarschijnlijk meer nog dan de inhoudelijke kwaliteit. Verder bleken van belang – niet verrassend – de politieke of ambtelijke besluitvormingscontext en de timing en tijdigheid van oplevering van evaluaties.

[20] Verder benoemen Nielsen et al. de evaluatieprogrammering niet expliciet als onderdeel van evaluatievermogen.

Vraagzijde: *evaluatiedoelen*
Evaluaties worden verricht vanuit verschillende, uit de politieke context voortkomende, prikkels en motieven – die elkaar overigens niet altijd uitsluiten.[21]

- Voldoen aan verantwoordingseisen (beleidsdoorlichtingen)
- Wettelijke evaluatieverplichtingen
- 'Ad hoc' Kamervragen
- Uitstel van (ingrijpende) beslissingen
- Politiek-strategisch (vaak al bij start beleid)
- Leren van het tot dusverre verzette beleidswerk (bijv. voor inhoud maar ook conceptueel, of voor de uitvoering).

Evaluatiedoelen kunnen worden geformaliseerd in richtlijnen of handreikingen voor evaluatieonderzoek of in een meerjarige evaluatieplanning en –programmering. Nielsen et al. (2011) achten heldere evaluatiedoelen daarom belangrijk.
In eerdere studies over evaluatieonderzoek is wel onderscheid gemaakt tussen leren versus verantwoorden als evaluatiedoelstelling of -motief (Vedung, 1997). Dit onderscheid bepaalt wat de aanvrager motiveert en hoe een evaluatie eruitziet.
Bij leren gaat het erom, inzicht te krijgen in hoe een wet of een beleidsinterventie in de praktijk uitwerkt in relatie tot het oplossen van problemen (*enlightenment*). Afhankelijk van de uitkomsten van de evaluatie kan die kennisvergroting leiden tot aanpassingen aan beleid of wetgeving (*embetterment*) (o.a. Rossi & Freeman, 1993; Leeuw et al., 1994; Mackay, 2006). Het leermotief komt vaak tot uiting in formatieve evaluatie gericht op eventuele bijsturing – zijn we op de goede weg? Hoe kunnen we wet of beleid verbeteren? Formatieve evaluaties kunnen allerlei gedaanten aannemen; bruikbaarheid is het belangrijkste beoordelingscriterium (Scriven, 1994; Patton, 2008). Bij leren betreft de focus vaak de langere termijn en daarbij past een dynamische opvatting van beleid: loopt het (implementatie)proces zoals beoogd en zo niet, wat moet worden aangepast? Onder welke omstandigheden of condities pakt het beleid 'beter' of minder goed uit?
Niet alleen leren is een drijfveer van beleidsevaluatie, maar ook het afleggen van verantwoording over overheidsuitgaven.[22] In summatieve evaluaties wordt een 'eindoordeel' over de doelbereiking of doeltreffendheid van beleid gegeven, aansluitend op dit motief van verantwoorden (o.a. Vedung, 1997). Een mogelijk probleem met het verantwoordingsdoel is dat het departement dat verantwoordelijk is voor de ontwikkeling van de wet of het beleid, ook de evaluatie uitzet of verricht. Maar evalueren vanuit een verantwoordingsmotief sluit het stellen van kwaliteitseisen zeker niet uit. Zo geldt wel als beoordelingscriterium dat de conclusies zoveel mogelijk op een consistente, transparante analyse van feiten gebaseerd zijn (Vedung, 1997). Vaak staat de doeltreffendheid, het kunnen toeschrijven van resultaten aan beleidsinspanningen, centraal (attributie).[23] In dit perspectief worden andere evaluatievormen, zoals procesevaluaties of evaluaties van onderdelen (interventies) van de wet of beleid onvoldoende geacht om tot doeltreffendheidsuitspraken te komen.

[21] vgl. Klein Haarhuis en Niemeijer (2008) in verwijzing naar eerder werk.
[22] Het centrale Nederlandse evaluatie- en verantwoordingsstelsel bespreken we in hoofdstuk 3.
[23] Een in de literatuur veel voorkomend synoniem van deze term is 'effectiviteit' (*effectiveness*).

Tabel 1 Evaluatiedoelen, benaderingen en waarden

Evaluatiedoel	Benaderingen	Waarden
'Leren' ***(enlightenment, embetterment)***	• Meerdere mogelijke vormen van evaluatie • Toekomstgericht • Interne en externe evaluatie • Formatieve evaluatie • Lange termijn- perspectief • Dynamisch / contextgebonden	• Nut • Bruikbaarheid voor beleid
Verantwoorden	• Doelbereikings- of doeltreffendheidsevaluatie • Outputgericht • Externe evaluatie • Summatieve evaluatie • Korte-termijn perspectief • Statisch	• Doeltreffendheid • Objectiviteit • Transparantie

Ordening op basis van de aangehaalde literatuur.

Voor het 'sluiten van de cyclus' van wetgeving en beleid zou een lerend evaluatieperspectief de voorkeur genieten boven een verantwoordingsperspectief. Evaluaties die 'opener' zijn ingestoken (hoe pakt het uit? welke lessen kunnen we trekken?) bieden immers meer kans om nieuwe inzichten te genereren dan verantwoordingsexercities (is het beleid succesvol?). In de praktijk zal de scheidslijn tussen de evaluatievormen van Tabel 1 echter niet heel scherp zijn. Zo maakt verklarende evaluatie steeds vaker deel uit van handreikingen voor summatief onderzoek. Omgekeerd worden tussentijdse, op leren gerichte evaluaties soms ingezet voor verantwoordingsdoeleinden – al dan niet gerechtvaardigd.
Op langere termijn werkt het motief mogelijk door in de manier waarop structureel met evaluaties wordt omgegaan: in het evaluatievermogen.

Vraagzijde*: structuren en processen*
Evaluatiedoelen, zoals bijvoorbeeld politieke of uitgavenverantwoording of juist 'leren', kunnen doorwerken in de organisatie*structuren en -processen.* Bij structuren en processen denken Nielsen et al. (2011) aan de aanwezigheid van een evaluatieafdeling of –unit. Daarbij valt te denken aan een aparte stafdirectie of een unit bij een beleidsdirectie, financiële of kennisdirectie.
Bourgeois & Cousins (2013) benadrukken het belang van een (meerjarige) evaluatieprogrammering voor de oplevering van verschillende evaluaties van het ingezette beleidsinstrumentarium. De meerjarigheid kan de programmering 'weerbaarder' maken tegen incidenten en politieke invloeden. Tegelijkertijd wordt aanpassingsvermogen van de programmering aan nieuwe ontwikkelingen van belang geacht.
Voorts bepalen de tijd en de financiële middelen die de beleidsorganisatie vrijmaakt voor evaluaties het evaluatievermogen bij de aanvrager.
Volgens Cooksy (2012, p. 82) weerspiegelen deze structuren en processen de waarde die aan evaluatie wordt gehecht en hoe bereidwillig een organisatie is om goed evaluatieonderzoek te verrichten. Zij voegt kwaliteitscontrole op de onderbouwing van evaluaties toe aan het lijstje. Daarbij kan bijvoorbeeld gebruik worden gemaakt van 'checklists' en procedures.

Aanbodzijde*: menselijk kapitaal (en technologie)*
Aan de aanbodzijde, bij onderzoekers die evaluaties uitvoeren, zijn menselijk kapitaal (Nielsen et al., 2011; Bourgeois & Cousins, 2013) en technologie (Nielsen

et al., 2011) bepalend voor het evaluatievermogen. In de definitie van Nielsen et al. vertegenwoordigt het concept menselijk kapitaal de kennis, vaardigheden (opleiding, training) en ervaring van diegenen die een evaluatie uitvoeren. Het zijn noodzakelijke voorwaarden, zonder die kennis en ervaring geen evaluatievermogen. Tot vaardigheden rekent Cooksy behalve technische en methodologische vaardigheden óók interpersoonlijke vaardigheden, zoals communicatieve en projectmanagement-vaardigheden en reflectievermogen. Evaluatieonderzoekers moeten naast kwantitatieve en kwalitatieve methodologische kennis volgens Cooksy ook weten hoe de bevindingen te rapporteren en aan de man te brengen. Dat maakt praktijkervaring onontbeerlijk.[24] Mark (2012, p. 85) acht het bovendien een onmisbare competentie van de evaluatieonderzoeker om flexibel te zijn in relatie tot veranderende omstandigheden, en te kunnen doorgronden waar hij/zij zich bevindt in het (ambtelijke of politieke) spel.[25]
Technologie verwijst volgens Nielsen et al. naar de middelen en bronnen bij de uitvoerders van evaluaties: ICT-infrastructuur, betrouwbare en beproefde evaluatiemodellen, methoden voor dataverzameling, en systemen voor de monitoring van prestaties en uitkomsten. Deze factor wordt door de auteurs minder zwaarwegend geacht dan menselijk kapitaal.
De vraag- en de aanbodzijde zijn afhankelijk van elkaar (Cooksy, 2012; Nielsen et al., 2011; Preskill & Boyle, 2008; Taylor-Ritzler et al., 2013, p.200). Waar bijvoorbeeld de (politieke) prioriteit en motivatie voor evaluatie ontbreken lijdt de professionaliteit van de uitvoering daaronder. Er is dan bijvoorbeeld onvoldoende budget om de meest capabele onderzoekers in te huren, of de onderzoekers zetten zich minder in zodra ze beseffen dat hun werk vooral zal worden gebruikt ter legitimering van beleid of wetgeving.

Overzicht conceptueel kader evaluatievermogen
Op basis van de voorgaande inzichten is een conceptueel kader opgesteld. Tabel 2 geeft een overzicht van de belangrijkste aspecten van het vermogen om evaluaties te doen en te gebruiken, en hun respectievelijke kenmerken (in de kolommen ernaast). De keuze van die aspecten en kenmerken is gebaseerd op twee criteria: de mate waarin ze voorkomen in de literatuur en de relevantie van variabelen voor de institutionele situatie en wat we al weten over evalueren in Nederland.[26] Twee punten waarop de situatie bij Nederlandse beleidsdepartementen afwijkt van de algemene focus in de literatuur, is dat (1) evaluatievermogen in de literatuur vaak (lerende) organisaties betreft en niet zozeer de beleids- of wetgevingscyclus; (2) in Nederland het gros van het (wets)evaluatieonderzoek wordt uitbesteed aan onderzoeksbureaus, wetenschappelijke instituten en universiteiten (o.a. Veerman 2013), terwijl in de literatuur vaak organisatie-intern evaluatieonderzoek centraal staat (Bourgeois, 2011).

Tabel 2 Aspecten van Evaluatievermogen en hun kenmerken; topics voor de interviews (ontleend aan de internationale literatuur, aangepast aan Nederlandse situatie)

[24] Een andere vaardigheid is volgens Cooksy het kunnen destilleren van informatie uit de gigantische en uitdijende hoeveelheid aanwijzingen en leidraden op het Internet – om zicht te krijgen op wat bruikbaar is.
[25] '*By seeing that in a particular case, for instance, evaluation is a ritualistic exercise to the funder and program leaders, the evaluator can exercise adaptability and reformulate the emphasis of the evaluation in ways that might bring positive consequences nevertheless'.*
[26] Voor een meer gedetailleerde uitwerking en operationalisering op basis van de literatuur over *evaluation capacity,* zie Parapuf (2016), die binnen het kader van dit onderzoek een Masterscriptie voltooide.

Context				
	Aspecten	**Kenmerken**		
	• Politieke commitment	Motivatie voor al dan niet verrichten van wetsevaluaties (verantwoording, leren, anders)		
	• Andere contextfactoren	- Nationale / internationale evaluatiegemeenschappen - Overige (bijv. conjunctuur ...)		
	Aspecten	**Kenmerken**	**Bronnen**	
Organisatie-inrichting/ **vraagzijde**	• Structuur	- Formele doel(en) van evaluatieonderzoek - Aanwezigheid *evaluatie unit,* aantal medewerkers - Verhouding intern-extern uitgevoerde evaluaties	(Online) documenten Organigram-men Jaarverslagen Interviews	**Vermogen om evaluaties te doen en vermogen om evaluaties te gebruiken**
	• Evaluatie-programmering	- (Meerjarige) evaluatieprogrammering		
	• Middelen	- Omvang en toedeling financiële en andere middelen (bijv. tijd)		
	• Evaluatiekader (ten aanzien van zowel evaluatieproces als -design)	- Aanwezigheid en inhoud van evaluatiebevorderende programma's - Richtlijnen, handreikingen, procedures bij aanvraag/ uitvoering		
	• Borging van evaluatiekennis en –ervaring	Kennis (aanvrager) - Kennis van en ervaring met evaluatieproces en – inhoud - Communicatie van evaluatiekader in de organisatie & trainingen/cursussen voor beleidsmedewerkers Selectie - Competenties waarop aanvragers onderzoekers selecteren - Begeleidingscommissies en hun samenstelling Ervaring - Delen en verankeren van ervaringen met evaluaties		
	• Gebruik	Proces (bruikbaarheid) - Communicatie tussen onderzoekers en aanvragers – al dan niet via een evaluatie- of BC - (Wijze van) presentatie bevindingen; aanbevelingen - Rapportage (openbaarheid, toegankelijkheid) en verdere verspreiding Gebruik - Direct gebruik in beleids- en besluitvorming - Overige leeropbrengsten		

Individuele competenties/ **aanbodzijde**	Kennis & vaardigheden	➔ Borging (selectie)		

Gelet op het grote aandeel extern aanbestede evaluatiewerk in ons land, omvat menselijk kapitaal in onze definitie zowel de competenties – kennis, vaardigheden en ervaring – van beleids- en juridische medewerkers om evaluatieaanvragen *op te stellen,* als de competenties van onderzoekers om deze *uit te voeren*. De politieke context en het evaluatievermogen bij de aanvragers, inclusief hun competenties, werkt door in de selectie- en begeleidingsprocedures van evaluatieonderzoekers. Hiermee brengen we een verruiming aan van Nielsen's model, waarin competenties strikt genomen alleen het *aanbod* van evaluatieonderzoekers betreffen.
De externe onderzoekspartijen vallen buiten het directe bereik van dit onderzoek; de hoofdfocus ligt op de capaciteit bij de aanvragers van evaluaties bij de centrale beleidsdepartementen. Wel wordt de aanbodzijde belicht in die zin dat aanvragers uitvoerders *selecteren*, op grond van (individuele of team-)competenties. Beschreven zal worden, op basis van welke eigenschappen dat in de praktijk zoal gebeurt.
Aan bovenstaande aspecten van evaluatievermogen zou men een norm of meetlat kunnen ontlenen – als ware het ideale of wenselijke situaties. Dit onderzoek is echter niet bedoeld om te toetsen in hoeverre verschillende departementen voldoen aan het conceptuele kader; deze studie vervult dus zelf geen 'evaluatieve' functie. Zo betekent de aanwezigheid van vernieuwende evaluatieconstructies bijvoorbeeld nog niet dat een heel departement er 'betere' praktijken op nahoudt; dit hebben we niet steeds verder uitgezocht.

1.4 Aanpak

Aan elk van de verschillende aspecten van evaluatievermogen (en contextuele factoren) zijn, met behulp van genoemde literatuur en toegespitst op de Nederlandse situatie, een aantal meetbare kenmerken gekoppeld (Tabel 2).[27]
Wat de gegevensverzameling betreft, is gekozen voor een tweeledige aanpak:
(a) *Desk study en literatuuronderzoek* van (internationaal) bekende inzichten en ervaringen. Deze hebben gefungeerd als inspiratiebron voor
(b) *Semigestructureerde interviews* bij een negental Nederlandse beleidsdepartementen, met behulp van topiclijsten.

Ad (a) De *desk study* van kennis en ervaringen omvatte de volgende materialen:

1. Literatuur over evaluation capacity en aanverwant (dit hoofdstuk)[28];
2. Beschrijvingen van verschillende aspecten van evaluatievermogen bij enkele grotere internationale instituties en landen, inclusief de hier ontwikkelde leidraden voor de aanpak van (wets)evaluaties (hoofdstuk 2);
3. Desk study van dezelfde aspecten van evaluatievermogen bij Nederlandse beleidsdepartementen op basis van gepubliceerde (online) documenten. Naderhand zijn via respondenten aanvullende interne documenten verkregen (hoofdstuk 3).

Aan deze drie vormen van vooronderzoek is een belangrijk deel van de interviewtopics ontleend. De algemene topiclijst is tevoren zoveel mogelijk

[27] Zie Parapuf (2016) voor nadere uitleg en verantwoording van de operationalisering van variabelen naar indicatoren per variabele.
[28] Als zoektermen zijn gebruikt o.a. *evaluation policy, evaluation use, evaluation systems, evaluation culture* - en vervolgens sneeuwbalmethode.

toegespitst op de Nederlandse context.[29] Elk afzonderlijk interview is voorbereid aan de hand van online beschikbaar materiaal over het departementale onderdeel in kwestie.
Bijlage 2 bevat een overzicht van de 35 in 20 gesprekken geraadpleegde sleutelinformanten met het oog op onderzoeksvragen 2 en 3. Voor de selectie van de sleutelinformanten is aangesloten bij waar binnen de rijksoverheid ervaring met evaluatievermogen aanwezig is. Eerste ingangen waren WKB, het Interdepartementaal Overleg Directies Wetgeving en/of Juridische Zaken (IOWJZ)[30] en ervaringsdeskundigen bij beleids- en tevens kennisdirecties bij verschillende ministeries. Ook is gebruik gemaakt van het bestaande professionele netwerk van onderzoeker en aanvrager(s). In de periode juni-oktober 2015 is gesproken met (senior)-beleidsmedewerkers en onderzoekscoördinatoren, directies Financieel-economische Zaken (FEZ) en enkele directies Wetgeving e/o Juridische Zaken, en indien van toepassing, rijksonderzoeksinstituten of -organen. Het betreft uiteindelijk negen departementen: BZK, BZ[31], EZ, Financiën[32], IenM, OCenW, SZW, VenJ en VWS. Per departement is niet met alle typen respondenten gesproken; de samenstelling wisselt afhankelijk van waar relevante initiatieven zijn ontplooid. Andere respondenten zijn vertegenwoordigers van: VIDE[33], WRR, de Algemene Rekenkamer, de Raad van State en betrokkenen bij twee wetgevings-evaluatieprogramma's (Awb en ZonMw).
Om de respondenten te stimuleren hun ervaringen met evaluatievermogen feitelijk te onderbouwen, is steeds doorgevraagd naar geschreven bronnen, zoals taakomschrijvingen van organisatieonderdelen verantwoordelijk voor evalueren, evaluatieprogramma's, -leidraden of handreikingen. Acht (groepen) respondenten met een aanzienlijk aandeel in de eindtekst zijn enkele maanden na het interview in de gelegenheid gesteld om de betreffende tekstdelen op feitelijke onjuistheden te corrigeren en aan te vullen.[34]

1.5 Opbouw van dit rapport

Het conceptuele kader vormt de ruggengraat voor de beschrijvingen in elk van de drie nu volgende hoofdstukken. In het tweede hoofdstuk wordt onderzoeksvraag 1 verder beantwoord, op basis van aanpakken en ervaringen van internationale organisaties en enkele grotere landen. Het derde hoofdstuk vangt aan met een algemene beschrijving van het Nederlandse evaluatiestelsel en het wetsevaluatiebeleid. Daarna zijn de Nederlandse beleidsdepartementen aan de beurt en we volgen ook hier de aspecten van het conceptuele kader (onderzoeksvragen 2 en 3). Binnen het aspect 'evaluatiekader' gaat speciale aandacht uit naar ervaringen met het design, het methodologische ontwerp, van evaluaties. Voor dit derde hoofdstuk is behalve van desk study ook gebruik gemaakt van interviews met sleutelinformanten bij de negen beleidsdepartementen en daarbuiten.

---

[29] Een eerste versie van de topiclijst is getest in een interview met vier EWB-medewerkers (WODC-intern).
[30] Tijdens een vergadering van dit platform is al in een vroeg stadium van het onderzoek aangegeven dat men primair navraag zou moeten doen bij beleidsdirecties die belast zijn met evaluatie en niet zozeer bij wetgevingsdirecties zelf.
[31] Bij BZ worden geen wetsevaluaties verricht. Toch achten we de ervaringen van de IOB (Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie), het instituut verantwoordelijk voor de meeste evaluaties bij BZ, relevant met het oog op wetsevaluaties, met name rond het gebruik en de leeropbrengsten van evaluaties (zie par. 3.7).
[32] Bij Financiën is alleen het centrale evaluatiestelsel aan bod gekomen.
[33] Beroepsvereniging voor toezichthouders, inspecteurs, handhavers en evaluatoren.
[34] Het betrof de respondenten die zijn geïnterviewd rondom: EZ, OCW (Kennis), SZW (FEZ), IenM (FEZ/HBJZ), VenJ (WODC-EWB), Awb, ZonMw en BZK (FEZ).

Hoofdstuk 4 biedt een samenvattende, geïntegreerde terugblik op de voorafgaande drie hoofdstukken, gevolgd door een nabeschouwing en slotconclusies.

Hoofdstuk 2. Internationale ervaringen met evaluatievermogen

Er is internationaal veel ervaring opgedaan met het bouwen aan evaluatievermogen (*evaluation capacity building* of ECB). De inzichten specifiek gericht op het doen en gebruiken van *wetsevaluaties* zijn echter een stuk dunner gezaaid, ondanks dat veel evaluaties in de praktijk een wet- en regelgevingscomponent hebben. In dit hoofdstuk zetten we een aantal internationale ervaringen met het bouwen aan evaluatievermogen (al dan niet in relatie tot wetgeving) op een rij.
In de internationale beleidspraktijk zijn, zeker het afgelopen decennium, in het kader van evaluatievermogen overweldigend veel handreikingen en gidsen, checklists, leidraden enzovoorts verschenen, zo leert de online zoektocht. Het is onbegonnen werk om dat alles in dit hoofdstuk systematisch aan bod te laten komen. We beperken ons daarom tot het belangrijkste, relatief algemene materiaal en de ervaringen van grotere internationale instellingen en landen in met name de EU en de VS. Dit materiaal hebben we gestructureerd naar het conceptuele schema van hoofdstuk 1: structuur, programmering, middelen, evaluatiekader, borging van kennis/ervaring en gebruik.[35]

2.1 Structuur

Bij sommige internationale organisaties ligt de oorsprong van de aandacht voor evaluatievermogen in een sterke behoefte aan verantwoording (*transparency & accountability*; Wereldbank, 1992). Denk aan aanzienlijke uitgaven aan leningen en ontwikkelingsprogramma's. Mede daarom bestaat bij de Wereldbank een van de langste ECB-tradities (vgl. de Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (IOB) in Nederland). Maar juist hier is in de loop der jaren ook een sterke beleidsinhoudelijk gedreven functie ('leren', als in enlightenment, *embetterment*) naar de voorgrond gekomen, getuige bijvoorbeeld de aandacht die het *Wereldbank Institute* en de huidige *Independent Evaluation Group* (IEG) aan leren besteden (Box 1).

Box 1 Wereldbank: Independent Evaluation Group

Van de Wereldbank en de betrokken ontwikkelingslanden wordt een hoge mate van *accountability* vereist; het gaat immers om de besteding van donorgelden. Als sinds de jaren tachtig werkt de Wereldbank dan ook aan monitoring & evaluatiesystemen (m&e) in ontwikkelingslanden (o.a. Wereldbank 2002, 2004a,b). De Bank is ervaringsdeskundige op het vlak van ECB. Het doel van evaluaties bij de Wereldbank is zowel verantwoorden als leren: om een objectieve beoordeling te geven van resultaten van de Wereldbank en om lessen te trekken uit ervaringen en die te verspreiden.
De Wereldbank heeft (o.a.) een Independent Evaluation Group (IEG), die reguliere evaluaties verricht van wereldwijde of regionale programma's en beleidsstrategieën.[36] De IEG rapporteert rechtstreeks aan de Board of Directors van de WB Group. De IEG verricht behalve projectevaluaties en doeltreffendheidsstudies (*impact evaluation*) en onafhankelijke assessments, ook veel overige werkzaamheden, waaronder literatuurreviews, analytisch werk, landenstudies en surveys onder staf en stakeholders. Daarin vinden verschillende soorten evaluatie-analyses plaats: het toetsen van uitkomsten van verschillende typen beleid aan gestelde doelen, verwachtingen en benchmarks en voorts het onderzoeken wat gebeurd zou

[35] Door de aard van het materiaal – vaak integrale benaderingen of programma's – valt de inhoud niet altijd precies in deze zes topics in te passen.
[36] www.ieg.worldbankgroup.org .

zijn in de afwezigheid van dat beleid. Dit laatste wordt wel analyse van de counterfactual genoemd en kan gerealiseerd worden aan de hand van experimenteel onderzoek (zie verder par. 2.4).

ECB vindt ook in het centrale overheidsapparaat van de VS en Canada al meerdere decennia plaats, wat door Bemelmans-Videc (1992) is verklaard uit het feit dat sociale wetenschappers hier al geruime tijd deel van uit hebben gemaakt. Dit zou contrasteren met Europa, waar juristen lang hebben gedomineerd en waar bestendiging van evaluatiepraktijken pas sinds eind jaren '80 op gang is gekomen. Stern (2009) beschrijft in verwijzing naar Wollman (2003) drie 'golven' (*waves*) waarmee evaluatiepraktijken zich hebben verspreid in 16 landen, inclusief de EU: (1) proactieve beleidsvorming in de jaren 1960 en '70; (2) soberheid en kostenefficiëntie vanaf midden jaren '70 en (3) internalisering van evaluatiepraktijken sinds eind jaren '80. Ontwikkelingen in de eerste golf zijn in de jaren tachtig opgevolgd door de tweede golf landen die beleidsevaluatie inzetten als instrument voor rationeel bezuinigen cq. om overheidsuitgaven via onderbouwing te verantwoorden.[37]
De Europese Commissie (EC) heeft in veel lidstaten het evaluatiedenken (verder) aangejaagd via de verplichte landelijke evaluaties van de Structuurfondsen voor sociale en regionale ontwikkeling (Stern, 2009) – te weten het EFRO, het ESF en het Cohesiefonds. Daar passen netwerkmeetings in Brussel bij en verder vele evaluatiegidsen, met inbegrip van de Evalsed-gids (EC, 2013a) voor de evaluatie van sociaaleconomische programma's.

De aanvrager meer centraal
De evaluatie-aanvrager heeft internationaal het laatste decennium een steeds grotere verantwoordelijkheid toebedeeld gekregen in verhouding tot de onderzoeker.[38] Meer en meer dringt de notie door dat kwaliteit en gebruik van evaluatieonderzoek niet alleen afhangen van de onderzoekers, maar juist ook van de aanvragers ervan. Zij nemen immers bepalende beslissingen in en rond het evalueren, bijvoorbeeld over de reikwijdte en de onderzoeksvragen (in de opdracht) en gedurende de verdere begeleiding van het evaluatieproces. Bovendien vormen niet alleen het evaluatieproces zelf, maar ook de inhoud en opzet van *het te evalueren beleid* en de onderbouwing daarvan, randvoorwaarden voor evaluatiekwaliteit (o.a. HM Treasury 2011a; EC, 2013a).
De bepalende rol van de vraagzijde wordt al langere tijd erkend door grotere internationale organisaties zoals de EU, de VN en de Wereldbank en de laatste jaren ook door steeds meer landen. Dit blijkt zowel uit handboeken te gebruiken bij opdrachtformulering, aanbesteding en benutting (EU, VK) als ook uit '*ethical guidelines*' voor aanvragers. Zo heeft de Britse evaluatiegemeenschap, de UK Evaluation Society (UKES) in 2013 in haar *Guidelines for good evaluation practice* een aparte lijst *for Commissioners* opgenomen met daarin een aantal voorschriften (Bijlage 3). Bijvoorbeeld dat in het voortraject van de evaluatie verschillende stakeholders geraadpleegd moeten worden, de aanbestedingstrajecten eerlijk en open zijn[39] en dat aanvragers vertrouwen dienen te stellen in de (methodologische) kennis en vaardigheden van de onderzoekers.

[37] Vgl. voor een Nederlandse tijdlijn de box in Bijlage 4 bij dit rapport.
[38] Kushner (2005) constateerde alweer een decennium geleden dat een in het Verenigd Koninkrijk breed geïntroduceerd evaluatieraamwerk niet alleen beperkt toegesneden was op de context van beleidsmakers, maar ook '*any obligation on the sponsoring department'* ontbeerde. De onderzoeker kreeg volgens deze auteur alle eisen en verplichtingen op zijn bord.
[39] Met betrekking tot het offertetraject in publiceerde de Nederlandse Vereniging voor beleidsonderzoek overigens iets vergelijkbaars (VBO, 2004).

2.1.1 Evaluatie-instituties (organisatieonderdelen)

Dat er instituties zijn die zich specifiek op evaluatieonderzoek richten, is beslist niet nieuw. Wel veranderen ze vaak van vorm en zijn er periodes waarin de evaluatiefunctie meer in de belangstelling staat. In een bundel zetten Mayne et al. (red., 1992) alweer meer dan twintig jaar geleden internationale modellen en ervaringen met evaluatiebeleid en –instituties bijeen. Bemelmans-Videc concludeert in haar bijdrage op basis van een globale landenvergelijking dat belangrijke condities waaronder evaluatie-instituties zich blijken te ontwikkelen, zijn: algemene politieke en economische context (bijv. bezuinigingsimpulsen) en de politiek-bestuurlijke structuur, bijvoorbeeld: de aanwezigheid van centrale instituties zoals Rekenkamer of ministerie van Financiën die als *change agent* fungeren. Administratieve evaluatieprocedures zijn volgens haar met name veelbelovend als ze gekoppeld zijn aan het begrotingsproces. Maar de bestuurlijke *cultuur* – het kritisch tegen het licht willen houden van beleidsproducten – zou het meest bepalend zijn voor de snelheid waarmee evaluatievermogen zich ontwikkelt.
De VS en Canada kennen al sinds de jaren '70 gecentraliseerde evaluatie-units en een systeem waarin evaluatie een duidelijke plek heeft in de bestuurlijke structuur en het budgetteringsproces.[40] Canada hanteert sinds 1977 de norm dat alle departementen en agentschappen hun programma's evalueren en de bevindingen gebruiken om programma's te verantwoorden, te verbeteren of op te heffen (Bemelmans-Videc, 1992:11; Lahey, 2010). Het adjuncthoofd van de federale regering kreeg hiervoor de verantwoordelijkheid en de coördinatie is belegd bij de *Program Evaluation Branch* van de *Comptroller General*. Uitzonderingen daargelaten kende reeds rond 1990 elk Canadees beleidsdepartement en agentschap een *corporate evaluation unit*.
In 2002 is door academici voor het eerst een *International Atlas of Evaluation* uitgebracht (Furubo et al., 2002). Hierin zijn 21 OESO-landen met elkaar vergeleken in termen van de mate waarin structureel aandacht bestaat voor beleidsevaluatie, en in hoeverre men een zekere 'volwassenheid' (maturiteit) heeft bereikt in termen van cultuur, systemen en organisatiecapaciteit voor evaluatie. De negen indicatoren uit de Atlas zijn:

1. Evaluatie vindt plaats in meerdere beleidsdomeinen
2. Er is aanbod van evaluatieonderzoekers vanuit verschillende disciplines
3. Discussies en debatten naar aanleiding van evaluaties vinden plaats
4. Er is een nationale evaluatiegemeenschap
5. In de overheid bestaan institutionele voorzieningen voor het doen van evaluaties en voor de verspreiding van evaluatieresultaten
6. Binnen het parlement bestaan evaluatievoorzieningen
7. Pluralisme: binnen elk beleidsdomein houden verschillende mensen/ units zich met evaluatie bezig
8. De (Algemene) Rekenkamer voert evaluatieactiviteiten uit
9. Evaluaties richten zich niet alleen op *input* en *output,* maar ook op de *outcome* (einduitkomst).

De Atlas-exercitie is in 2015 herhaald voor 19 OESO-landen, gebaseerd op een survey onder vier tot vijf evaluatie-experts per land, in verschillende domeinen en functies (Jacob et al., 2015).
In dit laatste onderzoek vallen niet alleen de grote internationale verschillen[41] in evaluatiematuriteit op, maar ook die tussen beleidsdomeinen. Er is binnen

[40] Hoewel die in de VS van de jaren '80 weer deels werden 'wegbezuinigd'.
[41] Landen met een grote mate van evaluatie' maturiteit' (n = 15; 79%) in termen van de negen indicatoren zijn: Australië, Canada, Denemarken, Finland, Frankrijk, Duitsland, Israël,

overheden bijvoorbeeld veel tot zeer veel evaluatie in ontwikkelingssamenwerking en op het vlak van onderwijs en onderzoek; maar minder op het domein van 'het recht', financieel beleid, politiek, defensie, transport en buitenlands beleid. Evaluatiekennis en -ervaring sijpelen door van de meer geëvalueerde domeinen naar andere. Contextfactoren, zoals evidence-based beleidsinitiatieven en managementpraktijken, beïnvloeden de maturiteit.

Rekenkamers
In de VS, Nederland, Canada en Zweden spelen algemene rekenkamers al jaren een belangrijke rol als producent van evaluatieonderzoek en in discussies daarover. Dit in tegenstelling tot andere landen, zoals Frankrijk. Hier is de *Cour des Comptes* lang met positiebepaling bezig geweest in relatie tot de juridisch-bestuurlijke omgeving, waarin gaandeweg meer nadruk op evaluatieonderzoek en prestatiemanagement is komen te liggen (Jacob, et al., 2015).

Nationale en internationale evaluatiegemeenschappen
Volgens Cooksy (2012) en Jacob et al. (2015) kunnen evaluatiegemeenschappen in een land of regio verschil maken voor evaluatievermogen en de algehele evaluatiekwaliteit. Het aantal evaluatie-gemeenschappen en –netwerken groeit gestaag, wat wel gezien wordt als een globale trend in de professionalisering van (beleids)evaluatie.[42] Deze meer informele evaluatiegemeenschappen worden sinds 2001 overkoepeld door de IOCE, de organisatie voor internationale samenwerking in evaluatie.[43]
Qua beginperiode is net als bij de formele evaluatie-instituties ook een verschil te zien tussen de oorsprong van de evaluatiegemeenschappen in Europa (jaren 1990-2000) en de VS en Canada (jaren 1980). In de VS speelt de *American Evaluation Association* (AEA), in 1986 ontstaan uit een samenvoeging van de vroegere Evaluation Research Society en het Evaluation Network, een relatief grote rol als aanjager van normen en standaarden. De AEA heeft als missie om evaluatiepraktijken en – methoden in de VS en wijder te verbeteren, gebruik van evaluaties te bevorderen, het vak verder te professionaliseren en bij te dragen aan theorie- en kennisontwikkeling over effectieve interventies (AEA, 2013). Daartoe werkt de AEA onder meer met richtlijnen – *Guiding Principles* – en met de *Evaluation Policy Task Force* (EPTF), die sinds 2007 als doel heeft de doeltreffendheid van het federale evaluatiebeleid te beïnvloeden, onder meer via de *Evaluation Roadmap for a more effective government* (AEA, 2013). Dit stappenplan ter versterking van de rol van evaluatie in alle fasen van het beleidsproces overlapt sterk met het conceptuele model in het eerste hoofdstuk van dit rapport. De AEA heeft sinds kort een strategisch plan[44] met concrete acties en deadlines dat de ambities moet helpen realiseren. De website van de AEA bevat een indrukwekkende lijst van al dan niet wetenschappelijke evaluatiebronnen en handboeken.[45]
De *European Evaluation Society* (EES) bestaat sinds 1994 – Nederland heeft een belangrijke rol gespeeld in de oprichting ervan (Leeuw 2009). Het mandaat van de EES is om theorie, methodologie, praktijk en gebruik van evaluatie te stimuleren,

---

Japan, Nederland, Noorwegen, Zuid-Korea, Zweden, Zwitserland, Verenigd Koninkrijk en de Verenigde Staten. Van de onderzochte landen bleken die met een middelmatige gradatie minder talrijk (n = 4; 21%): het gaat om Ierland, Italië, Nieuw-Zeeland en Spanje. Geen enkel onderzocht land is ingedeeld bij de categorie lage evaluatiematuriteit (n = 0).
[42] www.ilo.org dec 2015.
[43] www.ioce.net.
[44] www.eval.org/p/cm/ld/fid=13.
[45] www.eval.org/p/cm/ld/fid=70. In Canada is er eveneens een ambitieuze evaluatiegemeenschap (CES) waarvan men gecertificeerd lid kan worden (www.evaluationcanada.ca). Tevens kan een gecombineerd Amerikaans/Canadees lidmaatschap worden aangevraagd.

begeleiden en te bevorderen in Europa en verder.[46] Zo geeft de EES een Nieuwsbrief uit, genaamd *Evaluation Connections* en zijn onder de tab *resources* net als bij de AEA evaluatiestandaarden te vinden van landelijke/regionale evaluatiegemeenschappen en daarnaast van internationale en supranationale organisaties. De EES organiseert onder andere trainingen en een tweejaarlijks congres.
In Nederland is VIDE de Beroepsvereniging voor toezichthouders, inspecteurs, handhavers en evaluatoren. Deze vereniging is in 2001 opgericht (Jacob et al., 2015, p. 17) ongeveer tegelijk met de Noorse en Zweedse. Binnen VIDE organiseert sinds enkele jaren het Evaluatorennetwerk bijeenkomsten over evaluatieproces en -aanpak. Ook is de brancheorganisatie van bureaus en instituten, de Vereniging voor Beleidsonderzoek (VBO), actief op het vlak van kwaliteitsbevordering en professionalisering.[47]

Intern of extern?
Bourgeois et al. (2011) concludeerden op basis van een literatuurreview dat interne onderzoekers geschikter lijken voor formatieve (tussentijdse, op leren gerichte) evaluaties, terwijl summatieve (ex post, eind-)evaluaties beter passen bij externe partijen. Sommige internationale evaluatiehandreikingen gaan kort in op de relatieve voor- en nadelen van interne dan wel externe evaluatie. Externe onderzoeksteams zullen vaker over specialistische expertise beschikken en als onafhankelijk worden beschouwd, wat de geloofwaardigheid van de evaluatie ten goede komt. Interne evaluatoren zijn zich meer bewust van institutionele kaders en vereisten en hebben sneller toegang tot informatie en sleutelinformanten. Maar zij beschikken mogelijk niet over specialistische expertise en zullen minder snel gezien worden als onafhankelijk (vgl. EC, 2013a, p. 39).

2.2 Programmering en evaluatieverplichtingen

Evaluatie als metgezel van beleid: beleidstheorie en monitoring
Meer en meer worden beleidsprogramma's vergezeld van evaluatieplannen (EC, 2013a) en wordt benadrukt dat al bij het ontwerpen van het programma zelf rekening dient te worden gehouden met mogelijkheden voor latere evaluatie. Zo moeten programma's voldoende gefocust zijn, met heldere en niet teveel doelstellingen (EC, 2015c; Witte Huis, 2014; HM Treasury 2011a).
Veel van de evaluatiekwaliteit en precisie hangt af van de condities die worden gecreëerd door beleidsmakers.
Tot die condities behoort allereerst het expliciteren van de beleidstheorie: hoe een beleid of wet tot de beoogde doelen moet leiden en wat daarbij de onderliggende aannames of principes zijn. Daarbij dienen de belangrijkste proces- en resultaat-indicatoren te worden benoemd. '(...) *the more reliable the programme theory, the clearer the indicators system, the easier and more precise will be the evaluation*. (...)' (EC 2015a, p. 30). Patton (2010) stelt dat evaluatie onderdeel is geworden van de ontwikkeling van het beleid en de achterliggende theorie: (...) '*Whether evaluators are present or not, the very notion of theories of change has become so prominent that evaluative thinking becomes built into the program design process (...)'*.
Niet alleen via de beleidstheorie dragen beleidsmedewerkers bij aan de kwaliteit van evaluaties, zij doen dat ook via monitoring. Daarbij worden uitkomsten op cruciale indicatoren vanaf het begin van de looptijd periodiek bijgehouden. *Monitoring embodies the regular tracking of inputs, activities, outputs, reach, outcomes, and*

[46] http://www.europeanevaluation.org/.
[47] www.beleidsonderzoek.nl .

impacts of (...) activities—at the project, program, sector, national (....) levels (Wereldbank, 2002).
Met name in de VS wordt evaluatie steeds vaker ingebouwd in de structuur van nieuwe en soms bestaande beleidsprogramma's. Het gebeurt bijvoorbeeld steeds vaker dat deelnemers resultaatgegevens registreren en monitoren (Witte Huis, 2014). Als monitoringsystemen niet al bij het begin van de uitrol van beleid of wetgeving zijn aangelegd, dan riskeert men bij de evaluatie gaten in de gegevens. Moet men na enkele jaren bijvoorbeeld terugvallen op reconstructie via navraag onder sleutelinformanten, dan bedreigen geheugeneffecten de validiteit. Lacunes in voor de evaluatie benodigde gegevens kunnen dus worden ondervangen door het tijdig – parallel aan het beleidsprogramma – aanleggen van een monitoringsysteem. In veel landen, inclusief het onze, en bij alle grotere internationale instituties worden periodieke monitors van beleidsresultaten vervaardigd op uiteenlopende terreinen.

Handreikingen bij de programmering van evaluaties
Enkele internationale instituties bieden handreikingen voor het programmeren van evaluaties. De EC heeft zichzelf het laatste decennium herhaaldelijk expliciet gecommitteerd aan ex post evaluatie van beleid en wetgeving op EU-niveau, bijvoorbeeld dat *al het beleid systematisch wordt geëvalueerd* om het probleemoplossend vermogen en legitimiteit te versterken (Mastenbroek et al., 2015).[48] Mastenbroek et al. constateerden dat die laatste toezegging is gerealiseerd voor in totaal 31% in de periode 2000-2002, weliswaar met grote verschillen in afdekking tussen DG's. Verder vonden zij dat nog veel is aan te merken op inhoud en methodologische evaluatiekwaliteit. Niettemin gaat ruim de helft van de evaluaties over einddoelen – en houdt dus meer in dan alleen een procesevaluatie of monitor.
Wat de *timing* betreft, worden doeltreffendheidsevaluaties idealiter – maar het is een uitdaging volgens de EC (2013a) – zo laat als nodig gepland om de beoogde resultaten de tijd te geven om zich te kunnen voordoen, maar zo vroeg als nodig voor bevindingen om terug te kunnen vloeien naar het beleidsproces. Dit verschilt per type interventie: bij sommige programma's vergt het 'materialiseren' van resultaten veel tijd en volgt de ex post evaluatie pas aan het einde van de levensduur, bij andere is de evaluatie van de doeltreffendheid of doelbereiking relatief vroeg mogelijk. Het Witte Huis geeft in dit verband voorbeelden van strategieën die het al tijdens de uitrol van beleid mogelijk maken om interventie- en vergelijkingsgroepen te creëren – zoals gebruik maken van gefaseerde invoering, of van 'lotingen' voor de bepaling van deelnemers bij beperkte beleidscapaciteit. In het laatste geval kunnen zelfs gerandomiseerde experimenten worden gerealiseerd.
Volgens de EC (2013a, p.29) zit de waarde van beleidsevaluatie net zoveel in het *stellen van de juiste vragen* als in het geven van precieze antwoorden. De gids beschrijft eisen aan centrale onderzoeksvragen naast verschillende mogelijke soorten vraagstellingen – beschrijvend, verklarend, voorspellend en kritisch.
Het betrekken van de (mogelijk vele) nauw betrokken en maatschappelijke stakeholders van het beleid in het vraagstellingsproces kan een last zijn, maar ook richting geven via de inventarisatie van kennisbehoeften.

Verplichtingen tot evaluatie van wet- en regelgeving

[48] Op het vlak van sociaaleconomische EU-evaluaties wil de EC voor de periode 2014-2020 het *behoefte-gedreven* programmeren handhaven, maar uitbreiden met een vereiste dat elk beleidsspeerpunt gedurende haar levensduur op zijn minst een keer ex post wordt geëvalueerd (EC, 2013a, p.37). Het evaluatieplan dient inzichtelijk te maken hoe elk speerpunt van beleid bijdraagt aan diens doelen – het is een strategisch document dat het beleidsprogramma vergezelt.

Nederland is niet het enige land met evaluatiebepalingen in wetten (Ar artikel 164). Volgens Jacob et al. (2015) is wereldwijd het gebruik van evaluatiebepalingen in zwang geraakt in landen als Denemarken, Duitsland, Nieuw Zeeland en Zwitserland. Soms maken deze deel uit van politieke onderhandelingen en helpen het tot stand brengen van controversiële wetgeving of coalitieakkoorden.
Het werken met een evaluatiebepaling is in de VS al sinds de jaren '60 van de vorige eeuw gebruikelijk. De aanvragers (of andere actoren) worden in de VS soms zelfs wettelijk verplicht tot het beschikbaar stellen van een budget en het vrijgeven van data. Ook worden de 'ontvangers' van beleid (met name subsidies) soms verplicht een rol in de evaluatie te vervullen (Witte Huis, 2014).
Twintig OESO-landen zeiden in 2014 een soort van standaard evaluatieverplichting te hebben (OESO, 2015a,b). Volgens de OESO is sinds 2008-2009 over het geheel van aangesloten landen geen opwaartse ontwikkeling meer zichtbaar in automatische evaluatieverplichtingen in wet- en regelgeving – het werken met horizonbepalingen laat wel een groei zien.
De OESO stelt verder dat ex post evaluatieverplichtingen in lidstaten vaak alleen van toepassing zijn op primaire wetten, niet op het totale regelgevingsbestand. Bovendien zijn wetsevaluatieverplichtingen in de helft van de landen te smal om ex post evaluatie van doelbereiking of doeltreffendheid te heten, omdat ze zich bijvoorbeeld alleen richten op specifieke (neven)effecten, zoals administratieve lasten. In nog minder landen is het verplicht om de externe consistentie van wettelijke normen met nationale en internationale regelbestanden te toetsen (in ons land evenmin). De OESO concludeert dat over de hele linie maar weinig OESO-landen een ontwikkeld systeem hebben voor ex post evaluatie van wet- en regelgeving.[49]

2.3 Middelen

Evalueren kan niet voor niets. Internationale instituties (onder andere) hebben strategieën en aanwijzingen geformuleerd om een evaluatiebudget vast te stellen en binnen de perken te houden.
De EC acht het evaluatiebudget afhankelijk van beleid en evaluatieopzet, waarbij de bepalende factoren zijn: de omvang, duur, reikwijdte en complexiteit van de interventie, de omvang en aard van de doelgroepen, de kwaliteit van de beschikbare monitoringsystemen en reeds beschikbare -gegevens en beoogde evaluatiemethoden (EC, 2015c, p. 257). Om tot een budgettair-proportionele evaluatie te komen, kan het volgens de Commissie helpen om een paar in termen van reikwijdte en methodologie verschillende evaluatie-opties tegen elkaar af te wegen. Bij sociaaleconomische EU-programma's (EC, 2013a, p. 38) acht men het lastig van tevoren in te schatten, hoeveel een evaluatie zal kosten. In geval van grootschalige maar relatieve 'standaard'-beleidsprogramma's geldt dat de evaluatie een klein aandeel van het totale beleidsbudget inneemt – meestal minder dan 1%. Bij innovatieve interventies (of pilots) is het leer- en participatieve element groter en kan het evaluatiebudget oplopen tot 10%. Het meest bepalend voor het budget acht de EC-gids de aard en reikwijdte van het vereiste evaluatiewerk.

[49] Het uitwendige, formele systeem zegt echter niet alles. Er zijn enkele landen die, ondanks een algemene evaluatieverplichting, geen enkele ex post evaluatie hebben opgeleverd; omgekeerd zijn er ook landen die zonder evaluatieverplichting wel aan ex post evaluatie hebben gedaan. In 2014 publiceerde de OESO het *Framework for Regulatory Policy Evaluation*. Dit raamwerk staat landen bij in het systematisch evalueren van het ontwerpen en evalueren van wetgevings(kwaliteits)beleid, afgezet tegen strategische reguleringsdoelen. Daartoe behoren ook de systemen voor ex ante en ex post evaluatie (naast bijv. administratieve lasten-reductieprogramma's).

In geval van budgetbeperkingen acht de Commissie het raadzaam om: (a) een indicatief evaluatiebudget te ramen en daarna de concurrerende offranten te laten aangeven wat zij daarbinnen zouden kunnen doen. Een alternatief is (b) om de offranten voorstellen te laten doen voor de op zijn minst te verrichten evaluatietaken (EC, 2015a). Ook wordt uitsplitsing van kosten naar taken aangemoedigd; net als naar de verschillende rollen binnen het onderzoeksteam (junior, senior).
De EC benadrukt in het kader van evaluatieprogrammering dat het te onderzoeken beleidsprogramma een *beperkt aantal doelstellingen* heeft (2015a, p.7). Dit om het aantal evaluaties en de proportionaliteit van evaluatie-inspanningen ten opzichte van het beleidsbudget te beteugelen. Deze handreiking bevestigt Patton's observatie dat evaluatie meer en meer wordt 'ingebakken' in het beleid: de anticipatie op ex post evaluatie maakt dat minder complexiteit wordt nagestreefd.
In een publicatie van het Amerikaanse Witte Huis (2014) worden voorbeelden gegeven van hoe met behulp van wettelijke voorschriften de evaluatie van federale programma's kan worden ondersteund en de evaluatiekwaliteit kan worden bevorderd. Het wettelijk voorschrijven van evaluatie*budgetten* als percentage van het programmabudget hoort daarbij.
Experimenteel-kwantitatief evaluatieonderzoek, gebaseerd op grote test- en controlegroepen, maakt harde uitspraken over doeltreffendheid mogelijk, maar is meestal duur en dat is een reden waarom dit type designs vaak niet wordt toegepast. Het Witte Huis (2014) beschrijft betaalbare strategieën om ze toch in te zetten. Zo kan men via het subsidiëren van relatief kleinschalige gedragswetenschappelijke experimenten bestaande programma's verrijken; zoals een (effectief gebleken) experiment uit 2011 met het aantrekkelijk presenteren van fruit in schoolkantines ter bevordering van gezonde voedingskeuzes. Deze inzichten kunnen vervolgens in de verdere beleidsvorming worden gebruikt – in casu in de versteviging van het *School Lunch Program*. Een tweede strategie is het federaal subsidiëren van experimenten met beleidspilots op lokaal of regionaal niveau, waar het evaluatiebudget vaak te klein is. Daarbij kan gebruik gemaakt worden van bestaande administratieve data. Zo is een drugsreductieprogramma onder voorwaardelijk vrijgelatenen in Hawaii experimenteel getoetst op doeltreffendheid in termen van drugsgebruik en recidive (Witte Huis, 2014, p. 276).[50] Succesvol gebleken aanpakken kunnen vervolgens worden toegepast in andere regio's of gemeenten.

2.4 Evaluatiekader

2.4.1 Standaarden en programma's

Uit verschillende meta-onderzoeken (ingeluid door Forss & Carlsson, 1997) blijkt dat in een aanzienlijke hoeveelheid evaluatiestudies conclusies worden getrokken over doeltreffendheid, terwijl achteraf tekortkomingen worden geobserveerd in termen van algemene geaccepteerde kwaliteitsstandaarden. Schwartz & Mayne (2005) wezen alweer geruime tijd geleden op toenemende politieke of organisatorische druk op evaluaties, bij achterblijvende professionaliteit. Desondanks werd volgens deze auteurs schijnbaar vaker dan voorheen geclaimd dat evaluatie en audits een objectief (eind)oordeel over de werkzaamheid van beleid geven.
Inmiddels hebben de meeste grote internationale organisaties algemene kwaliteitsstandaarden voor zowel evaluatieproces als -product gepubliceerd en ingevoerd (bijv. EC, 2013a, zie Tabel 3).

[50] "Evaluations employing quasi-experimental and randomized controlled trial designs were implemented at a cost of only $150,000 and $230,000, respectively (2014:276)."

Tabel 3 Criteria voor evaluatievermogen: product en proces

Kwaliteitscontrole: criteria voor het evaluatie*product*	Kwaliteitsborging: criteria voor het evaluatie*proces*
Tegemoetkomen aan de vereisten in de *Terms of Reference* (ToR) of startnotitie	Coherente en evalueerbare doelen
Relevante scope en reikwijdte	Goed opgestelde startnotitie / ToR (zie bijlage 3)
Verdedigbaar onderzoeksdesign en methoden	Adequaat proces van aanbesteding
Betrouwbare data gebruikt	Effectieve dialoog en feedback gedurende het hele evaluatieproces
Gedegen analyse	Adequate informatiebronnen beschikbaar
Geloofwaardige conclusies die gebaseerd zijn op data en analyse	Goed management en coördinatie door het evaluatieteam
Objectieve conclusies zonder bias en op basis van juiste beoordeling	Effectieve verspreiding van rapporten / outputs aan 'evaluatiecommissie' en beleidsmedewerkers / programmamanagers
Helder rapport met managementsamenvatting en ruwe data in de bijlagen	Effectieve verspreiding aan (andere) *stakeholders*

Bron: EC (2013a).

In het Verenigd Koninkrijk vervult het *Magenta Book* (HM Treasury, 2011b) een belangrijke rol in het tot stand brengen van evaluatievermogen. Het eerste deel is ontworpen voor beleidsmakers en legt de vereisten van een goede evaluatie uit, naast stappen die beleidsmakers kunnen zetten om die (beter) te faciliteren; deel B is technischer. Britse overheidsorganen / directies dienen ervoor te zorgen dat hun eigen handboeken of richtlijnen voor evaluatie *consistent zijn* met de principes in het *Magenta Book*.[51]
De United Nations Evaluation Group (UNEG) publiceerde de *Norms for evaluation* voor binnen het VN-systeem (2005a), aangevuld met de *Standards for evaluation* (UNEG, 2005b)[52] over wat een evaluatie geacht wordt te zijn. De primaire verantwoordelijkheid voor goed en zinvol evalueren wordt bij de (deel)afdelingen binnen de verschillende VN-organisaties gelegd: zij dienen in de naleving van deze normen hun eigen evaluatiebeleid te voeren. Daar hoort onder meer bij: een expliciete evaluatie 'missie-statement', deugdelijke en voortuitziende evaluatieprogrammering, borging van onafhankelijkheid en onpartijdigheid. Ook dienen zij capaciteits(functie)eisen voor interne evaluatieonderzoekers vast te stellen en dient het hoofd van de evaluatie-unit over ervaring te beschikken met het uitvoeren en managen van evaluaties.
De UNEG heeft de normen en standaarden in 2010 uitgewerkt in twee checklists, die zowel door onderzoekers als door aanvragers gebruikt kunnen worden bij startnotities – *Terms of Reference* (ToR) genoemd – en afgeronde

[51] Het aanvullende *Green Book* (2011b) zit dichter op de beleidscyclus en benadrukt eveneens dat beleidsmakers al rekening kunnen houden met latere evaluatie vanaf het ontwerpstadium van beleid.
[52] Het document met 'standards' is qua centrale boodschap vergelijkbaar als de 'norms', maar meer prescriptief en gedetailleerder. Hieronder vallen ook vereisten voor *Terms of Reference* (ToR) en vormvereisten voor de uiteindelijke evaluaties (schutblad, samenvatting) en voor ToR's.

evaluatierapporten (UNEG, 2010a,b; zie bijlage 3). Ook de EC (2015a) heeft dergelijke checklists uitgegeven.
De UNEG vereist dat – door onderzoekers – aan ethische standaarden zoals integriteit en bronbescherming wordt voldaan. Elke VN-evaluatie dient verder aan de volgende kwaliteitscriteria te voldoen.

''(…) Each evaluation should employ design, planning and implementation processes that are inherently quality oriented, covering appropriate methodologies for data collection, analysis and interpretation. 8.2. Evaluation reports must present in a complete and balanced way the evidence, findings, conclusions and recommendations. They must be brief and to the point and easy to understand. They must explain the methodology followed, highlight the methodological limitations of the evaluation, key concerns and evidenced-based findings, dissident views and consequent conclusions, recommendations and lessons. They must have an executive summary that encapsulates the essence of the information contained in the report, and facilitate dissemination and distillation of lessons.' (UNEG, 2005:9).

Veel van wat landelijke evaluatiegemeenschappen aanbieden aan criteria vindt zijn oorsprong in de sinds 1981 bestaande en steeds verder ontwikkelde (onderwijskundige) *Program Evaluation Standards* (PgES) (Yarbrough et al. 2011). Dat geldt bijvoorbeeld voor de Zwitserse en Duitse evaluatiegemeenschappen (SEVAL en DEgEval).
**Accuracy:* het vergroten van het waarheids- en betrouwbaarheidsgehalte van evaluaties;
**Feasibility*: het bevorderen van de doorwerking van evaluaties;
**Utility:* zorgen dat belanghebbenden iets hebben aan de evaluatie;
**Propriety*: de evaluatie ondersteunt dat wat eerlijk is, wettig, ethisch en in het belang van belanghebbenden en
**Accountability*: aanmoedigen van de juiste documentatie en focussen op verbetering en berekenbaarheid van evaluatieprocessen en -producten.

Elk criterium bestaat uit meerdere kenmerken, in totaal 30. De PgES zijn ontwikkeld in een uitgebreid proces van consultatie en review en zijn internationaal erkend (onder meer door Europese landen, de regering van de VS en door de evaluatiegemeenschappen van Australië/Azië en Afrika). Oorspronkelijk zijn de PgES sterk georiënteerd op het onderzoeken van de causale relatie tussen instrumenten en uitkomsten – zoals via experimenteel onderzoek.[53] Later zijn ook aanwijzingen voor constructivistische, democratische en gebruiksgerichte evaluaties gepubliceerd. In wat volgt beschrijven we nationale en EU-standaarden en programma's specifiek gericht op de evaluatie van wet- en regelgeving.

Post implementation review
In het Verenigd Koninkrijk en Australië wordt wet-en regelgeving sinds de jaren '10 geëvalueerd aan de hand van *Post Implementation Reviews* (PIRs). Zo zijn beleidsdirecties in Zuid-Australië verplicht om een PIR te doen voor alle wetten en wetswijzigingen met een grote impact op economie en bedrijfsleven. Voorts zijn ze tot een PIR verplicht bij het introduceren, schrappen en substantieel wijzigen van wetten als een ex ante analyse, de *Regulatory Impact Statement*, achterwege is gebleven. Ambtenaren beschikken over een handboek – het *Better Regulation Handbook* – voor zowel ex ante als ex post evaluatie van wet- en regelgeving (Regering van Zuid-Australië, 2011). Specifiek voor ex post evaluatie is een extra begeleidende notitie, de *PIR Guidance note* ontwikkeld (Regering van Zuid-Australië, 2014). De voorgestelde evaluatiebenaderingen zijn gebaseerd op 'best practices'. Op deze manier moet de Zuidaustralische regering voldoen aan de '*best practice*

[53] vgl. AEA, 2004; zie voor designs par. 2.4.2.

regulatory principles' van de overkoepelende Raad van Australische regeringen (Regering van Australië, 2011, p.2).
Het is de bedoeling dat beleidsambtenaren zelf de PIR opstellen. De *guidance note* bevat zowel korte als uitgebreide aanwijzingen voor het evaluatieproces, de minimum inhoudsvereisten (waaronder relevantie, impact in termen van (administratieve) kosten en einddoel) en vereisten ten aanzien van objectiviteit, zoals de aanwijzing om zowel positieve als negatieve uitkomsten te benoemen. Alle PIR's worden na beoordeling gepubliceerd. Als doelgroepen worden beschouwd besluitvormers en andere stakeholders.
Op de resulterende rapporten vindt stevige *compliance monitoring* plaats door een eenheid die rechtstreeks onder de centrale regering valt: the *Office of best practice Regulation* (OBPR).[54] Naast controleren staat het OPBR departementen bij in het voldoen aan vereisten en vergroten van de kwaliteit van evaluatie van regelgeving – het vervult dus een tweeledige rol.
In het Verenigd Koninkrijk in ongeveer dezelfde periode eveneens een format voor PIRs van wetgeving tot stand gekomen; sinds 2007 neemt men zich voor om het aantal ex post studies te vergroten, onder meer door middel van een *Impact Assessment Guidance* te gebruiken door de evaluatie-aanvrager. De aanleiding hiervoor is eerder onderzoek door de Britse Rekenkamer, gebaseerd op 233 wetgevingsinstrumenten (*National Audit Office* (NAO), 2009). Hieruit bleek dat in maar de helft van de vooraf gedane evaluatietoezeggingen daadwerkelijk een PIR was uitgevoerd door beleidsdirecties.[55] Er zijn anno 2016 op Internet slechts twee beknopte verslagen van Britse PIR's terug te vinden.[56] De kernvragen daarin zijn: (1) wat waren beleidsdoelen, succescriteria en beoogde effecten? (2) Hoe en in hoeverre zijn deze bereikt – geef daarbij ook onbedoelde effecten aan (3) Voor wanneer diende de evaluatie van dit beleid worden opgeleverd? Is die datum niet gehaald, waarom niet?[57]

Better/ Smart Regulation (EU)
De EC-website over het programma *Better Regulation* – in 2010 omgedoopt tot *Smart Regulation* (Smismans, 2015) – kent een apart evaluatiegedeelte. De EC publiceert een algemene evaluatieplanning inclusief voor evaluatie van regelgeving, geordend naar DG.[58] Door te evalueren onderzoekt de Commissie naar eigen zeggen kritisch of EU-beleidsinspanningen 'fit-for-purpose' zijn en of ze, tegen minimale kosten, de beoogde veranderingen bij bedrijven en burgers teweeg brengen en bijdragen aan de rol van de EU in de wereld. De Commissie past naar eigen zeggen het *evaluate first*-principe toe om zeker te stellen dat eerdere beleidservaringen worden meegewogen in de totstandkoming van nieuwe beleidsbesluiten (Smismans, 2015, p. 11). Evaluaties zijn daarmee een essentieel onderdeel geworden van de beleidscyclus. Ze zijn onder meer vereist op grond van de *Financial Regulation* (art. 30 4e lid) en wettelijke bepalingen.[59]

---

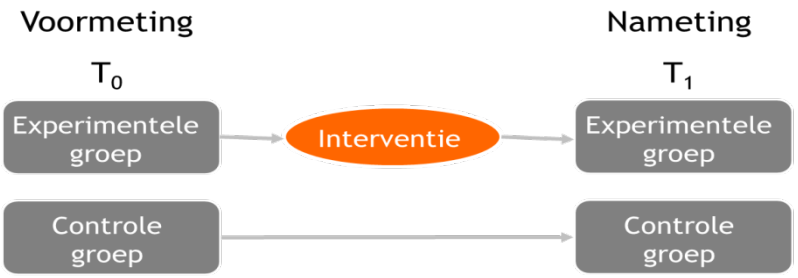
[54] De oordelen zijn: *compliant* of *non-compliant*. http://ris.dpmc.gov.au. Het is deze website waarop ook de PIRs worden gepubliceerd.
[55] In 45% van de verplichte (ex ante) *Regulatory Impact Assessments* (RIA) was die toezegging gedaan. Anderzijds is ook in 5% van de gevallen waarin geen toezegging was gedaan, toch ook een PIR verricht. In totaal is in ongeveer 54% van alle dossiers een soort van evaluatie verricht; vaak naar aanleiding van externe prikkels, zoals EU-verplichtingen.
[56] www.legislation.gov.uk/ukia?stage=Post%20Implementation%20Review (aug. 2015).
[57] In een aparte sectie wordt de aanpak van de PIR toegelicht, gevolgd door een beschrijving van het probleem en de rationale voor interventie, de kosten, de risico's/aannames, en de 'wider impacts'. Dit alles wordt gevolgd door een conclusie. Een samenvatting van de antwoorden op de hoofdvragen is op het voorblad van de PIR te vinden.
[58] Voor de programmering van 2015 zie: http://ec.europa.eu/smart-regulation/evaluation/docs/evaluation_forward_plan_2015_en.pdf.
[59] Bijvoorbeeld de *Council Regulation* (EU) 2015/323, van toepassing op het elfde Europese sociale ontwikkelingsfonds (ESF).

Alle EC-evaluaties dienen van hoge kwaliteit te zijn en in lijn met de principes zoals beschreven in de *Better Regulation guidelines* (EC, 2015b). Deze richtlijnen dekken niet alleen evaluaties maar de hele 'regulatory cycle' af. De *Toolbox* (EC, 2015c) vergezelt de *guidelines* met aanvullingen op specifieke onderdelen.[60] Sinds 2010 bevat de *Toolbox* het instrument van de 'fitness-checks'. Dit zijn integrale evaluaties van een beleidsterrein, waarin wordt nagegaan hoe meerdere samenhangende wettelijke bepalingen bijdragen aan beleidsdoelen – vormgegeven in termen van kosten en baten. Deze checks zijn geschikt in het aanwijzen van overlap, inconsistenties, synergie en cumulatieve impact van regelgeving. Hoewel zeer intensief, kunnen de fitness-checks op langere termijn helpen bezuinigen op 'losse' evaluatiestudies (EC, 2015c).[61] De fitness-checks passen in het bredere 'REFIT'-programma van de EU, om de hele regelgevingsinventaris te blijven herzien om te zorgen dat administratieve lasten tot een minimum beperkt blijven en alle versimpelingsopties bekend zijn en toegepast worden (vgl. OESO, 2015a; Smismans, 2015).[62]

2.4.2 Evaluatiedesigns

Het barst inmiddels van de ex post evaluatiehandboeken en –leidraden. Dit is ook in de leidraden zelf meermaals gesignaleerd (o.a. EC 2013a).

'Stromingen' in de aanpak
Cooksy & Caracelli (2009) leggen uit dat kwaliteitsstandaarden afhankelijk zijn van *welke evaluatieaanpak* men volgt. Zo wordt in niet-experimentele benaderingen de betrokkenheid van stakeholders bij de evaluatie hoger gewaardeerd dan in de experimentele.



Figuur 1. Black-box benadering: basistype experiment. *Werkt het?*

In het design, het methodologische ontwerp van ex post beleids- en wetsevaluaties onderscheiden we twee hoofdstromingen, die samenhangen met het doel van evaluaties (leren dan wel verantwoorden) (vgl. EC, 2015a en hoofdstuk 1 van dit rapport). Het is een ideaaltypisch onderscheid, in de praktijk kunnen de stromingen elkaar aanvullen. Tabel 4 zet de verschillen in benadering op een rij.
Summatieve evaluaties zijn gericht op de vorming van een (eind)oordeel over de werking van beleid of wetgeving; aansluitend bij het verantwoordingsdoel dat veel

[60] Wat evaluatie betreft gaat Hoofdstuk 3 van de *Toolbox* over specifieke effecten (bijv. administratieve lasten of milieu) en hoofdstuk 6 over ex post evaluatie in brede zin.
[61] Zie: Tool #38, http://ec.europa.eu/smart-regulation/evaluation/docs/fitness_checks_2012_en.pdf.
[62] Overigens heeft een netwerk van vijf landelijke 'ACTAL's', genaamd RegWatchEurope, zich enigszins kritisch geuit over REFIT (RegWatchEurope, 2014). Het programma luidt volgens deze groep weliswaar een cultuurverandering in, maar toont weinig commitment aan duidelijke en tastbare resultaten. Die laatste zouden een meer integrale aanpak vergen, die aansluit op wat er binnen lidstaten gebeurt.

evaluaties hebben. Komt het door de wet dat we xx% doelbereiking hebben, of zijn andere factoren hiervoor verantwoordelijk?
Daarentegen houden lerende (formatieve) evaluaties zich bezig met kennis- en inzichtvergroting en hebben daardoor vaak een verklarende component. Hoe kan het dat ...? Die verklaring valt bijvoorbeeld te bereiken via reconstructies van gebeurtenissen, of te onderzoeken of zich patronen voordoen van resultaten van beleid in verschillende contexten.

Tabel 4 *Black box* versus *clear box* evaluatiebenaderingen

	Black box (toetsende, experimentele designs)	**Clear box** (verklarende designs)
Causaliteit	Attributie: in welke mate werkt het beleid?	Contributie: *hoe* draagt het beleid bij?
Verklaring	Weinig aandacht voor *hoe* het beleid werkt	Focus op mechanismen, ingebed in context
Uitkomsten	Beoogde uitkomsten worden van tevoren gedefinieerd	Idem, en: afhankelijk van context
Context	De context wordt 'gecontroleerd' om netto-effect van beleid te destilleren *Does it work?*	Context heeft verklarende waarde *What works for whom and under what circumstances?*
Leren van evalueren	Ja, maar ceteris paribus (a-contextueel)	Ja, maar moeilijk overzicht te bewaren over diverse C, M, en O

Bron: Pattyn & Verweij (2014), met aanpassing.

Summatieve evaluatie (geassocieerd met verantwoorden)
Aan de overkant van de oceaan, met name in de VS, wordt in ex post evaluatie traditioneel de voorkeur gegeven aan designs waarmee resultaten aan beleid kunnen worden geattribueerd – ook wel *rigorous methods* gedoopt (Wereldbank, 2006; GAO, 2012; Witte Huis, 2014). Voor evaluatie van maatschappelijke (welvaarts)effecten wordt *counterfactual*-analyse voorgestaan: het vergelijken van resultaten bij een groep die wel aan de te evalueren beleidsinterventie(s) is 'blootgesteld' (de experimentele groep) met een groep die dat niet is geweest (de controlegroep). Daarmee kan de *counterfactual*-vraag worden beantwoord: wat zou de uitkomst zijn geweest *zonder* het gevoerde beleid? Dit vergt experimentele of quasi-experimentele evaluatieaanpakken. Experimentele designs waarbij proefpersonen willekeurig (gerandomiseerd) aan de experimentele of de controlegroepen worden toegewezen, worden wel *randomised controlled trials* (RCT's) genoemd. Sommigen beschouwen dit design als het hoogst haalbare, omdat dankzij die randomisatie zowel de theoretisch kenbare als niet-kenbare externe invloeden op de beleidsresultaten worden uitgeschakeld. De Wereldbank laat – met behulp van praktijkvoorbeelden – zien hoe behalve met RCT's ook met quasi-experimenten kan worden gewerkt, en hoe daarbij twee belangrijke methodologische problemen, contaminatie en selectie-bias[63] kunnen worden bestreden.
Hoe en onder welke contextuele omstandigheden effecten tot stand zijn gekomen, blijft een *black box*; de kracht van experimentele designs zit hem in de relatieve zekerheid waarmee resultaten aan beleidsinspanningen kunnen worden toegeschreven. Natuurlijk kan hier ook van geleerd worden, in de zin van te weten

[63] Bij contaminatie krijgt de controlegroep onbedoeld iets mee van het beleid; bij selectiebias verschillen controle- en experimentele groep systematisch van elkaar. In beide gevallen kunnen de netto-effecten van het beleid niet valide worden vastgesteld.

komen hoe succesvol een aanpak is geweest en of en hoe een beleidsproject kan worden voortgezet. Afhankelijk van het specifieke design kan men in deze stroming ook achterhalen in hoeverre alternatieve aanpakken – blijkend uit dezelfde[64], of parallelle analyses – meer opleveren.
De Wereldbank (2006) heeft *best practices* van dergelijke doeltreffendheidsevaluaties op een rij gezet. Ook de EC (2013b) heeft een gids over ex post *counterfactual impact evaluations* of CIE met zowel procesmatige als inhoudelijke aanwijzingen voor (met name) aanvragers. De aanleidingen voor de EC-gids, die betrekking heeft op het ESF, zijn: gebrek aan kennis van (quasi-)experimentele benaderingen binnen de EU en de lidstaten, gebrek aan aanbod van externe, gekwalificeerde onderzoekers om ze uit te voeren, juridische barrières, bijvoorbeeld bij het verkrijgen van de benodigde data en de behoefte aan een meer gedegen evaluatieprogrammering.

Tabel 5 Evaluatiedesigns voor het achterhalen van de doeltreffendheid van verschillende typen beleid*

Soort design	**Vergelijking(-sgroep) ter controle voor invloeden van buitenaf**	**Het meest geëigend voor:**
Monitoring en evaluatie van proces en uitkomsten	Prestatie- indicatoren, gestelde doelen en standaarden, bijvoorbeeld • op het vlak van R&D: relevantie, kwaliteit en prestaties • standaarden van productiviteit, kostendoeltreffendheid en efficiency	Onderzoek, handhaving, informatie en statistische programma's, en 'rijpe' doorlopende beleidsprogramma's met: • een landelijke en complete dekking • weinig alternatieve verklaringen voor einduitkomsten
Voor- en nameting / tijdreeksanalyses Eén groep	Uitkomsten voor deelnemers vóór en na de interventie • verzamelen van verschillende resultaatgegevens op verschillende momenten in de tijd • statistische correcties of modellen ter controle voor alternatieve verklaringen voor resultaten	Beleid en wetgeving met: • duidelijk afgebakende interventies op verschillende startmomenten • landelijke en complete dekking • waarbij het 'random' toewijzen van deelnemers aan controle- en experimentele groep(en) niet uitvoerbaar is, of onethisch
Quasi-experimenten: Vergelijkingsgroepen	Uitkomsten voor deelnemers en op de volgende kenmerken g*ematchte* vergelijkingsgroep : • de kenmerken zijn aannemelijke (of theoretische gefundeerde) alternatieve verklaringen voor verschillen in uitkomsten tussen groepen • er is een nulmeting en een nameting	Beleid met: • duidelijk afgebakende interventies die kunnen worden gestandaardiseerd en gecontroleerd • beperkte dekking waarbij random' toewijzing van deelnemers aan controle- en experimentele roep(en) niet uitvoerbaar of onethisch is
Gerandomiseerde experimenten (RCT): Controlegroepen	Uitkomsten voor een 'random' toegewezen experimentele en een niet-deelnemende controlegroep • er is een nulmeting en een nameting	Beleid met: • duidelijk afgebakende interventies die kunnen worden gestandaardiseerd en gecontroleerd • beperkte dekking waarbij random' toewijzing van deelnemers aan controle- en experimentele groep(en) uitvoerbaar en ethisch is

* Vergelijkbaar, maar enigszins anders ingedeeld: EC (2013a, p. 24-25) en Wereldbank (2004, p. 24).
Bron: GAO (2012) – aangepast met oog op de Nederlandse context.

[64] Behalve een controlegroep kan binnen één design ook gewerkt worden met vergelijkingsgroepen die een andere *treatment* ontvangen.

De EC voegt hieraan nog een op de praktijk toegesneden quasi-experimenteel design toe, dat handig gebruik maakt van *drempelwaarden* in 'scores' ter afbakening van doelgroepen van het beleid, waardoor die deelgroepen toch heel sterk op elkaar lijken. Zo is een valide vergelijking mogelijk van groep A (bijv. score 39, komt niet in aanmerking voor subsidie) met groep B (score 40, komt wel in aanmerking) waarbij groep A als controlegroep kan fungeren.
Soms is het vormen van een controlegroep niet mogelijk. De GAO (2012) beschrijft twee alternatieve designs om toch inzicht te krijgen in de doeltreffendheid:

1. *Tijdreeksanalyse*: door met behulp van meerdere metingen voor en na de beleidsinterventie statistisch te onderzoeken of er systematische verschillen zijn opgetreden na het introduceren van de interventie bij (alleen) de doelgroep. Naarmate er meer metingen in de tijd zijn, worden toevallige invloeden op de uitkomstmaat steeds meer uitgesloten.
2. Een *cross-sectionele studie* (niet-longitudinaal), die gebruik maakt van uitkomstmaten bij deelgroepen van de doelgroep, die *in verschillende mate* aan de interventie zijn blootgesteld. Statistische controles dienen ervoor te zorgen dat alternatieve verklaringen voor verschillen in uitkomsten zoveel mogelijk worden uitgeschakeld (vgl. AEA, 2004).

Niet-experimentele evaluatie
Dat gerandomiseerd experimenteren in het toetsen van doeltreffendheid wel een nastrevenswaardige, maar niet altijd realistische evaluatieopzet is, blijkt onder andere duidelijk uit een krachtig *Policy Statement* van de Europese Evaluatiegemeenschap (EES) van december 2007[65], waarin wordt gereageerd op de groeiende neiging in de evaluatiewereld om gerandomiseerd experimenteel onderzoek als de gouden standaard (o.a. Farrington, 1997) van doeltreffendheidsonderzoek te beschouwen. De EES geeft weer onder welke praktijkomstandigheden experimenteel werken vaak niet haalbaar is; zoals:

- Bij complexe projecten, waarbij meerdere gelijktijdige, en/of snel veranderende interventies in het spel zijn. Wetgeving kan daar ook toe worden gerekend (vgl. Bussmann, 2010).
- Als de contexten waarop hetzelfde beleid wordt toegepast (bijv. in geval van meerdere regio's), onderling sterk verschillen. Dat maakt het uitschakelen van contexteffecten niet alleen minder haalbaar, maar ook minder relevant omdat resultaten maar beperkt generaliseerbaar zijn. Want als een effect is gevonden na controle voor context met betrekking tot deelbevolking A, wat weten we dan over het effect op deelbevolking B?
- Als het vormen van experimentele en controlegroepen vanwege bijvoorbeeld ethische bezwaren niet haalbaar is (bijv. ontzeggen van hulp aan groepen slachtoffers; of het ontzeggen van wettelijke rechten aan een groep burgers, teneinde een controlegroep te vormen; vgl. AEA, 2004)

Wijzend op de literatuur pleit de EES in dit soort situaties voor evaluatieonderzoek waarbij niet (of niet alleen) wordt ingezet op attributie maar op 'plausibele contributie': het inzichtelijk maken en zo goed mogelijk staven van verklarende mechanismen achter behaalde resultaten, zowel met betrekking tot beleid als externe factoren. De beleidstheorie is daarbij leidend.
Dit verklaart het wel gebruikte antoniem *clear box* of *white box*–evaluatiebenaderingen (o.a. Scriven 1994; Pawson & Tilley, 1997; Mayne, 2008; Astbury & Leeuw, 2010), die kunnen dienen als alternatief of ter aanvulling of vervanging van de experimentele, 'black box' benadering. Wat werkt *hoe* en in welke contexten cq. onder welke condities?

[65] http://www.europeanevaluation.org/sites/default/files/EES%20Statement_0.pdf.

Deze benaderingen worden vaker geassocieerd met 'leren' en kunnen op nagenoeg ieder moment in de beleidscyclus worden toegepast. Vooral het afgelopen decennium zijn ook de evaluatiebenaderingen in deze 'familie' doorgedrongen tot evaluatiesystemen over de wereld (o.a. Wereldbank 2006, p.10; AEA, 2004; EC, 2015a, p. 30). Meerdere internationale organisaties laten belang en functie zien van dit type ex post evaluaties, die licht werpen op de causale keten van input, output en (tussen)uitkomsten: *'focus on the factual besides the counterfactual'*. Bij voorkeur wordt hierbij gebruik gemaakt van een combinatie van kwalitatief en kwantitatief onderzoek (Wereldbank, 2006).

Participatieve benaderingen
Daarnaast zijn er participatieve evaluatie-aanpakken, waarbij de nadruk vaak op de 'stakeholders' ligt en hun verschillende perspectieven op de werkelijkheid (Wereldbank, 2004a; Bovens et al., 2008; GAO, 2012). Met behulp van participatieve methoden kunnen verschillende (doel)groepen aangeven wat veranderingen zijn geweest als gevolg van een beleidsproject, wie heeft geprofiteerd en wie niet, en wat volgens hen sterke en zwakke punten zijn (GAO, 2012). Dit kan overigens ook een aanvullende strategie zijn voor verklarende benaderingen.
Voor een deel komen deze benaderingen voort uit frustratie over de soms beperkte relevantie en het politiek-strategische gebruik van andere typen evaluaties, zoals de experimentele (Bovens et al., 2008). Een participatieve aanpak zou de bruikbaarheid en de relevantie van evaluatiebevindingen vergroten en zo uiteindelijk bijdragen aan een breed gedragen beleidsstrategie. Hier ontstaat echter wel een spanningsveld met de wetenschappelijkheid die andere benaderingen kenmerkt (vgl. Wereldbank, 2004a).
Leeuw (2003) beschreef een participatieve methode voor het reconstrueren van de beleidstheorie. Het latere Arena-model (de Boer en Larsen, 2013), vaak onderdeel van ex-ante (bijv. MKBA-) exercities, past in deze lijn.

Procesvaluatie
De Amerikaanse federale Rekenkamer (GAO, 2012) biedt handvatten voor '*implementation evaluations'* ofwel procesevaluaties (Tabel 6). De centrale vraag bij procesevaluatie is of beleidsactiviteiten zijn uitgevoerd als beoogd en de implementatie daadwerkelijk de beoogde populaties of problemen bereikt. Procesevaluaties vergen een maatstaf waartegen implementatie-prestaties worden afgezet. Die maatstaf kan men ontlenen aan de betreffende wet- en regelgeving, de beleidstheorie, professionele standaarden of aan verwachtingen bij stakeholders.

Tabel 6 Procesevaluatie: welke aanpakken passen bij welke vraag?

Evaluatievraag	**Evaluatiedesign**
Is het programma uitgevoerd zoals beoogd?	Vergelijk implementatie-activiteiten / -prestaties met aan beleid of stakeholders ontleende maatstaf(-aven)
Zijn uitvoerings- of managementproblemen opgetreden?	- Vergelijking van prestaties met verwachtingen in termen van kwaliteit, kosten of efficiëntie - Onderzoek naar variatie in termen van kwaliteit /prestatie tussen contexten (bijv. gemeenten), verstrekkers, of sub-doelgroepen
Waarom worden de beoogde programmadoelen niet – of niet meer – bereikt? Wat zijn knelpunten in de	- (Diepte-) case studies - Onderlinge vergelijking van de implementatie(-typen) door bijv. gemeenten – zoeken naar patronen

implementatie?	
Gegevensbronnen	
- Beleidsstatistieken (bijv. gemiddelde inkomensniveau uitkeringsdoelgroep) - Administratieve gegevens (bijv. ter toetsing van kwaliteit of snelheid van aanmeldings- en selectieprocedures) - *Surveys* onder een representatieve selectie van (lokale) uitvoerders (bijv. is het wettelijk kader / de verstrekte toelichting erop voldoende duidelijk?) - Combinaties van gegevenstypen verdienen de voorkeur (validiteit, diepgang waarmee vragen worden beantwoord)	

Bron: GAO (2012, p. 32).

Fit-for-purpose
Ongeacht de genoemde 'stromingen' is in de praktijk elk evaluatiedesign weer anders, omdat het wordt toegespitst op het voorliggende programma en de gestelde evaluatievragen (GAO, 2012). De EC (2013b, p. 27) stelt in een handreiking dat een solide evaluatieprogrammering, naast experimentele of *black-box* evaluatie óók bevat: theorie-gestuurde evaluatie, procesevaluatie, en doelmatigheids- of efficiency-analyse. Deze verschillende aanpakken worden gezien als complementair. *Fit-for-purpose* is steeds vaker het credo: internationale instituties en landen pleiten niet meer zozeer voor een bepaald type design, maar voor keuzes hieromtrent die goed aansluiten op, met name: (1) de gestelde beleidsdoelen en -prioriteiten, (2) het type instrumentarium, (3) het stadium waarin het beleid zich bevindt en (4) het evaluatiedoel (o.a. EC, 2013a). Het bewerkstelligen van een op de situatie toegesneden evaluatiedesign vergt een actieve rol van de opdrachtgever, die een bewuste afweging dient te maken. Onder meer in het Verenigd Koninkrijk (HM Treasury, 2011a,b) en bij de EC (2013a) wordt expliciet onderscheid gemaakt tussen verschillende te kiezen evaluatietypen afhankelijk van kennisbehoefte en situatie. Behalve de genoemde experimentele, proces- en verklarende evaluatiestudies op basis van de beleidstheorie staan 'rapid' evaluaties in het rijtje op basis waarvan relatief snel kan worden bijgestuurd (RIPI, Kautto & Similä, 2005; en recenter ook *real-time evaluation*, o.a. Wereldbank, 2010).

Design van wet- en regelgevingsevaluatie
In een expert paper voor de OESO pleit Coglianese (2012) voor meer robuuste uitspraken over de doeltreffendheid van wet- en regelgeving *(regulation and regulatory policv).*[66] Overheden hebben betrouwbare indicatoren nodig die de volle breedte van zowel positieve als negatieve uitkomst(indicator)en afdekken. Gepleit wordt voor evaluatiedesigns met een experimentele inslag – of anders met statistische controles voor externe factoren – om uitkomsten te kunnen attribueren aan de wetgeving in kwestie. Daarnaast is er ook een pleidooi voor vaker (ex ante) experimenteren met wetgeving. De auteur maakt echter niet expliciet duidelijk voor welke specifieke uitdagingen de evaluator van wet- en regelgeving komt te staan in vergelijking met programma's of projecten in het algemeen. Hoe valt bijvoorbeeld het zo veel bepleite experimentele design waar te maken als wetten al voor de hele bevolking zijn ingevoerd en het vormen van een controlegroep niet meer kan? Hoe om te gaan met eventuele ethische bezwaren – al dan niet geobjectiveerd?
Ook de wetenschappelijke literatuur bevat weinig expliciete handvatten voor onderzoeksdesigns specifiek gericht op het evalueren van wet- en regelgeving. De spaarzame academische publicaties laten vooral de worstelingen zien bij pogingen om de aanpak van de juridisch georiënteerde wetsevaluaties te verrijken met meer algemene inzichten en standaarden uit de evaluatiewereld.
Bussmann (2010) pleit voor wijze benaderingen in het evalueren van de complexen van regels en interventies die wetten heten. Maar ook al benadrukt hij het complexe

[66] Met de term 'regulatory' wordt normstellend, als in wet- en regelgeving bedoeld.

karakter van wetgeving als evaluatieobject, ook hij bespreekt de juridische component zeer globaal. Niettemin vallen een paar waardevolle handreikingen voor wetsevaluatie aan het betoog te ontlenen:

- Plausbiliteitsanalyse verrichten, in plaats van proberen de (mate van) doeltreffendheid van een wet als geheel te bewijzen. Dit komt neer op het bijeen puzzelen van evidentie uit nieuw en bestaand onderzoek tot de '*chain of effects*'
- Gebruik maken van methoden analoog aan multiple case study onderzoek en zo variatie in kenmerken van verschillende doelgroepen van dezelfde wet te benutten in het bestuderen van de werking.

De Amerikaanse federale Rekenkamer (GAO, 2012) biedt een aantal bruikbare handvatten voor designs van doeltreffendheidsevaluaties van complexe, grootschalige federale beleidsprogramma's (Box 2). Deze programma's vertonen overeenkomsten met veel Nederlandse (kader)wetgeving.

Box 2 Problemen en handvatten voor de evaluatie van complexe progrmma's (GAO, 2012)

- Vaak kan geen controlegroep meer worden gevormd. Juist bij complexe wetgeving wordt dit vaak als bezwaar genoemd.[67]
    - Voor 'hardere' statistische uitspraken zijn dan mogelijk nog wel tijdreeksanalyses mogelijk;
    - ...of analyses die gebruik maken van de variatie in de 'mate van interventie-intensiteit' tussen groepen ontvangers
- Er kunnen geen overkoepelende 'uniforme' indicatoren worden aangelegd voor de meting van einduitkomsten in verschillende deelcontexten (bijv. gemeenten)
    - De GAO suggereert drie oplossingen voor dit probleem: (1) samenwerking tussen de beleidsuitvoerders binnen regio's of gemeenten in de data-registratie, om te zorgen dat de benodigde evaluatiegegevens voldoende consistent zijn; (2) bij al 'vaste' lokale dataregistratiesystemen de data omzetten in een gedeeld format; (3) aanvullende dataverzameling (bijv. eenmalige survey) om lacunes aan te vullen
- Maatschappelijke einduitkomsten (de outcome) worden pas na heel lange tijd zichtbaar (bijv. incidentie van een bepaalde ziekte als gevolg van een vaccinatieprogramma)
    - Zoom in op tussendoelen of output-indicatoren; en combineer deze met evidentie over de relatie daarvan met de outcome
- Maatschappelijke einduitkomsten doen zich zelden voor (bijv. bij toezichtconstructies, contraterrorisme of beleid ter preventie van rampen)
    - Richt je op indicatoren zoals de kwaliteit van preventieplannen van uitvoerders
    - Doet zich (toch) een incident voor, richt daar de evaluatie op – op het incident zelf, of op de crisisbeheersing achteraf
- Het beleid is zeer 'indirect' van aard en opzet, bijv. een periodieke bijdrage aan een fonds, of een subsidieregeling met breed geformuleerde doelen, die bijdraagt aan al bestaande financieringsstromen van gemeenten of bedrijven. Een harde doeltreffendheidsstudie is dan niet passend.
    - Beschrijf de variatie in hoe lokaal wordt omgegaan met het beleid. Hier zal het vooral gaan om een *impressie* van doeltreffendheid, geen harde meting

[67] Hoewel dit niet gerechtvaardigd is, als men al in een vroeg stadium met het vormen van controle- en experimentele groepen had kunnen beginnen. Bovendien kan in een vroeg stadium van (totstandkoming van) wetten – vaker dan nu – zou kunnen worden geëxperimenteerd; eventueel gefaciliteerd door een speciale experimenteerbepaling.

    - Meet nationale verbeteringen in overkoepelende output-indicatoren – deze gegevens kunnen dan ook fungeren als benchmark (richtpunt) voor gemeenten of bedrijven
    - (...)
- Beleid en wetgeving houden vaak veelomvattende, algemene hervormingen in: dit vergt gecoördineerde actie in gemeenten in meerdere gebieden, bijvoorbeeld gericht op service-verbetering of systeemhervormingen, waarbij verschillende doelgroepen moeten worden betrokken.
    - Combineer meerdere evaluatiedesigns: (1) kwantitatief-experimentele, om een indruk te krijgen van de mate van (a) doeltreffendheid (b) nationale 'dekkendheid' van beleidseffecten; met (2) representatief gekozen kwalitatieve *case study*- analyses, die het algemene beeld aanvullen door in te zoomen op 'hoe' de beoogde hervormingen zijn opgepikt en via welke processen het beleid een rol heeft gespeeld in de doelbereiking.
- Sommige programma's raken in de tenuitvoerlegging volledig 'vermengd' met de context (bijv. lokaal beleid binnen raamwetgeving). Hoe die context dan te scheiden van de effectmeting?
    - Verricht exploratieve case studies (bijv. representatief verdeeld over het land), gebaseerd op diepte-interviews en observaties, om een beeld te krijgen van wat per context werd bereikt, en hoe.
- Er lopen verschillende beleidsprogramma's tegelijk, gericht op dezelfde uitkomst. Dit vertroebelt de bijdrage van de afzonderlijke programma's
    - Stel een logisch model op voor elk afzonderlijk programma
    - Specificeer daarna (a) de scope van de uitkomstmaat per programma *zo smal mogelijk*. Het vinden van resultaten op die smalle uitkomstmaat valt met meer zekerheid toe te schrijven aan het betreffende programma dan algemene uitkomsten; (b) bezie of het aannemelijk is dat de programma's elkaar onderling beïnvloeden, en toets de resulterende hypothesen gericht (door variatie in uitkomsten aangevuld met 'waarom' surveys – bijvoorbeeld, waarom sommige groepen minder gemakkelijk aan gezondheidszorg komen.

Ook uit de voorgestelde oplossingen blijkt dat evaluatiedesigns in de toepassingspraktijk vaak een mix zijn van kwantitatief-experimentele en kwalitatieve aanpakken.

2.5 Borging van evaluatiekennis en – ervaring

Communicatie van het evaluatiekader
Al dan niet in aansluiting op het aangehaalde hulpmateriaal worden wereldwijd vele cursussen over evaluatie gegeven – die we hier niet allemaal kunnen noemen.[68]

Selectie van onderzoekers: competenties
Kennis over evaluatieve vaardigheden van onderzoekers is voor aanvragers relevant, omdat zij – via aanbestedingsprocedures of anderszins – proberen te achterhalen hoe competenties in een team van onderzoekers verdeeld zijn en hoe die zich verhouden tot de uitdagingen die met de evaluatie worden verwacht (vgl. UKES).
De EC (2013a, p. 71) benadrukt het belang van goede contacten met evaluatieonderzoekers voor de ontwikkeling van evaluatievermogen en hoogwaardige evaluatiestudies. De term *professionalisering* van evaluatie en van

[68] Zie alleen al bij de VN bijv. het overzicht op de ILO website: www.ilo.org . Voorts biedt het Britse EPPI een workshop aan over programma-evaluatie, mede ontwikkeld voor aanvragers. Bijvoorbeeld hoe goede uitvoerders van een evaluatie te selecteren en hoe een evaluatievoorstel op te stellen: http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=704.

evaluatieonderzoek valt de laatste jaren ook vaker, bijvoorbeeld in het Verenigd Koninkrijk (UKES) en bij de AEA.
De Europese EES-website bevat een overzicht van competenties van evaluatieonderzoekers (Bijlage 3). De AEA (2004, 2013) duidt de verantwoordelijkheden van evaluatie-onderzoekers op de volgende vijf terreinen aan: systematisch onderzoek, competenties, eerlijkheid en integriteit, respect voor mensen en verantwoordelijkheid voor het algemene en publieke belang.
Een verschil – volgens sommigen – tussen de VS en Europa is dat de functie van evaluatie-onderzoeker in de VS duidelijker omlijnd is; in Europa betreft het vaker algemeen onderzoekers, wiens evaluatievaardigheden niet op hun visitekaartje staan (comm. 8 december 2015).
Ook de evaluatieliteratuur biedt uitwerkingen van benodigde vaardigheden voor evaluatieonderzoekers (o.a. Cooksy, 2012; Stevahn et al., 2005; King et al., 2001). Zo onderscheiden King et al. (2001) zes competenties: professionaliteit, systematische onderzoeksvaardigheden, gevoeligheid voor de beleidscontext en vaardigheden om beleid te reconstrueren, projectmanagement, reflexiviteit/ persoonlijke ontwikkeling en interpersoonlijke vaardigheden. Deze gelden feitelijk ook (beleids)onderzoekers in het algemeen. Mark (2012) benadrukt daarom dat juist evaluatieonderzoekers over aanvullende competenties dienen te beschikken, zoals (1) het kunnen beoordelen en tevens verdedigen wat passende evaluatiedoelen in een bepaalde situatie zijn (...) (2) aanpassingsvermogen in geval van veranderende omstandigheden, zoals budgetwijzigingen en (3) het door hebben van het spel waarin men zich bevindt; bijvoorbeeld of een evaluatie voor strategische doeleinden wordt ingezet. '*The evaluator can exercise adaptability and reformulate the emphasis of the evaluation in ways that might bring positive consequences nevertheless'* (Mark, 2012, p. 85).

Begeleidingscommissie of stuurgroep en samenstelling
Vooral met het oog op de benutting en terugkoppeling van evaluatiebevindingen naar het beleid werkt de EC (2013a, p. 46-7; 2015c) uit, waarom het zinvol is een *Steering Committee* (SC), een adviserende stuurgroep of begeleidingscommissie, in te schakelen en hoe die kan worden samengesteld. Voor de samenstelling van de SC worden vier categorieën leden genoemd: strategisch en operationeel management, sociale partners (andere stakeholders) en experts, met inhoudelijke dan wel methodologische kennis. Benadrukt wordt dat het de taak van de voorzitter van de SC is om ervoor te zorgen dat de leden geen beïnvloeding toepassen of bepaalde evidentie negeren. Andere taken zijn: het waarborgen van kwaliteit en bruikbaarheid door faciliteren van de evaluatie, bijvoorbeeld via toegang tot informatiebronnen en contacten en het uitwerken van evaluatievragen en speerpunten. De SC kan ook een rol vervullen in de communicatiestrategie (zie verder 2.6).

2.6 Gebruik

Zoals aangegeven lijkt evaluatie internationaal steeds meer onderdeel te worden van verschillende fasen van de beleidscyclus; het wordt met de tijd minder beschouwd als een losse activiteit die uitmondt in een rapport. Deze visie op evaluatie werkt ook door in het *gebruik* ervan, dat vaker beschouwd wordt als een proces dat zich niet alleen over het evaluatie- maar ook over het beleidstraject uitstrekt.
Evaluatiedeskundige Patton (2010, p.7) stelt in een Wereldbankpublicatie:
'*[E]verything that will be done from beginning to end will affect use. (....) We have learned that use is a process, not an event, and that it needs to be facilitated. It involves an interaction, not just a report, to interpret findings and apply them. It involves training for use, not just the delivery of results. The intended users have to have some help in knowing what to*

do with findings. It is not apparent or natural to go from data to action and decision making, and use will mean different things for different evaluation purposes.'
De door Patton geschetste interactie vergt wel een adequate timing van evaluaties in de beleidscyclus. Het tijdig beschikbaar komen van evaluatieresultaten is in de praktijk niet evident. Het Witte Huis (2014, p. 277-79) geeft aan dat het soms lange tijdsverloop tussen de implementatie en het uitkristalliseren van eventuele effecten een belemmerende factor kan zijn, vooral voor het gebruik van 'rigorous' evaluaties. Men doet verslag van een praktijk, de *Rapid Cycle*- benadering, waarbij via vroegtijdige terugkoppeling van data – eerst alleen over proces- maar daarna ook resultaatindicatoren – de aanvrager zich zo snel mogelijk een beeld kan vormen van de werking van nieuwe interventies. Die resultaten worden vergeleken met die van alternatieve (bijv. de oude) aanpakken.
Jacob et al. (2015) concluderen dat systematisch evaluatiegebruik in de 19 door hen onderzochte OESO-landen lastig te bewerkstelligen is. Wel noemen ze de Britse overheid als voorbeeld, waar het integreren van onderzoekers in de beleidsdirecties en strategische units van departementen een succesvolle manier is gebleken om evaluatiebevindingen in het doorlopende beleidsproces – van conceptie tot en met implementatie – te verwerken. Wel blijkt uit OESO-onderzoek (2015b) dat in nationale wet- en regelgeving steeds vaker gebruik wordt gemaakt van horizonbepalingen.

Voor onderzoekers
Ondanks de sterkere nadruk op de verantwoordelijkheid van de aanvrager is de evaluatie-onderzoeker nog steeds een belangrijke actor in de bevordering van gebruik. De UNEG (2005a) hanteert als norm dat evaluaties bijdragen aan structurele kennisopbouw en organisatieverbetering. Daartoe dienen ze in voor de gebruiker begrijpelijke taal te zijn opgesteld.
Voor de sociaaleconomische EU-evaluaties vereist de EC (2013a, p. 48) niet alleen dat analyses en gevolgtrekkingen inhoudelijk gedegen en verantwoord zijn, maar ook dat rapporten helder en bondig zijn – rond de 100 pagina's inclusief managementsamenvatting. Diepgaande case studies en kwantitatieve analyses horen in bijlagen en het gebruik van samenvattende tabellen en figuren wordt aangemoedigd. Aanbevelingen dienen duidelijk te zijn over de vereiste vervolgacties.

Follow-up en kennisverspreiding
De UNEG-normen (2005a) vergen dat (a) het management van de VN-organisatie die het betreft, expliciet reageert op aanbevelingen uit de evaluatie (b) dat in geval van resulterende actieplannen duidelijk verantwoordelijkheden worden weergegeven) (c) dat er systematische opvolging is in de implementatie van aanbevelingen (...) en (d) dat er een periodieke rapportage komt met betrekking tot de status van die implementatie aan het management.
Soms worden programma's afgeschaft als gevolg van negatieve evaluaties. Het Witte Huis (2014, p. 282 e.v.) beschrijft een voorbeeld hiervan, voortvloeiend uit een experimenteel evaluatiedesign.
De UNEG hanteert voorts een bepaling die aanstuurt op het aanleggen van een repositoir van evaluaties vergezeld van een helder verspreidingsbeleid, ter bevordering van verdere kennisvergaring en leren (UNEG, 2005a, p.11).
Ook in het *Smart Regulation* beleid van de EC dienen de lessen uit evaluaties beschikbaar te zijn en terug te vloeien naar het beginstadium van de beleidscyclus. Dit zou moeten gebeuren via het langer bestaande systeem van ex ante *Regulatory Impact Assessments* (RIA) (EC, 2015b,c; Smismans, 2015).

Parlementen en gebruik

Parlementen vervullen per land een verschillende rol ten aanzien van evaluaties, maar volgens Jacob et al. (2015) besteden maar weinig van 19 onderzochte OESO-parlementen aandacht aan de kwaliteit van door hen uitgezette evaluaties. In Ierland of Spanje vinden discussies naar aanleiding van evaluaties zelden plaats, in Noorwegen en Engeland vaker. In Noorwegen worden evaluatieresultaten in de vorm van *White Papers* in het parlement besproken. Het Zwitserse parlement heeft als enige sinds 1991 een evaluatie-unit. Dit wordt toegeschreven aan het feit dat de Zwitserse parlementariërs geen professionele politici zijn en in vergelijking met andere landen relatief veel ondersteuning nodig hebben.
In Nederland is het gebruik van evaluatieresultaten door parlementsleden het afgelopen decennium sterk toegenomen, vooral op het punt van ontwikkelingssamenwerking (Jacob et al., 2015). In ons land bestaat sinds 2007 de BOR: het Bureau Onderzoek en Rijksuitgaven. Dat levert een bijdrage aan de inhoudelijke ondersteuning van de Tweede Kamer op het terrein van onderzoek en rijksuitgaven – ter versterking van de controlerende taak van de Kamer.

Overige initiatieven ter ondersteuning van gebruik
In sommige landen wordt het gebruik van ex post evaluaties in de beleidscyclus ondersteund met initiatieven voor *evidence-based policy* en meta-analyses. Voorbeelden zijn het Amerikaanse *What Works Clearinghouse* voor wat werkt in het onderwijs, de (Nordic) *Cochrane Collaboration* en de *SFI-Campbell Collaboration*, de Scandinavische tak van Campbell, specifiek gericht op het systematisch onderzoeken en verzamelen van wat werkt in termen van sociale zekerheidsinterventies en –voorzieningen.[69] In het Verenigd Koninkrijk bestaat het EPPI-Centre, de *Evidence for Policy and Practice Information and Co-ordinating Centre*.[70] EPPI is onderdeel van een universitair onderzoekscentrum en gericht op het informeren van beleid en professionele praktijk met wetenschappelijke kennis. Dat uit zich in twee werkwijzen: (1) *systematic reviews* en trainingen daarin (Gough et al., 2012) en (2) het bestuderen van gebruik en trainingen daarin, onder meer ten behoeve van de ambtelijke en politieke besluitvorming. De site bevat lijsten van *systematic reviews* waarin kennis uit (evaluatie)onderzoek op een specifiek beleidsterrein of instrument bijeen gebracht is. In veel landen echter ontbreken dergelijke aanvullende initiatieven (Jacob et al., 2015).

2.7 Werkt evaluatievermogen?

Er is (zeer) veel literatuur over evaluatievermogen, maar de vraag: leidt evaluatievermogen tot betere beleidsuitkomsten? is vooralsnog in de internationale literatuur niet of nauwelijks beantwoord, zo stellen Suarez-Balcazar & Taylor-Ritzler (2013, p. 97):

> *'Most of the empirical research involves descriptive case study designs and little use of reliable and valid instruments. As a result, although ECB holds much promise in improving program outcomes for program participants and accountability for organizations, there is, as yet, little evidence that ECB practices actually help organizations systematically improve services that lead to positive outcomes for their program participants.'*

Ook Jacob et al. (2015) concluderen op basis van hun vergelijking van OESO-landen dat een positieve invloed van evaluatievermogen niet aan te tonen is, omdat systemen van verschillende landen te veel van elkaar verschillen. Het valt dus niet in zijn algemeenheid te stellen dat ECB 'werkt'. Uit praktijk komt volgens eerder aangehaalde auteurs wel naar voren dat evaluatievermogen een kwestie van een

[69] www.sfi.dk/sfi_campbell-432.aspx.
[70] http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=53.

lange adem is. Er zijn geen snelle oplossingen en het gaat naast om technische vaardigheden en toepassing van administratieve instrumenten en richtlijnen óók om het kweken van een evaluatiecultuur. Vertrouwen tussen aanvrager en onderzoeker is daarbij van belang.

Neveneffecten
Eerder is in relatie tot (ex ante) *Regulatory Impact Assessments* ook al gesignaleerd dat in de EU en een aantal van haar lidstaten in aanzienlijke mate sprake is van *justificatory use* (Hertin et al., 2007). Men betwijfelt in hoeverre evaluatie een integraal onderdeel uitmaakt van de beleidscyclus, zoals in veel professionaliseringshandboeken als streven wordt benoemd.[71]
In 2005 publiceerden Schwartz & Mayne een internationaal overzicht van praktijken van jurisdicties om de kwaliteit van beleidsevaluaties (en die van audit-onderzoek) te monitoren en te verbeteren. Ze bestudeerden daartoe niet-representatief vier landen, Nederland, Frankrijk, Canada en Zwitserland, en twee internationale organisaties (de Wereldbank en de EU).
In navolging van Power (1997) onderscheiden de auteurs twee perverse effecten van kwaliteitsmonitoring, te weten ontkoppeling en kolonisatie. *Ontkoppeling* houdt in dat kwaliteitsverhogende activiteiten losgekoppeld blijven of worden van de intrinsieke activiteiten. Er is dan sprake van ritualistische naleving van kwaliteitscriteria, zonder dat dit impact heeft op de dagelijkse praktijk. Bij *kolonisatie*, daarentegen, raken de kwaliteitsstandaarden juist ingebakken in de werkzaamheden. Hierdoor kan een soort van uitholling ontstaan, die zich uit in problemen zoals tunnelvisie, orthodox werken en gebrek aan innovatie. De organisatie zet alles in op voldoen aan de standaarden, ten koste van niet gemonitorde of niet meetbare kwesties.
Een ander ongewenst neveneffect van evaluatievermogen – als dit de vorm aanneemt van resultaatgeoriënteerde M&E systemen – is dat ze beleidsontwikkelaars kunnen prikkelen om *minder* ambitieuze doelen te formuleren (Clements et al. 2008). Dit kan ontstaan als voorbij wordt gegaan aan de intrinsieke (maatschappelijke) waarde van het bereikte: *'a program that reaches all its timid targets may be less cost effective than one that fails to reach ambitious goals'.*

2.8 Samenvatting

In dit hoofdstuk zijn internationale ervaringen met bouwen aan evaluatievermogen bij een aantal grotere internationale instituties en westerse landen op een rij gezet. We bespraken ook enkele internationale studies (Jacob et al. en OESO) die in helicopterview belangrijke kenmerken van (wets)evaluatievermogen bespreken, gevolgd door verschillen en overeenkomsten tussen landen daarin.

Evaluatie als onderdeel van het beleidsproces: de aanvrager als sleutelfiguur
Meer en meer dringt de notie door dat de kwaliteit en het gebruik van evaluatieonderzoek niet alleen afhankelijk zijn van de onderzoekers, maar juist ook van de *aanvragers* ervan – de beleidsmedewerkers. Zij nemen voor evaluaties bepalende beslissingen, bijvoorbeeld over het budget, de reikwijdte en de onderzoeksvragen (in de opdracht) en ten aanzien van de verdere begeleiding van het evaluatieproces.
Niet alleen als een evaluatie aanstaande is, maar ook daaraan voorafgaand worden randvoorwaarden voor evaluaties geschapen. De inhoud en opzet van het te

[71] Anderzijds zijn evaluatiebevindingen slechts een uit een reeks factoren die meewegen in de ontwikkeling van beleid. Dit is zeker zo in landen/culturen waar draagvlak en stakeholderbetrokkenheid veel gewicht hebben – zoals Nederland (vgl. Leeuw, 2009).

evalueren beleid en de onderbouwing daarvan ziet men steeds vaker als een factor die medebepalend is voor evaluatiekwaliteit. Ontbreekt bijvoorbeeld de argumentatie voor beleidskeuzes (beleidstheorie) dan valt het lastiger de evaluatie te focussen. En die zou daardoor ook wel eens duur kunnen uitvallen.
Ook in de monitoring van beleid heeft de evaluatie-aanvrager een cruciale rol te vervullen. Meerdere instituties vinden het een gemiste kans voor hoogwaardige evaluaties als niet tijdig wordt begonnen met nulmetingen en de verdere monitoring van belangrijke proces- en uitkomstindicatoren. Ook doelgroepen, zoals gemeenten, burgers en bedrijven kunnen in de monitoring een rol vervullen.
Al met al wordt evalueren steeds vaker beschouwd als onderdeel van de beleidscyclus, en minder als een activiteit die daar los van staat.

De aanpak: stromingen en fit-for-purpose
In handreikingen voor de evaluatieaanpak zijn verschillende stromingen te onderscheiden; illustratief is het onderscheid tussen de 'attributieve' (*black box*) en de 'contributieve' (*clear box*) stroming. In de eerste staat het kunnen toeschrijven van resultaten aan beleid centraal: heeft het gewerkt en in welke mate? De tweede stroming is gericht op het begrijpen van hoe en waarom (onder welke condities) veranderingen zijn opgetreden en resultaten zijn geboekt.
De stromingen sluiten elkaar overigens niet uit, ze kunnen wederzijds aanvullend worden gebruikt. Het uitgangspunt dat steeds vaker de stromingen overstijgt is '*fit for purpose'*; keuzes in het evaluatiedesign worden afhankelijk gesteld van de situatie, o.a. soort beleid, beleidsfase, evaluatieonderwerp, hoofdvragen en soort doelgroep. In dit licht ontwikkelde onder meer de Amerikaanse federale Rekenkamer een aantal slimme handgrepen voor de evaluatie van complexe federale programma's. Net als (Nederlandse) wetten zijn ook deze meestal geen 'cleane' interventies waarvan de werking eenvoudig gemeten en bovendien geïsoleerd kan worden van de context.
Als het te laat of onmogelijk is om nog een controlegroep te vormen, suggereert men bijvoorbeeld om (1) kwantitatieve tijdreeksanalyses te doen, om te toetsen of het beleid tot 'trendbreuken' heeft geleid (met controle voor andere mogelijke verklaringen), en/of (2) variatie in kaart te brengen tussen regio's of gemeenten, met case studies en eventueel kwantitatief, om te ontrafelen *hoe* beleid in de praktijk is uitgepakt en waarom het zo is gegaan.

Ook evaluatiegebruik is een proces
De visie op evaluatie als integraal onderdeel van het beleidsproces werkt door in die op het *gebruik* ervan; grote instituties en landen beschouwen gebruik steeds meer als het resultaat van het hele evaluatieproces en niet van alleen een eindrapport. Dit alles neemt niet weg dat de evaluatie-onderzoeker nog steeds als een belangrijke schakel wordt gezien in de bevordering van gebruik. Dit blijkt bijvoorbeeld uit leidraden voor onderzoekers om bevindingen toegankelijk en bruikbaar te presenteren. Verder blijkt uit competentiebeschrijvingen dat niet alleen methodologisch-inhoudelijke vaardigheden van onderzoekers belangrijk worden gevonden maar ook hun interpersoonlijke en communicatieve vaardigheden, flexibiliteit en een sterke gevoeligheid voor de (politieke en beleids-)context.

Werkt evaluatievermogen?
Hoewel de literatuur over evaluatievermogen omvangrijk is, is vooralsnog niet wetenschappelijk bewezen dat ECB tot betere beleidsresultaten leidt. Wel zijn uiteenlopende (positieve) case studies opgeleverd. Tegelijkertijd is gewaarschuwd voor neveneffecten, zoals dat het aanleggen en handhaven van standaarden opportunistisch gedrag of ritualisering in de hand kan werken waarbij het eigenlijke doel uit het oog wordt verloren.

Hoofdstuk 3 Praktijkervaringen met evaluatievermogen bij beleidsdepartementen

3.1 Evaluatiebeleid en wetgevingskwaliteitsbeleid

3.1.1 Verantwoording binnen het rijksbrede evaluatiestelstel

In het centrale verantwoordingsstelsel dient evaluatieonderzoek bij te dragen aan een antwoord op de vraag, of en in hoeverre het beleid of de wet doelmatig en doeltreffend is geweest en welke gewenste en ongewenste effecten teweeg zijn gebracht.
Op basis van een chronologische reconstructie constateert Leeuw (2009) dat centrale evaluatie-instituties en het centrale evaluatiebeleid – primair gericht op verantwoording van begrotingsuitgaven – zich in Nederland over de afgelopen decennia aanzienlijk hebben ontwikkeld. Voor een op basis van meerdere bronnen samengestelde tijdlijn zij verwezen naar bijlage 4.

Beleidsdoorlichtingen en evaluatieonderzoek
Met het in de Regeling Periodiek Evaluatieonderzoek (RPE) van 2006 geïntroduceerde instrument van de beleidsdoorlichting wordt elk departement geacht om bevindingen uit verschillende doeltreffendheids- en andere evaluaties hierin samen te brengen (synthetiseren) op het niveau van de beleidsdoelen (artikelen) in de begroting.[72]
Departementen verantwoorden dus hun uitgaven aan de volksvertegenwoordiging op basis van de beschikbare informatie in de beleidsdoorlichting. De beleidsdoorlichting werkt op deze manier door in de ambtelijke evaluatiepraktijken waar beleidsevaluaties plaatsvinden. Idealiter stapelt zich hier in de loop van (maximaal) 7 jaar het basismateriaal op voor synthese tot de uiteindelijke beleidsdoorlichting.

> ‘Het is (...) noodzakelijk ervoor te zorgen dat de evaluatieprogrammering aansluit bij de programmering van de beleidsdoorlichtingen. Idealiter is op het moment dat een beleidsdoorlichting gepland staat voldoende informatie beschikbaar (in de vorm van evaluatieonderzoek), zodat een goede beoordeling van de doeltreffendheid en doelmatigheid van het gehele beleidsartikel mogelijk is’ (Financiën, 2014).

Volgens artikel 3 van de recentere RPE uit 2014 hoort een beleidsdoorlichting informatie over de volgende aspecten te bevatten: afbakening van de onderzochte beleidsterreinen; motivering voor het beleid en de beoogde doelen die de overheid ermee wil bereiken; beschrijving van het beleidsterrein en de financiële middelen die zijn ingezet; overzicht van het al bestaand onderzoek naar doeltreffendheid en doelmatigheid en de onderbouwing van de evaluatieprogrammering; de effecten van het gevoerde beleid, de doeltreffendheid en doelmatigheid hiervan en de effecten van het beleid op economische groei en regeldruk.[73]
In de eerste (tussen)evaluatie van het instrument beleidsdoorlichting – alweer een aantal jaren geleden – constateren externe onderzoekers dat ‘op een aantal ministeries het instrument beleidsdoorlichting goed blijkt te zijn opgepakt. Het gaat

[72] De rapporten zijn openbaar op de website van de Rijksbegroting; www.rijksbegroting.nl/beleidsevaluaties/evaluaties-en-beleidsdoorlichtingen/2015.
[73] In de beleidsdoorlichting horen ook suggesties voor de verhoging van de doeltreffendheid en doelmatigheid van het gevoerde beleid. Bovendien dienen de opstellers budgettair onderbouwde beleidsopties in kaart te brengen – en de consequenties daarvan af te wegen – voor als zich de hypothetische situatie van een daling van middelen met 20 procent voordoet.

dan veelal om ministeries waar een cultuur bestaat waarin betrokkenen alert zijn op de doeltreffendheid van het ontwikkelde beleid en open staan voor aanpassingen om dat beleid verder te verbeteren. In andere ministeries is die cultuur minder duidelijk aanwezig, met als gevolg dat beleidsdoorlichtingen daar veel meer als een verplicht nummer gezien worden, als een extra taak die niets oplevert' (Von Meyenfeldt et al., 2008, p. 74).

Methodologische standaarden
Wat betreft methodologische standaarden voor evaluatie constateert Leeuw (2009) dat die in het Nederlandse verantwoordingssysteem niet concreet omlijnd zijn; de invulling wordt goeddeels aan het evaluatie-en onderzoeksveld overgelaten.[74]
Er zijn in de RPE wel kwaliteitseisen opgesteld voor de overkoepelende beleidsdoorlichtingen (Financiën, 2014):

a) het onderzoek maakt duidelijk welk beleid wordt onderzocht en wat de doelstellingen van dat beleid zijn;
b) het onderzoek beoogt de vraag te beantwoorden in hoeverre het beleid, alsmede de daarmee samenhangende uitgaven, doeltreffend of doelmatig is;
c) de conclusies van het onderzoek worden onderbouwd door onderliggende bevindingen;
d) de in het onderzoek gebruikte onderzoeksmethode is valide en betrouwbaar, het rapport geeft inzicht in de gebruikte evaluatiemethode en in de mogelijkheden en onmogelijkheden om de doeltreffendheid en/of de doelmatigheid van het betreffende beleid vast te stellen.

'Handhaving' van verantwoordingsplichten
De begrotingsdoelen die steeds het richtpunt zijn van de beleidsdoorlichtingen, worden vaak (te) breed gevonden (o.a. Leeuw, 2009) en open voor interpretatie. Niet-naleving van de evaluatieverplichtingen in het centrale stelsel heeft geen grote formele gevolgen. Trekt de Algemene Rekenkamer bijvoorbeeld stevige conclusies op basis van de hoeveelheid en kwaliteit van verantwoordingsevaluaties binnen departementen – zoals gebeurde in 2012 en 2013 – dan leidt dit los van eventuele reputatieschade formeel niet tot consequenties, buiten de kans op een opvolgingsmonitor door de AR (interview 2 september). De 'openheid' van het stelsel kan positief of negatief uitgelegd worden. Zo kan teveel nadruk op het voldoen aan evaluatieverplichtingen perverse effecten veroorzaken, zoals uitholling van inhoud en evaluatiekwaliteit (vlg. Van Thiel en Leeuw, 2002 en par. 2.7 van dit rapport). Deze negatieve effecten kunnen eventuele positieve effecten van verplichtingen al dan niet overstemmen.

3.1.2 Wetgevingskwaliteitsbeleid en wetsevaluaties

De zorg voor de kwaliteit van wetgeving is sinds de jaren tachtig expliciet onderdeel van het regeringsbeleid. De wortels van het huidige wetgevingskwaliteitsbeleid liggen in de nota *Zicht op wetgeving* uit 1991.[75] Deze gaf de visie van de overheid inzake kwaliteit van wetgeving weer. De nota schetst de kwaliteitseisen die ten grondslag liggen aan het huidige wetgevingskwaliteitsbeleid: rechtmatigheid, doeltreffendheid en doelmatigheid, subsidiariteit en evenredigheid, uitvoerbaarheid en handhaafbaarheid, onderlinge afstemming, eenvoud, duidelijkheid en toegankelijkheid.

---

[74] Leeuw schrijft dit toe aan de sterke rol van inspraak van belangengroepen in beleid: sociale acceptatie van instrumenten en standaarden gaat vóór methodologische kwaliteit (2009, p. 95).
[75] TK 1991-1992, 22008 nr. 2.

Evaluatiebepaling en wetsevaluatiebeleid
In 1993 is in de Ar een evaluatieartikel (art. 164) opgenomen in de vorm van een standaardformulering voor een in de wet op te nemen evaluatiebepaling. In een evaluatiebepaling wordt opgenomen dat – indien wenselijk – een wet éénmalig of periodiek wordt geëvalueerd en dat verslag wordt gedaan van de *doeltreffendheid en de effecten* van die wet, of van aspecten of onderdelen ervan. Volgens deze Aanwijzing ligt een evaluatietermijn van vijf jaar in de rede. De toelichting vervolgt dat voor evaluatie in aanmerking komen de verwezenlijking van gestelde doelen en neveneffecten en daarnaast ook *voornoemde rits van kwaliteitseisen* in de nota Zicht op Wetgeving. Welke evaluatievorm wordt gekozen, is volgens de Ar afhankelijk van 'onder meer het gewicht van de betrokken regeling, de maatschappelijke betekenis ervan en de aan evaluatie verbonden lasten.' De toelichting stelt dat 'er een breed scala denkbaar is, van diepgaand wetenschappelijk onderzoek tot rapportage door uitvoeringsorganen'.
In 2000 verscheen de nota Wetgevingskwaliteitsbeleid en wetgevingsvisitatie - onder meer naar aanleiding van Europese kwaliteitseisen en ICT-ontwikkelingen (TK 2000-2001 27 475 nr. 2, p. 4). Wetten moeten vanwege deze ontwikkelingen regelmatig worden herzien. Daarnaast benadrukt het kabinet in de nota dat wetgeving slechts één van de instrumenten is waarover de politiek en het ambtelijke apparaat beschikken om veranderingen in de samenleving te bewerkstelligen. Het kabinet benadrukt voorts het belang van samenwerken en het *benutten van de kennis* die vanuit alle niveaus wordt aangereikt om de kwaliteit van de wetgeving te borgen.[76]
In de jaren nul van deze eeuw verschoof het accent van een integrale blik op kwaliteit naar het belang van *uitvoerbaarheid* (ibid, p. 5), met name dat nieuwe regelgeving geen grote *lasten* voor de uitvoering en handhaving mag veroorzaken. De organisatie van de wetgevingsfunctie per departement kan in theorie de inhoud en kwaliteit van wetsevaluaties mede bepalen (onafhankelijke Visitatiecommissie Wetgeving 2000). Dit zou kunnen leiden tot verschillen in de kwaliteit van wetgeving tussen beleidsdomeinen.
Een van de onderdelen van het tien jaar later geïntroduceerde Programma Versterking Juridische Functie Rijk (2010)[77] is het beter laten samenwerken van juristen en niet-juristen - zoals probleemoplossend vermogen integreren met gedegen juridische analyse en het kweken van wederzijds begrip. Dit zou ook van toepassing kunnen zijn op de totstandkoming van wetsevaluaties.

Wetsevaluatiebeleid
De uitgangspunten voor het huidige wetsevaluatiebeleid zitten besloten in de al genoemde oudere nota Wetgevingskwaliteitsbeleid en wetgevingsvisitatie (2000): destijds had, buiten de genoemde evaluatiebepaling om een algemeen wetsevaluatiebeleid geen prioriteit en werd niet nodig geacht. Inmiddels is de aandacht voor wetsevaluaties weer toegenomen, zoals ook beschreven in de inleiding van dit rapport.
In de afgelopen 15-20 jaar zijn 'aan de voorkant' (ex ante) steeds meer aanwijzingen, checklists en toetsen geïntroduceerd om wetgevingsjuristen te helpen bepaalde afwegingen te maken bij het opstellen van wetgeving, variërend van administratieve lasten- tot uitvoerings- en handhavingstoetsen. Veel van die eerdere instrumenten zijn samen- en teruggebracht in het Integraal Afwegingskader

[76] Zo dienen wetgevingsjuristen bij het opstellen van nieuwe wetgeving rekening te houden met *alle betrokken actoren* die een rol spelen in het wetgevingsproces: internationale actoren, actoren bij de centrale overheid, decentrale overheden zoals provincies, gemeenten, lokale rekenkamers, uitvoerders, handhavers en wetenschappers.
[77] Dit programma is onderdeel van het programma versterking Rijksdienst en vloeit voort uit het rapport 'Met recht verbonden' (Commissie Hoekstra, 2007) – om het hoofd te kunnen bieden aan nieuwe ontwikkelingen, over de grenzen van beleidsvelden heen.

voor beleid en regelgeving (IAK), een overkoepelend instrument met aandachtspunten en criteria voor kwaliteitsborging. Het IAK dient te worden toegepast bij de voorbereiding van zowel beleid als wet- en regelgeving. Het IAK is de Nederlandse invulling van het Europese *Impact Assessment* dat ook veel andere lidstaten kennen. Het omvat zeven vragen[78] en achttien kwaliteitseisen, die sinds 2011 verplicht door beleidsmakers en wetgevingsjuristen moeten worden toegepast bij de voorbereiding van nieuw beleid en regelgeving. Ieder nieuw voorstel van beleid en regelgeving moet worden begeleid door een toelichting naar aanleiding hiervan.[79]

Verdere opbouw van dit hoofdstuk
Dit hoofdstuk volgt verder zo gericht mogelijk de indeling van het conceptueel model uit hoofdstuk 1.

3.2 Evaluatiestructuur

3.2.1 Doelen van evaluatieonderzoek

Evaluaties en wetsevaluaties worden verricht vanuit verschillende motieven – die elkaar overigens niet altijd uitsluiten. Deze motieven kunnen variëren van kennisvergroting tot politiek-strategische doelen tot verantwoording van doeltreffendheid en doelmatigheid. Evaluatiedoelen zijn dus sterk verbonden met de politieke context en eventuele andere contextuele factoren.
Buiten de doeltreffendheids- en doelmatigheidsdoelstellingen uit de Cw en de RPE om, bestaan maar weinig departementale documenten waarin doelen van evaluaties of *wets*evaluaties formeel worden vastgesteld. Bij EZ is één van de doelen sinds enkele jaren bijvoorbeeld het verhogen van de kwaliteit van doeltreffendheidsstudies (via een nieuwe organisatiestructuur, zie 3.2.2) en daarmee uiteindelijk het bewerkstelligen van effectiever beleid.
Het programma Evaluatie Regelgeving (PER) van de gelijknamige Commissie (CER), specifiek gericht op gezondheidswet- en regelgeving, biedt 'een helder en doorlopend kader waarbinnen wetsevaluaties kunnen plaatsvinden. (...) Het streeft uniformiteit na bij uit te voeren wetsevaluaties.' (ZonMw, 2013, p.3-4).

3.2.2 Organisatieonderdelen met evaluatietaken

Departementen, en daarin op hun beurt DG's en directies, verschillen in de mate waarin structureel aandacht wordt besteed aan het tot stand brengen van evaluatieonderzoek.
Globaal is de evaluatiefunctie bij de meeste departementen als volgt georganiseerd: het initiatief voor (ex post) wets- en beleidsevaluaties ligt bij de beleidsdirecties, die hiervoor de verantwoordelijkheid dragen. Binnen de betreffende beleidsdirectie fungeert een van de beleidsmedewerkers als *dossierhouder* betrokken bij de evaluatieaanvraag, het evaluatieproces en de oplevering van het rapport (interviews 15 juni; 15 juli; 26 augustus; 2 september). Doorgaans is er op DG- en soms directieniveau een *onderzoeks- of kenniscoördinator*, die dossierhouders inhoudelijk en/of procesmatig ondersteunt.
De meeste departementen hebben geen aparte evaluatieafdeling en al helemaal geen afdeling die zich uitsluitend op *wets*evaluatie richt. Wel zijn er stafafdelingen,

[78] (1) Wat is de aanleiding (2) wie zijn betrokken (3) wat is het probleem (4) wat is het doel (5) wat rechtvaardigt overheidsinterventie (6) wat is het beste instrument (7) wat zijn de gevolgen? https://www.kcwj.nl/kennisbank/integraal-afwegingskader-beleid-en-regelgeving?ip_login_no_cache=0825a1cc6c28626dcc80cf0a9a50b389
[79] Het inhoudelijke verslag van toepassing van het IAK hoeft niet te worden gepubliceerd.

commissies of directies die fulltime of parttime een evaluatie-ondersteunende functie vervullen. Deze bespreken we in wat volgt.

Directies Financieel-economische zaken
De Directie Financiële en Economische Zaken (FEZ) hebben departementen gemeenschappelijk als het gaat om evaluaties. Dit is geen toeval: sinds de jaren '90 van de vorige eeuw zijn Directies FEZ verantwoordelijk voor het stimuleren en coördineren van de evaluatiefunctie (en voor het nagaan van kosten en opbrengsten van evaluaties). De FEZ-directie houdt zich primair bezig met verantwoordingsdoelen en –processen.
Desalniettemin verschilt de rol die FEZ vervult in de evaluatiefunctie tussen de bestudeerde departementen. Zo heeft de directie bij zowel EZ als SZW een aanjagersrol inzake de planning en kwaliteitsborging van het evaluatieonderzoek in brede zin (interview 15 juli). Anders is het bij bijvoorbeeld VenJ, VWS en OCW, waar de rol van FEZ dichter bij de formele control-taken blijft in relatie tot het centrale evaluatiestelsel (interview 1 juli, 2 en 15 september). Het instrument van de beleidsdoorlichting is daarbij leidend.
Bij OCW hebben medewerkers van de Auditdienst Rijk (ADR) in de paar jaar voorafgaand aan dit onderzoek de beleidsdoorlichtingen op zich genomen.[80]
FEZ-SZW, met name het cluster Evaluatie en Beleidsinformatie, waar zes personen werkzaam zijn, borgt de kwaliteit van de beleidsinformatie en –evaluaties en is verantwoordelijk voor het systematisch evalueren van SZW beleid/wetgeving (interview, 15 juli). Dit proces loopt net als bij andere departementen voor een belangrijk deel via de onderzoekscoördinatoren bij de beleidsdirecties. Voor afzonderlijke evaluaties en beleidsdoorlichtingen worden de inhoudelijk betrokken beleidsmedewerkers aangewezen.[81] Ook ziet FEZ toe op de programmering, kwaliteit en de benutting van beleidsevaluaties en -doorlichtingen (SZW, 2009). Zo maakt de directie een evaluatieagenda voor evaluaties en beleidsdoorlichtingen, adviseert ze al dan niet in een begeleidende rol over de keuze van onderzoeksmethoden en draagt ze zorg voor de procesmatige kant van evaluatieonderzoek en beleidsdoorlichtingen.
Ook bij EZ vervult FEZ een spilfunctie. Een aantal relatief jonge organisatieonderdelen richt zich in samenwerking met FEZ op de kwaliteitszorg voor beleid en evaluatie: de Beleidskwaliteit- en Evaluatie Commissie (BEC, sinds 2013), een tot de BEC-structuur behorende Regiegroep Monitoring en Effectmeting (sinds 2011)[82] binnen het DG Bedrijfsleven & Innovatie (DG B&I) en het *Behavioural Insights Team* (BIT, sinds 2014). De BEC houdt zich (onder meer) bezig met kwaliteitsborging, kennisdeling en kennisopbouw ter bevordering van de kwaliteit van effectevaluaties. Het secretariaat van de BEC is in handen van de Directie Algemene Economische Politiek (AEP) en FEZ – de directeur van laatstgenoemde directie is BEC-voorzitter. In de BEC zitten voorts portefeuillehouders van elk van de DG's (B&I, Energie, Telecom en Mededinging en Agro & Natuur, de sector beleidsinformatie van de Rijksdienst voor Ondernemend Nederland (RVO.nl) naast vier externe leden: van CPB, PBL, de VU en de speciaal door EZ aangestelde VU-hoogleraar beleidsevaluaties (BEC-secretariaat, p.32). De Nederlandse Voedsel- en Warenautoriteit (NVWA) (toezicht) en DWJZ zijn agendalid.

[80] Bijvoorbeeld in 2015 Cultuur, en 'Prestaties van leerlingen en studenten omhoog' en in 2014: Emancipatie en Mediabeleid. Zie www.rijksbegroting.nl/beleidsevaluaties/evaluaties-en-beleidsdoorlichtingen.
[81] Men heeft gebrainstormd over het centraliseren van de onderzoekscoördinatie in een afzonderlijke unit, maar uiteindelijk is de structuur gehandhaafd dat elke SZW-directie een onderzoekcoördinator heeft.
[82] In 2016 is de Regiegroep M&E opgegaan in het beleidsanalistenteam (BAT) dat dezelfde naast andere functies vervult.

De BEC moet zorg dragen voor versterking en verbetering van de beleidskwaliteit en evaluatie over het hele departement; hoewel de DG's hiervoor zelf de verantwoordelijkheid blijven dragen. 'BEC is een netwerkstructuur op basis waarvan je iedereen kunt bereiken, aan elkaar kunt blootstellen, de do's en don'ts kunt delen' (interview, 9 juli), met andere woorden: men beoogt te leren en verbindingen te zoeken, gebruik makend van de interne deskundigheid en daarnaast externe expertise (BEC-secteratiaat, z.j., p.2). De BEC is ontstaan na publicatie van het rapport *Durf te meten* van de expertgroep onder leiding van wijlen Jules Theeuwes, waarin meer aandacht wordt gevraagd voor de doeltreffendheid van beleidsinstrumenten en voor het gebruik van kwantitatieve onderzoeksmethoden om die vast te stellen (Expertgroep Effectmeting Theeuwes, 2012; interview 9 juli).[83]
Bij ex post evaluaties geldt: 'voordat je met een offertetraject begint en een onderzoeksbureau betrekt, moet je de probleemstellende [of start-] notitie – en het logisch model van je interventie – goed hebben opgesteld' (interview 9 juli).
Onderdeel van de BEC-structuur is de voormalige Regiegroep Monitoring en Effectmeting (M&E), bestaande uit vertegenwoordigers van elke directie en voorts van AEP, FEZ en RVO, die offerteaanvragen voor beleidsevaluaties bespreekt. In 2016 is de Regiegroep opgegaan in het beleidsanalistenteam (BAT). De vertegenwoordiger of 'linking pin' van de directie in de Regiegroep fungeert als eerste aanspreekpunt voor vragen vanuit de beleidsdirecties over onderzoeksmethoden en over wat men van een onderzoeksbureau kan verwachten (BEC-secretariaat, z.j., p. 26). Voorts: al in de ex ante fase, wanneer één van de directies van DG B&I aan een nieuw beleidsinstrumentarium werkt, biedt de Regiegroep begeleiding bij het opstellen van een monitoring- en effectmetingsplan waarbij de dossierhouder wordt uitgedaagd te bedenken wat hij nodig heeft om de evaluatie op een later moment in goede banen te kunnen leiden – om te achterhalen 'wat het de samenleving oplevert' (interview 9 juli). De Regiegroep M&E functioneert anno 2015 alleen voor het DG B&I en vergadert gemiddeld eens in de zes weken (interview 9 juli). Ook binnen de andere DG's wordt gekeken naar verdere institutionalisering van de regiefunctie.
Het BIT-EZ ten slotte, wordt door respondenten gezien als een innovatief team dat ernaar streeft om gedragswetenschappen een plek te geven in het beleid – de *evidence based* beleidsaanpak. In deze aanpak worden gedragsinterventies getest in beleidsexperimenten (RCT's) voordat ze worden opgeschaald. Dit heeft zijn doorwerking in het evaluatieonderzoek, hoewel het zwaartepunt van BIT aan de voorkant, bij de beleidsvorming ligt.
Bij het ministerie van IenM is de hoofddirectie Financiën, Management en Control (FMC) betrokken bij de meeste evaluaties en bij alle beleidsdoorlichtingen (interview 16 juli). Het initiatief voor beleids- en wetsevaluaties ligt bij de beleidsdirecties, maar wordt soms ook aangejaagd door FMC.
Bij BZK is sinds september 2014 de coördinatie van de evaluatiefunctie bij DFEZ belegd. Voorheen vervulde de inmiddels ontmantelde directie Kennis (later BO-KIEM) deze functie (interview 13 oktober). FEZ coördineert de planning van de beleidsdoorlichtingen, monitort de voortgang en ondersteunt dit proces waar nodig of gewenst. Dit gaat verder dan de reguliere control-taak van FEZ bij het beoordelen

[83] Het DG B&I heeft de Expertwerkgroep Effectmeting gevraagd, voor zes beleidsinstrumenten mogelijke evaluatieontwerpen uit te werken, waaronder innovatiekredieten en topconsortia voor Kennis en Innovatie (TKI). Het rapport van de commissie, de hoofdopdracht van de expertwerkgroep, was tot een advies te komen over de meest geschikte evaluatieontwerpen om de effectiviteit van bovengenoemde instrumenten in kaart te brengen gegeven het doel, de vormgeving, de werking van de instrumenten en de beschikbaarheid van data. Daarnaast is de werkgroep gevraagd een oordeel uit te spreken over de onderlinge kwaliteit van de verschillende evaluatieontwerpen en deze te rangschikken in volgorde van inzichtelijkheid en robuustheid op de effectiviteit van het instrument (Commissie Theeuwes, 2012:V).

van doorlichtingen. Daarnaast wordt gewerkt aan een dekkende programmering van onderliggende evaluaties en aan kennisdeling tussen de verschillende directies.

Tabel 7 geeft een overzicht van evaluatie-ondersteunende afdelingen bij departementen (exclusief externe onderzoeksinstituten). Niet expliciet genoemd in de tabel zijn de beleidsdirecties en daarbinnen de onderzoekscoördinatoren, aangezien die doorgaans aanhaken op genoemde organisatieonderdelen.
Hetzelfde geldt voor directies WJZ bij andere departementen. Voor zover we hebben kunnen nagaan zijn wetgevingsdirecties nergens *structureel* betrokken bij ex post evaluatieonderzoek. De Hoofddirectie Bestuurlijke en Juridische Zaken (HBJZ) bij IenM heeft in het verleden wel wetsevaluaties gecoördineerd – de nadruk lag veelal op de juridische kant. Met de komst van het IAK in 2011 zijn ze wel vaker betrokken aan de voorkant, bij de totstandkoming van wetgeving en beleid.

Tabel 7 Organisatieonderdelen bij beleidsdepartementen met taken op het vlak van evaluatie en evaluatiekwaliteit

Departement	**Afdelingen / directies nauw betrokken bij evaluatieonderzoek**
BZ	• FEZ • IOB - Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie
BZK	• FEZ: coördineert sinds september 2014 de evaluatiefunctie • Voorheen vervulde de sinds 2011 ontmantelde directie Kennis & Strategie deze functie

EZ	• Directie AEP - Algemene Economische Politiek
	• FEZ - Financiële en Economische Zaken
	  o BEC - Beleidsevaluatiecommissie
	• *Behavioural Insights Team* (BIT)
	• DG B&I – Directoraat-Generaal Bedrijfsleven & Innovatie
	  o Regiegroep Monitoring & Effectmeting (sinds 2016 opgegaan in het Beleidsanalistenteam (BAT)

Financiën (stelsel)	• DG Rijksbegroting o IRF – Inspectie der Rijksfinanciën
IenM	• FMC - Financiën, Management en Control • BIT - *Behavioural Insights Team* (vooral ex ante) • KiM – Kennisinstituut voor Mobiliteitsbeleid e.v.a. 'externe' kennisinstituten
OCW	• FEZ • Directie Kennis
SZW	• FEZ – Financiële en Economische Zaken • *Chief Science Officer* (CSO) met werkgroep en andere kennisonderdelen

VenJ	• FEZ (vooral gericht op beleidsdoorlichtingen)
	• Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC)
	  o Afdeling EWB, Extern-wetenschappelijke betrekkingen (betrokken bij uitbesteding evaluatie- en overig onderzoek)
	  o Interne onderzoeksafdelingen

VWS	• FEZ • ZonMw o CER - Commissie Evaluatie Regelgeving. Via het gelijknamige programma (PER) wordt een deel van de externe evaluaties van wet- en regelgeving extern verricht

Kennisdirecties

Met name daar waar FEZ een minder centrale rol vervult, is de stimulans en ondersteuning voor evaluatieonderzoek deels bij andere afdelingen belegd. Deze maken vaak deel uit van een groter netwerk van kennis- of onderzoeksinstellingen.

Bij OCW heeft de stafdirectie Kennis een belangrijke rol in de borging van de kwaliteit van (wets)evaluaties en ander onderzoek. De vaste accounthouders bij deze directie ondersteunen de dossierhouders en de beleidsdirecties zowel op eigen initiatief als op verzoek (interviews 1 juli; 26 augustus).[84] Een van de doelen is om de kwaliteit van evaluatieonderzoek, waaronder quasi-experimenteel doeltreffendheidsonderzoek, te bevorderen. De directie Kennis zit het overleg van onderzoekscoördinatoren voor. De onderzoekscoördinatoren hebben binnen hun directie of DG relatief veel ervaring met evaluatieonderzoek. Zij helpen (toekomstige) dossierhouders bij inhoudelijke en procesmatige zaken, zoals de formulering van de vraag- en probleemstelling, de opzet van de startnotitie en de keuze van BC-leden (interview 26 augustus).
Om quasi-experimenteel doeltreffendheidsonderzoek te stimuleren heeft de Kennisdirectie de afgelopen jaren een beperkt aantal onderzoeksopzetten laten opstellen door CPB en SCP, als richtlijn voor startnotities en offerteaanvragen. De beslissing om dit design wel of niet te gebruiken, ligt bij de directie (interviews 1 juli en 26 augustus).
In aanvulling op de centrale coördinerende rol van FEZ beschikt SZW over een *Chief Science Officer* (CSO), een wetenschapper die inhoudelijke ondersteuning en richting biedt aan (evaluatie)onderzoek en uiteenlopende andere kennisinitiatieven.[85] De CSO geeft advies aan de onderzoekscoördinatoren bij de beleidsdirecties, waar de beslissingen vallen over uit te voeren evaluatieonderzoeken. Hij kan benaderd worden door beleidsdirecties, of zitting nemen in een BC. FEZ is ook betrokken bij overleggen met de CSO, uit oogpunt van de coördinatie van evaluatie- en overig onderzoek (interview, 15 juli).[86] Daarnaast heeft SZW een interne Regiegroep Kennis – ondersteund door de CSO en zijn werkgroep – die zich buigt over de Hoofdlijnennotitie Kennisvragen SZW, de strategische kennisagenda. Verder zijn er 'kennisplatformen'[87] en is er het kennispartnerberaad, een overleggroep met externe kennis- en onderzoeksinstituten. De twee DG-kolommen hebben sinds 2015 elk een club medewerkers die zich bezig houden met (strategisch) onderzoek. Binnen het DG Werk is dat de directie Arbeidsmarkt en Sociaal-Economische Aangelegenheden (ASEA) en onder Sociale Zekerheid en Integratie (SZI) valt het 10 personen tellende Strategie Stelsel en Onderzoek (SKO).
Iets vergelijkbaars geldt voor EZ, met de directie Agro-en Natuurkennis (bij DG Agro & Natuur) en de directie Kennis en Innovatie (KI) bij DGB&I. Het DG ETM heeft een cluster Kennis en Strategie. Met name dit laatste cluster heeft een specifieke rol in evaluatieonderzoek. Voorts speelt de Rijksdienst voor Ondernemend Nederland (RVO.nl) een belangrijke rol in de kennisfunctie via de afdeling Kwaliteit, Monitoring en Effectmeting (KME). KME werkt in afstemming met beleid aan het opstellen van monitoring- en effectenplannen, beziet welke data beschikbaar zijn en denkt via deelname aan de voorm. Regiegroep M&E bij B&I en deelname aan de BEC mee over de opzet van evaluatieonderzoeken. Net als bij andere departementen is ook

[84] De accounthouders leveren evidentie en onderbouwing, brengen beleidsmedewerkers in contact met externe experts, of denken en schrijven mee aan een beleidstraject. Zij zijn ook agenderend wat betreft (grotere) kennisvragen en geven een signaal af zodra zij een bijdrage van de directie Kennis zinvol of nodig achten (e-mail 9 juli 2015).
[85] Vgl. de speciaal benoemde Hoogleraar beleidsevaluaties bij EZ.
[86] Een vergelijkbare functie werd in het verleden vervuld door de stafdirectie Analyse en Onderzoek – waar echter een duidelijke relatie met de beleidsdirecties werd gemist (interview 15 juli, 2015).
[87] Zoals het Kennisplatform Werk en Inkomen (KWI) en Kennisplatform Integratie en Samenleving (KSI).

sprake van een deels extern kennisnetwerk, zoals Wageningen UR met daaronder onderzoeksinstituten zoals Alterra of het LEI.[88] Deze onderzoeksinstituten voeren soms evaluatieonderzoeken uit voor EZ.
De situatie bij IenM laat zich vergelijken met EZ; er is een kennisprogramma gekoppeld aan de strategische kennis-en innovatieagenda (IenM, 2013). Een belangrijke rol – eveneens aan de voorkant – vervult voorts het *Behavioural Insights Team* (BIT), een netwerk van kennisinstellingen op het gebied van gedrag dat IenM ondersteunt bij de toepassing van gedragsinzichten in het maken en implementeren van beleid en wetgeving (Kennisplatform Crow, 2014). Verder is er een leergang Kennis voor Beleid.[89]
De Directie Kennis, innovatie en strategie voert de regie op de kennisfunctie bij IenM. Deze stuurt op een groot netwerk van kennisinstituten en wetenschappers bestaande uit een binnenring (KiM, RIVM, PBL, KNMI, RWS-WVL) en buitenringen.[90]
Bij het Kennisinstituut voor Mobiliteitsbeleid (KIM) is beleidsevaluatie de laatste van zes kennislijnen (KiM, 2015). Het PBL en het KiM verrichten voornamelijk ex ante evaluaties en monitoring en zijn minder actief op het vlak van ex post evaluatie (vgl. Berveling et al., 2009; interview 16 juli).[91]
Verbeterpunten die in een online presentatie genoemd worden t.a.v. de kennismanagementfunctie bij IenM zijn: landing van kennis in beleid, versnippering van budgetten en vraagsturing, zwakke vraagarticulatie, te veel focus op de korte termijn en beperkte samenwerking en concurrentie tussen kennisinstellingen (Riedstra, IenM Kenniscongres 2 april 2014). Dit lijken aandachtpunten die verder reiken dan alleen het domein van IenM.
Onder het DG Bestuur en Koninkrijksrelaties van het ministerie van BZK valt in het huidige organogram het Bureau Verkenning en Onderzoek (BVO), dat naast strategische kennisontwikkeling en kennismanagement ook een rol vervult ten aanzien van evaluaties (en wetsevaluaties). BVO draagt zorg voor de DG-brede onderzoeksprogrammering en – budget en voor de procesmatige en inhoudelijke begeleiding va het onderzoek. Bij het DG Wonen en Bouwen is er de directie Kennis & Verkenningen, met een vergelijkbare functie (interview 13 oktober).[92]
Hoewel veel van de genoemde instituten vaak ook evaluaties verrichten en bijdragen leveren aan evaluatiekennis en -inzichten voor departementen, bevatten de verschillende geraadpleegde *kennisagenda's* zelden verwijzingen naar evaluatieonderzoek.

Onderzoeksinstellingen bij en voor departementen
Onderzoeken waaronder evaluaties voor de ministeries van VenJ – en BZK – worden verricht door of via het ongeveer 95 fte tellende Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC), dat een zelfstandige positie heeft ten opzichte van het aanvragende ministerie.[93] Het WODC doet drie soorten onderzoek:

---

[88] www.wageningenur.nl/nl/Expertises-Dienstverlening.htm biedt een overzicht.
[89] Presentatie Riedstra, IenM Kenniscongres d.d. 2 april 2014 (online gepubliceerd).
[90] SWOV, CROW, CPB, SCP, TNO, Deltares, NLR, IMARES en verder o.a. universiteiten en NWO.
[91] In deze KiM-publicatie 'Na het knippen van het lint' pleiten de auteurs voor het vaker doen van (lichte) ex post evaluaties van infraprojecten om te bezien wat hier de meerwaarde van kan zijn. De nadruk ligt bij infrastructuurprojecten sterk op ex ante onderzoek ('Wie een beslissing moet nemen over de aanleg van infrastructuur gaat niet over één nacht ijs').
[92] www.rijksoverheid.nl/ministeries/ministerie-van-binnenlandse-zaken-en-koninkrijksrelaties/inhoud/organisatie/organogram/directoraat-generaal-wonen-bouwen-en-integratie. Met de reorganisatie van BZK, vanaf april 2016, moeten BVO en K&V geïncorporeerd zijn binnen de dan nieuwe directies.
[93] www.wodc.nl/organisatie. Soms verricht het WODC ook (evaluatief) onderzoek in opdracht van het ministerie van BZK, zoals bijvoorbeeld op het terrein van koninkrijksrelaties en als gevolg van verschuivingen tussen VenJ en BZK in departementale 'portefeuilles' tijdens de looptijd van evaluaties (bijv. op het vlak van migratie en politie).

probleemverkennend en -verklarend onderzoek, monitoring & evaluatie en R&D (WODC zelfevaluatie 2013; en -onderzoeksprogramma 2015). Het is zelf inhoudelijk betrokken bij de onderzoekprogrammering, zie par. 3.3 (WODC, 2015; interview 8 juni). Naast het verzorgen van uitbesteding aan externe partijen verricht het WODC een deel van het evaluatieonderzoek intern – minder dan de helft. Van de *wets*evaluaties wordt niettemin een relatief groot deel uitbesteed (interviews 8 juni). Het proces van uitbesteding en verdere projectbegeleiding loopt via de WODC-afdeling Extern Wetschappelijke Betrekkingen (EWB).
Beleidsevaluaties worden bij het ministerie van Buitenlandse Zaken (BZ) door de Inspectie Ontwikkelingssamenwerking en Beleidsevaluaties (IOB) uitgevoerd. De IOB verricht echter *geen wets*evaluaties. Tot de taken van IOB behoren de programmering en uitvoering, intern en extern, van onafhankelijke beleidsevaluatie en de verzameling van benodigde kennis om nieuwe interventies doelgericht voor te bereiden (BZ, 2009; 2015). De aanpak en methodologie die IOB bij beleidsevaluaties gebruikt is vastgesteld in de handreiking 'Evaluatiebeleid en richtlijnen voor evaluatie' (BZ, 2009).
Het Ministerie van VWS werkt voor specifiek *wetsevaluaties* samen met ZonMw[94] om precies te zijn met de Commissie evaluatie regelgeving (CER), de uitvoerder van het – in par. 3.5 te bespreken – programma Evaluatie Regelgeving (PER). ZonMw financiert de uitvoering van specifiek evaluaties van wet- en regelgeving door externe onderzoeksteams. Evaluatie-uitvoerders in spé kunnen zich inschrijven op openbare tenders. De CER adviseert over de opdrachtformulering, begeleidt het onderzoek en bewaakt de kwaliteit. De CER bestaat uit deskundigen op het terrein van beleid en bestuur van de gezondheidszorg, het gezondheidsrecht in brede zin, onderzoeksmethodologie en het patiëntperspectief. Alle commissieleden nemen op persoonlijke titel en zonder last of ruggespraak deel aan de commissievergaderingen (ZonMw, 2013, p.8).
Behalve gericht op (voornamelijk) één departement zijn er ook evaluerende instituten die die evaluatie- en ander onderzoek verrichten voor meerdere departementen, zoals het Sociaal Cultureel Planbureau (SCP). Het SCP voert bovendien geregeld wetsevaluaties uit, zoals op het vlak van de decentralisaties (SCP, 2014). Dat geldt ook voor het Planbureau voor de Leefomgeving (PBL), bijvoorbeeld met de evaluatie van de Wet ruimtelijke ordening (Wro). De instituten hebben vaak hun eigen kwaliteitsnormen of –richtlijnen, zoals blijkt uit bijvoorbeeld de Strategische Koers 2015-2020 van het SCP (2014). Het SCP heeft een overkoepelende BC met vertegenwoordigers uit wetenschap, beleid en praktijk, die waakt over zowel het wetenschappelijke niveau als de maatschappelijke relevantie. Het WODC heeft een Raad van Advies. Instituten zoals SCP, PBL en WODC worden regelmatig gevisiteerd of ge-audit.[95]

Directies Wetgeving en/of Juridische Zaken
De verantwoordelijkheid voor het initiëren en (laten) verrichten van wetsevaluaties ligt primair bij de *beleidsdirecties* en doorgaans niet bij de directies Wetgeving en/of juridische zaken (IOWJZ, alle interviews) – die bij elk van de onderzochte departementen behalve BZ vertegenwoordigd zijn. De voorzieningen voor wetsevaluatie verschillen (hierdoor) niet structureel van die voor evaluatie algemeen. Departementen hebben geen aparte afdelingen voor wetsevaluaties en evenmin een apart wetsevaluatiebeleid, zo leren desk study en de interviews. Net als bij 'reguliere' beleidsevaluaties wordt een belangrijk deel van het Nederlandse wetsevaluatieonderzoek uitbesteed.

[94] www.zonmw.nl. ZonMw financiert gezondheidsonderzoek en zorginnovatie. Hoofopdrachtgevers van ZonMw zijn VWS en NWO. Het programma Evaluatie Regelgeving is een van de programma's bij ZonMw.
[95] Vgl. o.a. WODC (2014) en van het PBL www.pbl.nl/publicaties/2012/zelfevaluatie-pbl-2012 .

Een overeenkomst tussen de bestudeerde departementen is dat de wetgevingsdirecties weinig betrokken zijn bij ex post evaluaties van wetten (alle interviews). Zo stelt een respondent bij IenM met ongeveer 30 jaar ervaring als wetgevingsjurist dat hij slechts enkele wetsevaluaties voorbij heeft zien komen, ondanks dat deze zijn warme belangstelling hebben (interview 24 augustus).
Wel is DWJZ nogal eens de formele opdrachtgever als meerdere beleidsdirecties bij een wetsevaluatie betrokken zijn (sectoroverstijgend) (interview, 1 juli). Een voorbeeld zijn de opeenvolgende evaluaties van de Awb (par. 3.5).
Directies Wetgeving zijn bovendien meer nadrukkelijk betrokken bij het begeleiden van ex ante onderzoek en met het toetsen van nieuwe wetsvoorstellen aan de IAK-richtlijnen (interviews 9, 16 en 24 juli).

Intern uitgevoerde versus extern uitbestede evaluaties
Zoals op basis van eerder onderzoek verwacht (H1), is uit alle gehouden interviews gebleken dat evaluaties en wetsevaluaties op rijksniveau voor het (over)grote merendeel uitbesteed worden aan externe partijen zoals academische onderzoekteams en onderzoeksbureaus.
Voor het uitbesteden van evaluaties zijn een drietal verklaringen gegeven door respondenten: (1) Het ontbreekt directies aan tijd en capaciteit; het verrichten van evaluaties komt bovenop de andere taken van beleidsmedewerkers. 'Het is eenvoudiger om er euro's tegenaan te gooien' (interview 26 augustus en 15 september); (2) De uitvoering van evaluaties wordt vaak aan externe onderzoekers overgelaten omdat specialistische kennis vereist is; en ten slotte (3) wordt extern onderzoek gezien als bevorderlijk voor objectiviteit (merendeel interviews).
Bij OCW worden evaluaties standaard extern uitgezet. Het grootste deel hiervan komt tot stand via het raamcontract. Raam- of mantelcontracten worden voor een vooraf bepaalde periode afgesloten met een vaste groep van onderzoeksinstellingen en –bureaus. De selectie, die verloopt via een meerledige procedure, is gebaseerd op de competenties en ervaring die de partijen bij aanvang in huis hebben.[96] VenJ en SZW werken niet met raamcontracten, binnen BZK gebeurt dit her en der wel (interviews 15 juli en 13 oktober).
Een klein deel van het onderzoek voor OCW wordt uitgezet via het Nationaal Regieorgaan Onderwijsonderzoek (NRO) (interview 1 juli). Het NRO is opgericht in 2012 met als doel praktijk, wetenschap en beleid samen te brengen; het valt onder de verantwoordelijkheid van de Nederlandse organisatie voor wetenschappelijk onderzoek (NWO) (interview 1 juli). Sinds 1 januari 2014 is het NRO volledig operationeel met een eigen programmering – van voornamelijk fundamenteel en praktijkgericht onderzoek – en een structureel jaarbudget van bijna 15 miljoen.[97]
Voor VenJ, BZK en VWS worden (wets)evaluaties ook door WODC en ZonMw en SCP uitgevoerd. Bij het WODC wordt een relatief groot deel van de evaluaties via de afdeling EWB weer 'door'-uitbesteed (interviews 8 juni).
Nieuwe organisatieonderdelen en structuren ter bevordering van ex post evaluaties vergen volgens meerdere respondenten een wenperiode: het is zoeken naar een rolverdeling die niet alleen op papier maar ook in de praktijk werkt (interviews 1 juli, 9 juli, 26 augustus).

3.2.3 Samenvattend over structuur

[96] Zo hebben beoordelaars bestaand onderzoekswerk van inschrijvende partijen bestudeerd of deze inhoudelijk voldoende 'in huis hebben' – toegespitst op één van de 15 expertisegebieden of 'percelen'. Dit gebeurde na een aanvankelijke beoordelingsronde met behulp van een reeks uitsluitingsgronden, minimumeisen en 'knock-out'-criteria (noodzakelijke voorwaarden) (comm. 16 feb 2016 en deskmateriaal). De OCW-Kennisagenda 2015 bevat overzichten van gecontracteerde en gesubsidieerde partijen.
[97] http://www.nro.nl/.

Er zijn in en rond elk onderzocht departement meerdere organen die zich (deels) met evaluatie en kennis bezighouden. De FEZ-directies en eventuele omringende organisatieonderdelen zijn primair gericht op beleidsevaluaties, met name in het kader van verantwoording en beleidsdoorlichtingen. Wat evalueren betreft, vullen directies FEZ hun rol per departement verschillend in. Zo heeft FEZ bij SZW een stimulerende en aanjagende rol in zowel de planning, de kwaliteitsborging als het gebruik van evaluatieonderzoek en beleidsdoorlichtingen, terwijl andere directies FEZ dichter op hun formele controletaken blijven. Binnen EZ is rond FEZ een netwerk van organisatieonderdelen opgezet dat zich specifiek richt op de kwaliteit van doeltreffendheidsevaluaties.
Sommige departementen hebben een onderzoeksinstituut dat onder meer evaluaties doet of uitzet, zoals het WODC bij VenJ en de IOB bij BZ.
Kennisorganen dragen zorg voor meer strategisch en toekomstgericht onderzoek. In de praktijk maakt evaluatieonderzoek in zeer beperkte mate deel van uit van het kennisgeoriënteerde werk. Het onderscheid tussen de taken van financieel-economische en kennisdirecties is echter niet zwart-wit. Neem bijvoorbeeld de rol die het CPB en een enkele kennisdirectie hebben gespeeld in de bevordering van quasi-experimentele doeltreffendheidsstudies; en omgekeerd de rol die Kennis-SZW vervult in relatie tot de kennisagenda en algemene evaluatie-initiatieven.
Evaluaties worden voor het overgrote deel uitbesteed aan externe partijen, vanwege interne capaciteitsrestricties en vermeende objectiviteit.
Veel van de organisatieonderdelen en -structuren zijn nog vers en vergen een wenperiode: het is zoeken naar een rolverdeling die niet alleen op papier, maar ook in de praktijk werkt.

3.3 Evaluatieprogrammering

In hoofdstuk 1 is beschreven dat evaluatievermogen ook betekent dat evaluaties daadwerkelijk worden uitgevoerd. Dit wordt bevorderd door een (meerjarige) evaluatieprogrammering. Die geeft volgens de literatuur ruimte aan evaluaties en helpt zo tegenwicht te bieden aan de (politieke) waan van de dag.

3.3.1 Beleidsdoorlichtingen

Elk departement is verplicht om jaarlijks (tenminste) één programmering van alle soorten onderzoek vast te stellen en in te dienen bij Financiën. Deze wordt als bijlage bij de Rijksbegroting opgenomen. Het onderzoek is ingedeeld in drie categorieën: (1) beleidsdoorlichtingen (2) ander (evaluatie)onderzoek naar doeltreffendheid en doelmatigheid en (3) overig onderzoek.
Er is geen apart kopje voor wetsevaluaties, wat laat zien dat deze geen afzonderlijk item vormen binnen het centrale verantwoordingsstelsel maar opgaan in de beleidsevaluaties (vgl. interview 9 juli).
De meeste bevraagde departementen maken, buiten de begrotingsprogrammering om, geen melding van een proactieve evaluatieprogrammering gericht op leren. Dit komt volgens een respondent vooral omdat evaluatie gezien wordt als ‘beleidsinstrument’ (interview 9 april). Evaluaties worden volgens hem gedaan als het goed uitkomt of als er om wordt gevraagd (vgl. interviews 2 en 15 september). Het creëren van een periodieke ‘feedback-loop’ naar het beleid lijkt over het algemeen niet de belangrijkste drijfveer.
Via het instrument van de beleidsdoorlichting wordt stelselbreed steeds vaker gehamerd op het belang van een adequate programmering van evaluatieonderzoek, om inhoudelijke lacunes in de doorlichting – die per begrotingsdoel iedere 4-7 jaar moet worden opgeleverd – tegen te gaan (RPE 2014; Financiën, 2016). Bij in elk geval BZ, EZ, IenM en SZW is de directie FEZ –in samenwerking met andere

organen – hier naar eigen zeggen gerichter op gaan sturen. Idealiter is de programmering dus zo opgebouwd, dat deze leidt tot evaluatierapporten waarin voldoende informatie staat om onderbouwde uitspraken over doeltreffendheid en doelmatigheid mogelijk te maken. Bij sommige departementen is dit de afgelopen jaren een worsteling gebleken, wat zich weerspiegelt in de inhoud en kwaliteit van beleidsdoorlichtingen (interviews 1 juli en 16 juli). Informanten bij de centrale evaluatie-instituties van het ministerie van Financiën en de Algemene Rekenkamer hebben dit bevestigd (26 augustus en 1 september).[98] Veel respondenten gaven aan dat niet al het onderzoek dat in de evaluatie- en onderzoeksprogrammering van de Rijksbegroting staat, ook daadwerkelijk bruikbaar is bij de invulling van de beleidsdoorlichtingen – bijvoorbeeld omdat het zich niet primair richt op doeltreffendheid en doelmatigheid. Dit laatste hoeft ook niet steeds het streven te zijn, er zijn immers vele typen evaluatieonderzoek – waaronder proces- en werkingsevaluaties – die bruikbare kennis voor beleid genereren.
Onder meer bij OCW bestond behoefte aan een overzicht van welk afgerond evaluatieonderzoek voor beleidsdoorlichtingen zou kunnen worden gebruikt; hier is dan ook met mensen van de ADR aan gepuzzeld (interview 1 juli). Maar deze respondenten stellen ook, dat veel vooruitgang is geboekt op het gebied van doeltreffendheidsevaluatie (interview, 1 juli).
FEZ-SZW voert gesprekken met beleidsdirecties over de evaluatieagenda en -planning. Dit gebeurt vooral in het verband van beleidsdoorlichtingen en met als doel gedegen doeltreffendheidsevaluaties te laten verrichten die daarvoor relevant en bruikbaar zijn. Ook helpt FEZ overzicht te scheppen in de programmeringen van verschillende directies (interviews 15 juli, 13 oktober).
Bij EZ heeft FEZ een interne Handreiking evaluatieonderzoek (FEZ-EZ, z.j.) opgesteld (zie verder par 3.5), waarin ook richtlijnen staan voor de evaluatieprogrammering. ‘Reeds bij het opzetten van een instrument wordt idealiter nagedacht over het evalueren van het instrument. Voor ieder instrument zou duidelijk moeten zijn wanneer de evaluatie plaatsvindt’ (p.9). De Regiegroep Monitoring en Evaluatie (M&E) van DG B&I – inmiddels BAT geheten – en de BEC-contactpersoon (bij de andere DG’s) vervullen daarbij een sturende rol. Een goede evaluatieprogrammering is vereist om de effecten van subsidies en beleid door middel van evaluatie te kunnen aantonen. De programmering van het evaluatieonderzoek wordt afgestemd op de beleidsdoorlichtingen, zowel qua timing als qua afdekking van beleidsartikelen.
Een subsidieregeling vervalt automatisch na maximaal 5 jaar – of wettelijk anders bepaalde termijn, tenzij de Tweede Kamer voorafgaand aan het aflopen van de subsidieregeling heeft ingestemd met een voorstel om de regeling voort te zetten. Het is dan ook zaak om, voorafgaand aan het voorstel tot eventuele voortzetting, een mede op een evaluatie gebaseerde onderbouwing te hebben. Ook voor beleidsdoorlichtingen wordt bij EZ in de praktijk een cyclus gehanteerd van 5 jaar. Omdat ex post evaluatie bij EZ primair als een beleidsaangelegenheid en onderdeel van de beleidscyclus wordt gezien, speelt DWJZ geen expliciete rol bij de evaluatieprogrammering – wel is de wetgevingsdirectie actief aan de voorkant, onder meer via betrokkenheid bij het IAK en door deelname aan de Monitorcommissie, waar nieuw en aangepast beleid wordt getoetst. De beleidsdirecties zijn verantwoordelijk voor hun eigen evaluatieplanning en stemmen dit af met FEZ (FEZ-EZ z.j., p. 9).
BZ heeft een evaluatieprogrammering voor vijf jaar, die jaarlijks wordt bijgesteld. Deze bevat alleen beleidsdoorlichtingen en andere evaluaties naar doelmatigheid en doeltreffendheid – geen ‘overige onderzoeken’ (interview, 2 september). Ook hier

[98] Zie ook http://verantwoordingsonderzoek.rekenkamer.nl/2014/rijksbreed/beleidsresultaten/beleidsdoorlichtingen-nog-te-weinig-zicht-op-doelmatigheid-en

houdt men rekening met de timing voor beleidsdoorlichtingen. De evaluaties betreffen geen wetgeving maar beleid en subsidies. FEZ en IOB werken nauw samen bij de programmering van evaluatieonderzoek met input vanuit de vier DG's van het ministerie: DG Europese Samenwerking, DG Internationale Samenwerking, DG Politieke Zaken en DG Buitenlands Economisch Beleid. IOB is verantwoordelijk voor de programmering van het DG Internationale Samenwerking en heeft een adviserende rol bij de evaluatieprogrammering van de andere drie DG's (IOB, 2009, p. 12). De meest voorkomende oorzaken van de bijstelling van de evaluatieprogrammering zijn veranderingen in de steun van het Kabinet voor een bepaald beleid en grote maatschappelijke veranderingen, zoals oorlog in een ontwikkelingsland (interview 2 september).
Een Panel van Advies (van de IOB) signaleerde in 2011 dat de nadruk in de evaluatieprogrammering sterk op de verantwoordingskant ligt (het afdekken van de begroting) en minder selectief, bijvoorbeeld via het consulteren van stakeholders, tot stand komt.

> 'Het punt van de optimale balans tussen de [verantwoordings- red.] en de leerfunctie van evaluaties wordt in het IOB-programmeringsproces niet geproblematiseerd. Aan de eerste functie lijkt een groter gewicht te worden toegekend dan de tweede, zoals tot uiting komt in het streven de programmering 'dekkend' te laten zijn in plaats van meer selectief te werk te gaan. Externe stakeholders en het maatschappelijk debat spelen niet of nauwelijks een rol in de huidige praktijk van de evaluatieprogrammering. (...) Het meer programmeren van gezamenlijke en samengestelde evaluaties zou van invloed kunnen zijn op de balans tussen leren en verantwoorden in de evaluatieprogrammering (...).' (Eerste Panel van Advies IOB, 2011:24).

Ondanks dat dit niet expliciet wordt benoemd, zijn meerdere beleidsdoorlichtingen voor een belangrijk deel op wetsevaluaties gebaseerd. Bij SZW zijn relatief veel beleidsdoorlichtingen ontleend aan één of meerdere evaluaties van sociale zekerheidswetten, al dan niet in combinatie met beleidsmaatregelen, zoals bij de beleidsdoorlichtingen van de Wet SUWI (2015), de WAO (2014), de Anw (2013) en de Arbowet (2012).[99] Ook bij andere departementen, zoals IenM en VenJ, worden wetsevaluaties gebruikt als belangrijk basismateriaal voor beleidsdoorlichtingen (interview 16 juli).
De minister van Financiën meldt in het Financieel jaarverslag van het Rijk over 2014, dat 17 van de 27 voorgenomen beleidsdoorlichtingen zijn uitgevoerd.[100]

3.3.2 (Evaluatie)onderzoek gericht op kennisvergroting

De onderzoekcoördinatoren bij SZW pogen dossierhouders te stimuleren om proactief na te denken over de evaluatie- en onderzoeksprogrammering; naast verplicht onderzoek ook over niet-verplichte nummers (interview 15 juli).
De kennisagenda van SZW (Hoofdlijnennotitie kennisvragen) van 2015 bevat tien thema's waaraan nieuw onderzoek wordt gekoppeld (interview 15 juli). Deze thema's ontstaan in overleg met directies, de CSO en zijn ondersteunende werkgroep. Vervolgens wordt dit voorgelegd aan de Regiegroep Kennis, waar het MT-SZW en drie directeuren in zijn vertegenwoordigd.
Bij het ministerie van OCW maken beleidsdirecties via interne rondes hun eigen onderzoeksprogrammering en zijn zij zelf hiervoor verantwoordelijk – hier valt ook het evaluatieonderzoek onder. De directie Kennis van OCW stelt ieder jaar de (strategische) kennisagenda op.[101]

---

[99] Deze en andere beleidsdoorlichtingen zijn te vinden via: www.rijksbegroting.nl/beleidsevaluaties/evaluaties-en-beleidsdoorlichtingen.
[100] Ibid.
[101] 'De kennisagenda is een poging een samenhangend verhaal te bieden over strategische uitdagingen voor het OCW-terrein. Daarmee is de kennisagenda niet alleen relevant voor de programmering van nieuwe kennisvragen, maar ook om de hoeveelheid kennis die

Het aandeel proces- en ex post evaluatie binnen de door ons geraadpleegde strategische kennisagenda's is verwaarloosbaar.[102]
De laatste jaren zou – volgens een onderzoekscoördinator bij tenminste één van de DG's van OCW– de onderzoeksprogrammering budgettair onder druk zijn komen te staan (26 augustus). Het niet teveel willen belasten van de organisaties in het onderwijsveld speelt daarbij eveneens mee.
Bij het ministerie van VenJ worden jaarlijks twee programmeringsrondes van evaluatie- en ander onderzoek geïnitieerd door het WODC (WODC, 2015). Voor iedere programmeringsronde worden de onderzoekwensen – al dan niet via de coördinatoren onderzoek – bij de beleidsdirecties en DWJZ geïnventariseerd aan de hand van een door het WODC opgestelde standaardvragenlijst. Het resulterende overzicht van onderzoeksvoorstellen wordt in het MT-WODC besproken. Dan wordt bepaald welke onderzoeken de komende periode zullen worden verricht en wordt de keuze gemaakt of het onderzoek WODC-intern wordt uitgevoerd of via de afdeling EWB wordt uitbesteed. De conceptprogrammering wordt in de Bestuursraad geaccordeerd en vervolgens gepubliceerd via de website van het instituut. De directie FEZ speelt geen nadrukkelijke rol in de programmering van individuele VenJ-evaluaties, behalve dat in de WODC-vragenlijst onder 'evaluatieonderzoek' een item is opgenomen waarin directies wordt gevraagd naar of en hoe het voorgenomen evaluatieonderzoek past binnen de meerjarenplanning van de beleidsdoorlichtingen – die uiteraard wel door FEZ worden gecoördineerd.

3.3.3 Evaluatiebepalingen in wetten

Tien tot 20% van recent geëvalueerde wetten bevat een evaluatiebepaling, een wetsartikel dat vaststelt wanneer de wet zal worden geëvalueerd om inzicht te geven in de werking ervan (Veerman, 2013).[103] Dat gebeurt in de regel vijf jaar na inwerkingtreding, hoewel een tussentijdse evaluatie ook mogelijk is in het geval dat de implementatie van de wet anders verloopt dan verwacht, of de Tweede Kamer eerder om informatie vraagt. Een wet kan eenmalig of periodiek worden geëvalueerd (Aanwijzing 164 voor de regelgeving).
Evaluatiebepalingen worden volgens ingewijden vaker dan voorheen ingezet als politiek 'wisselgeld'. Voor wetten die geen evaluatiebepaling bevatten, kan de Tweede Kamer, indien noodzakelijk, ad hoc opdracht geven tot evaluatie (Bemelmans-Videc, 2002). Dit laat zien dat ook wetsevaluaties zich in een politieke context bevinden en hun aanleiding soms direct ligt in de politieke realiteit.
Hoewel enkele respondenten menen dat evaluatiebepalingen in de praktijk vanwege uitstel of afstel niet altijd worden nageleefd, is hier geen systematisch onderzoek naar verricht. Bij OCW houdt DWJZ een kalender bij met daarin de evaluatietermijnen van wetten naar aanleiding van de evaluatiebepalingen (e-mail 14 juli).[104]
De termijnen in de evaluatiebepalingen – of naar aanleiding van een Kamertoezegging – worden door respondenten bij Hoofddirectie Bestuurlijke en Juridische Zaken (IenM) vaak te kort gevonden. Zo concludeerden onderzoekers naar aanleiding van de eerste – door artikel 5.9a verplichte evaluatie van de *effecten* van de Crisis- en Herstelwet ('binnen twee jaar') dat na die termijn slechts twee rechterlijke uitspraken te vinden waren en toch geforceerd conclusies zijn

---

beschikbaar is en komt, in context te plaatsen en beter te benutten.'(OCenW, 2015).
[102] Geraadpleegd zijn de (strategische) kennisagenda's van de ministeries van SZW (2013); OCW (2015b); IenM (2013) en BZK (2010).
[103] Veerman (2013), mondelinge communicatie. Volgens sommige respondenten neemt het aandeel wetten met een evaluatiebepaling toe (o.a. interviews, 16 juli 2015); dit is niet nader onderzocht.
[104] Dit gebeurt via PARIS, het departementale systeem van toezeggingen aan de TK.

getrokken (interview 16 juli). Een vergelijkbaar probleem was er bij de 'Taxiwet' [105], met als belangrijkste kenmerk dat die gemeenten de bevoegdheid heeft gegeven om verordeningen met kwaliteitseisen te maken in aanvulling op het landelijk kader. De staatssecretaris schreef de Kamer in 2013 dat een evaluatie van de wet gelet op de ontwikkelingen in het veld (pas) zou kunnen worden uitgevoerd vanaf de tweede helft van 2014.[106] Dit omdat het enkele jaren heeft geduurd tot de eerste verordening (Amsterdam) beschikbaar kwam. Deze evaluatie is gepresenteerd als een tussentijds onderzoek (Kort et al., 2014).
Voor VWS vormt ZonMw een van de wegen om wetsevaluaties uit te zetten. Daarnaast kunnen de betrokken beleidsdirecties ook zelf de wetsevaluatie uitbesteden (interview 9 april). Het uitbesteden van wetsevaluaties via ZonMw is een wat langduriger en intensiever proces, omdat veel aandacht wordt besteed aan de kwaliteit van zowel de aanbestedingsprocedure als aan de inhoudelijke kwaliteit van het rapport (zie verder par. 3.5). Dit kan het voor beleidsmedewerkers ook minder aantrekkelijk maken; niet alleen vanwege het vaak langere tijdpad, maar ook omdat beleidsdirecties soms een minder diepgravend onderzoek willen, of slechts enkele gerichte vragen hebben (interview 9 april). Sinds 2008 zijn ook *thematische wetsevaluaties* onderdeel van het pakket van de CER. Deze evaluaties overstijgen individuele wetten en richten zich op aspecten van het systeem van gezondheidswetgeving, zoals bijvoorbeeld de regelgeving rond wilsonbekwaamheid. Een wetgevingsjurist bij IenM meldt dat hij 'aan de voorkant' stuurt op latere evaluaties van AMvB's, door reeds bij het opstellen van de tekst hiervan de evaluatiebepaling gericht uit te werken (interview 24 augustus). Dit is bijvoorbeeld toegepast op het Wijzigingsbesluit bodemenergiesystemen van 2013, dat drie jaar na inwerkingtreding dient te worden geëvalueerd.[107] Tot die nadere invulling laat Aanwijzing 164 wetgevingsjuristen de ruimte; bijvoorbeeld, welke evaluatiecriteria men voor de toekomst van belang acht en wat de scope van de evaluatie is (doeltreffendheid of minder verstrekkend). Op het terrein van IenM bestond tot 2010 ten slotte een evaluatieprogramma specifiek voor milieuwetgeving, STEM. (Box 5 in par. 3.5 geeft de ervaringen met dit programma weer.)

3.3.4 Politieke context en evaluatieprogrammering

De politieke context werkt – zo blijkt ook uit de praktijk – door in de evaluatieprogrammering en ook in het werken met evaluatiebepalingen. Genoemde oorzaken van bijstellingen van evaluatieprogrammeringen zijn veranderingen in kabinetssteun voor een bepaald beleid, of nieuwe maatschappelijke en beleidsontwikkelingen (interviews 9 juli; 26 augustus; 2 september). Sommige respondenten relativeren het feitelijke belang van een evaluatieprogrammering, door aan te geven dat evaluaties vooral worden gedaan als erom wordt gevraagd of als het strategisch goed uitkomt (interviews 8 april, 15 juni, 9 juli en 15 september). Evaluaties zijn daarbij door een enkeling geduid als politiek instrument. Bovendien is gesignaleerd dat de behoefte van beleidsmedewerkers is gericht op korte-termijnopbrengsten, terwijl onderzoek doorgaans een lange(re) doorlooptijd heeft (interview 15 juni).

Box 3.2 Oorzaken van weinig animo voor ex post evalueren van infraprojecten

[105] Deze aanpassing van de Wet Personenvervoer van 2010 kende geen evaluatiebepaling, aanleiding voor de evaluatie was een toezegging in de behandeling door de Tweede Kamer.
[106] Brief Staatssecretaris van Infrastructuur en Milieu aan de Tweede Kamer, 26 juni 2013 IENM/BSK-2013/94127.
[107] Besluit van 25 maart 2013 tot wijziging van een aantal algemene maatregelen van bestuur in verband met regels inzake bodemenergiesystemen en enkele technische verbeteringen, Artikel VIII (https://zoek.officielebekendmakingen.nl/stb-2013-112.html). De evaluatie gaat vanaf begin 2016 van start (www.rwsleefomgeving.nl/onderwerpen/bodem-ondergrond/nieuws/2015/beleidsevaluatie/).

Berveling et al. (2009:5-6) zien vier typen oorzaken van weinig animo voor ex post evalueren – naar aanleiding van infrastructuurprojecten bij IenM (vgl. interviews 2 september ...).
1. Vanuit de politiek-bestuurlijke en beleidscontext speelt de wens om vooruit te kijken een belangrijke rol. Het ligt in de aard van beleidsmakers om bezig te zijn met de toekomst, niet met het verleden.
2. Daarnaast staan bepaalde psychologische processen de wil tot evalueren in de weg. Bij de start van een project kan een zogenaamde '*optimism bias'* een rol spelen. Dit houdt in dat betrokkenen (ex ante) vaak een te optimistisch beeld hebben van een project. Dit leidt ertoe dat men zich later vaak verzet tegen alles dat het positieve beeld verstoort. Meer in het bijzonder kan een negatieve evaluatie leiden tot reputatieschade.
3. Organisatorische belemmeringen kunnen de praktische uitvoering van evaluaties in de weg staan. Projectdirecties worden vaak direct 'na het knippen van het lint' opgeheven. Bovendien kan een gebrek aan geld en capaciteit er toe leiden dat ex-postevaluaties worden verdrongen door meer urgente zaken. Daarnaast kan het natuurlijk personeelsverloop, in combinatie met de lange doorlooptijden van infrastructuurprojecten, leiden tot verwatering van kennis en expertise.
4. De belangrijkste methodologische problemen zijn dat het moeilijk is om effecten van projecten te isoleren en om te weten hoe de wereld er zonder het project had uitgezien. Daarnaast is het moment van evalueren een punt dat aandacht verdient.

3.3.5 Samenvattend over programmering

Beleids- en wetsevaluaties vinden voor een belangrijk deel plaats in het kader van het centrale verantwoordingsstelsel. Daarnaast zijn er de (periodieke) evaluatieverplichtingen in de wetgeving zelf; die zo eveneens een programmering inhouden. Beleids- en wetsevaluaties kunnen, eenmaal afgerond, door directies worden gebruikt als basismateriaal voor een beleidsdoorlichting.
Eventuele extra of parallelle onderzoeksprogrammeringen, zoals strategische kennisagenda's, bevatten nauwelijks ex post evaluatieonderzoek en zijn vooral toekomstgericht, zoals ex ante- en probleemverkennend onderzoek.
In de context van verantwoording zijn de afgelopen paar jaar door genoemde organisatieonderdelen verschillende initiatieven genomen om evaluatieprogrammeringen proactiever te maken, vooral met het oog op aankomende beleidsdoorlichtingen en het meer doordacht onderzoeken van doeltreffendheid/doeltreffendheid en doelmatigheid. We hebben niet overkoepelend kunnen vaststellen wat hier de vruchten van zijn – bijvoorbeeld in termen van minder lacunes in beleidsdoorlichtingen.
Een Panel van Advies signaleerde in 2011 dat de nadruk in de evaluatieprogrammering van BZ en IOB sterk op verantwoorden ligt (het afdekken van de begroting) en weinig selectief tot stand komt, bijvoorbeeld via raadpleging van stakeholders. Dat verantwoorden in evaluatieprogrammeringen meer nadruk krijgt dan leren, lijkt relevant voor meer bestudeerde departementen.
De politieke context werkt door in de evaluatieprogrammering. Sommige respondenten relativeren het feitelijke belang van een programmering: zij stellen dat evaluaties vooral samenhangen met veranderingen in kabinetssteun voor een beleid of met beleidsaanpassingen en -ontwikkelingen.

3.4 Middelen

Voor (adequaat) evaluatieonderzoek zijn, naast de aanwezigheid van voornoemde aspecten van evaluatievermogen, ook middelen nodig.
De meeste departementen stellen eerst een integraal budget voor onderzoek vast. Alle uit te voeren evaluaties moeten binnen dat financieel kader passen. In de praktijk gaat men uit van de bestedingsplannen, voortkomend uit de jaarlijks in

oktober vastgestelde budgetten (interview 15 september). Kosten voor evaluaties komen ten laste van het betreffende beleidsartikel (althans bij EZ, interview 9 juli). Bij het uitzetten van evaluaties voor extern onderzoek bepaalt in veel gevallen de aanvrager, doorgaans de dossierhouder bij een beleidsdirectie, het budget voor offertes. Het is vervolgens aan hem of haar om binnen het budget van een concrete evaluatie te blijven en de voortgang van de uitputting te monitoren – met behulp van de eigen administratieve ondersteuning (interview FEZ-SZW 15 juli).[108]
Bij onderzoeksinstellingen gaat het anders. Het WODC beheert naast eventuele extra middelen van de aanvrager een eigen onderzoeksbudget van gemiddeld 12 mln. over de periode 2006-2012 (Van der Laan et al., 2013). Dit is in lijn met de zelfstandige status van het instituut.[109] ZonMw financiert een deel van de evaluaties van gezondheidswet- en regelgeving in opdracht van het ministerie van VWS. De CER, verantwoordelijk voor dit programma fungeert daarbij als 'doorgeefluik' van het VWS-budget (interview, 8 april).

Wat de niet-financiële middelen betreft, melden respondenten vaak dat beleidsdirecties weinig *tijd* hebben voor evaluaties en onderzoek (interviews 1 juli en 2 september). Dat blijkt onder andere uit beperkte capaciteit voor onderzoekscoördinatie. Een concreet voorbeeld betreft dat op papier 1 fte coördinatie beschikbaar is op een directie van 60 personen, maar dat het in de praktijk op minder dan een halve fte neerkomt. De financiële middelen hiervoor lijken in het betreffende geval verder terug te lopen (interviews 1 juli en 26 augustus). Dit zet volgens een beleidscoördinator de onderzoeksprogrammering onder druk. Een ander gevolg van dit tijdsgebrek is volgens respondenten dat beleidsdirecties vaak te laat beginnen aan evaluaties die wel op de agenda staan. Als een evaluatieopdracht te laat wordt opgesteld, snijdt dit de weg af voor tijdige gegevensverzameling, bijvoorbeeld in de vorm van een monitoringsysteem. Dit knelpunt is bij alle departementen herkenbaar.
Ondersteunende afdelingen (zoals FEZ bij SZW, FMC bij IenM, directie Kennis bij OCW, de onderzoekscoördinatoren bij zowel SZW als OCW, BEC en de Regiegroep bij EZ) zijn initiatieven gestart om beleidsmedewerkers te stimuleren eerder na te denken over wat nodig is om een evaluatie (of beleidsdoorlichting) op een gedegen manier te laten uitvoeren (en IRF bij Financiën in overkoepelende zin; interview 26 augustus).

3.5 Evaluatiekader

In deze paragraaf ligt de nadruk op de aanpak van evaluaties en wetsevaluaties – zowel procesmatig als inhoudelijk. De aandacht richt zich vooral op hulpmiddelen voor de aanvragers – meestal beleidsmedewerkers. Bekeken is welke programma's voor (wets)evaluaties bij departementen er zijn (geweest), en hoe daar in de praktijk uitvoering aan is gegeven. Daarbij is in kaart gebracht welke handreikingen, richtlijnen of leidraden zijn ontwikkeld voor het aanbesteden, begeleiden en uitvoeren van evaluatieonderzoek.
Tabel 8 biedt een overzicht van actuele programma's, handreikingen en richtlijnen. Behalve voor beleidsmedewerkers zijn deze hulpmiddelen[110] soms ook bedoeld voor onderzoekspartijen – bijvoorbeeld offranten – en andere belangstellenden.

[108] Bij SZW beheert FEZ wél een budget specifiek voor CBS-onderzoeken.
[109] Niettemin acht een onderzoekscoördinator binnen VenJ het vaak lange tijd onduidelijk wat het beschikbare onderzoeksbudget is waar men bij de afbakening en formulering van onderzoeks- en evaluatievragen van uit kan gaan (interview 15 juni 2015).
[110] Vooral de openbare.

Tabel 8 Evaluatiekaders (ex post) bij de onderzochte departementen (momentopname 2015; niet uitputtend)

Departement	Programma's, handreikingen en richtlijnen voor evaluatieonderzoek	Openbaar of intern
Centraal / overkoepelend	• Handreiking Evaluatieonderzoek ex post (Financiën, 2003) • Handreiking Beleidsdoorlichtingen (Financiën, 2016) • Cursussen bij Academie voor Wetgeving over IAK en over ex post evalueren • Rijksacademie voor Financiën, economie en bedrijfsvoering (2015). Beleid in perspectief	Openbaar
BZ	• Evaluatiebeleid en richtlijnen voor evaluaties, (IOB, 2009) • Meerdere evaluatiehandreikingen op het vlak van ontwikkelingssamenwerking	Openbaar
EZ	• Commissie Theeuwes (2012). *Durf te meten* Eindrapport Expertwerkgroep Effectmeting • FEZ-EZ (z.j.) Handreiking evaluatieonderzoek EZ • BEC-secretariaat (z.j.), Handreikingenboekje beleidskwaliteit en beleidsevaluatie • Evaluatiecursus op basis van deze en andere materialen – later overgeheveld naar de Rijksacademie voor Financiën, economie en bedrijfsvoering	Intern
IenM	• Tot 2010: Programma Structurele Evaluatie Milieuwetgeving (STEM) • Begrotingsevaluatieprogramma (nota 2013).	Openbaar Intern
OCW, directie Kennis	• CPB (2011). Notitie Zicht op effectiviteit. Evaluatieontwerpen voor onderwijs- en wetenschapsmaatregelen, bedoeld om de quasi-experimentele design te stimuleren • Kohnstamm Instituut, Pater et al. (2012). Methodiek voor Verklarende evaluatie	Intern Openbaar
SZW	• FEZ-SZW (2014). Handreikingen en sjablonen voor proces en inhoud beleidsdoorlichtingen. Met o.a. formats voor de uitbesteding van het onderzoek, het opstellen van de startnotitie, planning, aandachtspunten bij onderzoeksvragen, de hoogte van vergoeding voor onafhankelijk onderzoek, de afspraken met onafhankelijke deskundigen en het opstellen van de nota aan minister en de kabinetsreactie.	Intern Nb overlapt sterk met latere Handreiking beleidsdoorlichtingen (Financiën, 2016)
VenJ	*Procesmatig* • Wegwijzer WODC-EWB (2012) bij contractonderzoek in opdracht van het WODC • Handboek WODC-EWB (2008)- met name bedoeld voor nieuwe EWB-medewerkers *Inhoudelijk* • Kautto & Similä (2005). RIPI's: *Recently Introduced Policy Instruments and Interventions Theories* • Van Ooyen-Houben en Leeuw (2010). Evaluatie van justitiële (beleids)interventies • Van Gils (2014) *Real time evaluation* • Overige wetenschappelijke bronnen	Openbaar Intern Openbaar
VWS	• Programma Evaluatie Regelgeving (PER, 1997). Ontwikkeld door ZonMw. Evaluatie door Olsthoorn-Heim (2003) waarvan de aanbevelingen in vervolgprogramma zijn verwerkt. • Brochure Programma Evaluatie Regelgeving	Openbaar

	(ZonMw, 2013).	

3.5.1 Evaluatieprogramma's specifiek gericht op wetgeving

Wetsevaluaties worden soms 'op afstand' uitgevoerd– zoals via het WODC of ZonMw (Ringeling 1997; Olsthoorn-Heim, 2003). Bij dit laatste worden een aantal voordelen genoemd: behalve de afstandelijkheid is er minder last van 'Haagse haast', de evaluatie gebeurt systematisch, er kunnen vragen worden gesteld die beleidsmakers liever niet in discussie willen hebben, en het evaluatieproces is openbaar en navolgbaar, bijvoorbeeld omdat onderzoeken als achtergrondstudies worden gepubliceerd.
In wat volgt beschrijven we twee programma's ter stimulering van *wets*evaluatie: het Programma Evaluatie Regelgeving (PER) bij ZonMw voor VWS en het Programma Structurele Evaluatie Milieuwetgeving (STEM) bij het voormalige VROM. STEM is in 2010 afgerond. Vervolgens beschrijven we een derde 'programma': ervaringen met de wettelijk verplichte Awb-evaluaties, vooral met de derde.[111]

Het Programma Evaluatie Regelgeving (PER)
ZonMw, het onderzoeksinstituut dat wetsvaluaties in opdracht van het ministerie van VWS uitvoert en uitbesteedt, beheert sinds 1997 het *Programma Evaluatie Regelgeving* (PER), dat inmiddels tientallen wetsevaluaties en sinds 2009 vier thematische wetsevaluaties heeft opgeleverd. Het PER is samengevat in de brochure Evaluatie Regelgeving (2013), te vinden op de website van ZonMw. Het programma streeft uniformiteit na bij de uit te voeren wetsevaluaties om zo bij te dragen aan de kwaliteit van wetgeving. Veelal hebben die wetsevaluaties een verplicht karakter, vanwege een evaluatiebepaling of een toezegging van de minister aan de Tweede Kamer.
De brochure is onder meer bedoeld als richtlijn voor de potentiële aanvragers en uitvoerders van evaluaties en vermeldt de algemene criteria die de Commissie evaluatie regelgeving (CER) hanteert bij de selectie van de evaluatie-uitvoerders en hun voorstellen. Relevantie en kwaliteit zijn daarbij sleutelwoorden (interview, 8 april). Het PER beschouwt wetsevaluaties als een combinatie van juridisch en sociaalwetenschappelijk onderzoek; het is het enige programma dat dit punt expliciet noemt.
In 2003 is het PER geëvalueerd – vijf jaar na oprichting en acht wetsevaluaties verder (Olsthoorn-Heim, 2003), zie Box 4. De grote verschillen tussen wetsevaluaties vielen daarbij op, zowel in termen van kwaliteit en doorwerking als in termen van budgettoekenning en proces – waarin ook knelpunten werden gesignaleerd. De politieke context – per wet weer anders – lijkt een oorzaak voor de gevonden verschillen.

Box 4 Conclusies en aanbevelingen evaluatie eerste vijf jaar Programma Evaluatie Regelgeving (PER, 1998-2003)

Algemene conclusie in 2003: 'Het programma Evaluatie Regelgeving heeft in de eerste vijf jaar ten dele voldaan. Het vormt een kader waarbinnen geregeld en bijna altijd tijdig wetsevaluaties tot stand worden gebracht, op afstand van het departement, waarover de opdrachtgever meestal redelijk tevreden is. Het programma is ondanks de systematische aanpak niet geslaagd in het streven om uniformiteit en onderlinge *vergelijkbaarheid* van de wetsevaluaties te bereiken.'

[111] Er zijn zeker meer voorbeelden te geven, die hier niet zijn uitgewerkt. Zo noemde Olsthoorn-Heim (2003) de Commissie van advies inzake evaluatie van de Bopz (Wet bijzondere opnemingen in psychiatrische ziekenhuizen), ingesteld door de ministers van VWS en Justitie. Deze besteedde negen onderzoeken uit voor de eerste evaluatie, die elk zijn begeleid door subcommissies uit de evaluatiecommissie.

Belangrijkste aanbevelingen

- Heldere opdrachtverlening met de juiste deskundigheid en in de juiste infrastructuur
- Vroeg beginnen met het voorbereiden van een wetsevaluatie op het departement; bijvoorbeeld aan de hand van een nulmeting, vroeg experimenteren en ook het overwegen van thematische (wetsoverstijgende) evaluatie. Tegelijkertijd: geduld betrachten bij het plannen van een evaluatie van doelbereiking of doeltreffendheid – de werking moet zich eerst kunnen uitkristalliseren.
- Voor meer uniformiteit/vergelijkbaarheid: het beter bewaken van de (nieuwe) programmatekst met onderscheid tussen eerste en volgende evaluaties en nieuwe en bestaande wetten; en het door de BC laten toetsen van wetsevaluaties aan de programmatekst.
- Kansen en mogelijkheden voor de thematisering van wetsevaluaties
- Het uitbreiden van expertises in de CER van gezondheidsrecht en - zorg, met expertise op het vlak van wetgeving, bestuurskunde en rechtssociologie; en 'buiten' ook het werven van de verschillende disciplines voor het indienen van projectvoorstellen.
- Meer uitgewerkte eisen aan de aanpak of onderzoeksmethode – zonder de onafhankelijkheid en creativiteit van onderzoekers te bedreigen, zoals: schetsen van theoretisch kader, vaststellen van beoordelingsmaatstaven, heldere vraagstelling, verantwoord evaluatiedesign, voldoen aan vereisten van validiteit/betrouwbaarheid/consistentie.
- Bijeenkomst ter afsluiting – met samenvatting op de website – ten behoeve van bredere verspreiding en eventueel publicatieplan van betrokken onderzoekers.

De aanbevelingen voortvloeiend uit deze evaluatie zijn na de evaluatie in het nieuwe PER verwerkt.[112] Het PER vat procesmatige en inhoudelijke criteria voor evaluatie van wet- en regelgeving samen in de volgende punten:
(1) wetsevaluaties moeten door een multidisciplinaire groep worden uitgevoerd met aandacht voor zowel het empirische als voor het juridische deel;
(2) kennisnemen van de beginsituatie – zo mogelijk een nulmeting – moet onderdeel zijn van de evaluatie[113];
(3) wetsevaluatie beperkt zich niet tot de te evalueren wet, maar houdt ook rekening met de internationale en Europese context;
(4) wetsevaluaties kijken zowel naar gewenste als ook ongewenste effecten;
(5) wetsevaluaties moeten aandacht schenken aan de relatie tussen wetgeving en zelfregulering – een criterium verbonden aan het subsidiariteitsbeginsel;
(6) elke nieuwe evaluatie moet naast nieuw onderzoek ook gebruik maken van bestaande kennis, ter voorkoming van verspilling van tijd en geld; en tot slot
(7) wetsevaluaties besteden aandacht aan diversiteit en aan de toegankelijkheid van alle bevolkingsgroepen tot de wet (ZonMw, 2013, p. 4-5).

Op de website van ZonMw verschijnt een oproep tot het indienen van subsidie-aanvragen. Deze aankondiging bevat de project- of thematekst en de deadline voor het indienen van een aanvraag.[114] Bij de selectie van het winnende project hanteert de CER een aantal *wetenschappelijke* kwaliteitscriteria. Daaronder vallen de helderheid, originaliteit en haalbaarheid van de vraagstelling, de aansluiting van het plan van aanpak op de vraagstelling, de theoretische en de empirische onderbouwing. In het licht van de evaluatiekwaliteit zijn daarnaast van belang: de kennis en ervaring van de indieners, de haalbaarheid van het evaluatieproject en de begroting. Deze criteria zijn voor potentiële indieners toegankelijk en dienen als richtlijnen bij het schrijven van offertes (ZonMw, 2013; interview 9 april).

[112] Zo meldde de toenmalig minister Hoogervorst van VWS in een brief aan de Kamer (TK, 2003–2004, 29 200 XVI, nr. 281).
[113] Hiertoe wordt ook het expliciteren van de beleidstheorie gerekend (interview 8 april 2015).
[114] De 'project- of thematekst' is de goedgekeurde vertaling van de opdracht van de opdrachtgever aan ZonMw. Deze tekst vormt voor aanvragers het kader waarbinnen zij een aanvraag kunnen indienen. Indien er een specifieke aanleiding bestaat, kan voor een zg. *top-down* procedure worden gekozen. Hierbij worden volgens de 'procedures ZonMw' potentiële aanvragers van wetsevaluatieonderzoek rechtstreeks uitgenodigd om een aanvraag in te dienen.

Aanvragen voor ZonMw-subsidies doorlopen de volgende stappen:
– de wetenschappelijke kwaliteit van de aanvraag wordt door minimaal twee onafhankelijke externe referenten beoordeeld;
– de aanvrager krijgt de gelegenheid schriftelijk te reageren op de geanonimiseerde referentenoordelen;
– de CER geeft op basis van de aanvragen, de referentenoordelen en de reactie daarop van de aanvragers een advies over de relevantie, de wetenschappelijke kwaliteit en de prioritering van de aanvragen;
– op basis van dit advies neemt het ZonMw-bestuur een besluit over de toekenning van het wetsevaluatieonderzoek.

Voor een enkelvoudige wetsevaluatie wordt uitgegaan van één jaar looptijd; voor een thematische wetsevaluatie van 18 maanden. Voor de evaluatiekwaliteit is gedurende de looptijd naast de PER-selectieprocedure ook de rol van de BC, die altijd wordt ingesteld, van groot belang (zie verder par. 3.6).
De ervaringen met het STEM-programma worden weergegeven in Box 5.

Box 5 Ervaringen met het programma Structurele Evaluatie Milieuwetgeving (STEM)

Het programma STEM is de opvolger van de ECWM, de Evaluatiecommissie Wet Milieubeheer (2000-2003) en haar voorganger sinds 1980. Het onderzoeksprogramma STEM (eerst 2005-2008 en daarna het aanvullend programma over 2009-2010) stelde zich – naast het voldoen aan de wettelijke evaluatieverplichting in artikel 21.2 lid 1 Wet Milieubeheer (Wm) – ten doel: 'Het uitvoeren van evaluatieonderzoek, waarbij de onderzoeksresultaten:

- bijdragen aan kennis inzake de regulering op het milieubeleidsterrein, dat wil zeggen inzake de mogelijkheden om via wetgeving het milieu in brede zin te beschermen.
- gericht zijn op de verbetering van de kwaliteit van wetgeving in brede zin, waaronder in ieder geval worden begrepen vraagstukken van subsidiariteit (is regelgeving (in de gegeven vorm) wenselijk), de doeltreffendheid, efficiëntie, uitvoerbaarheid en handhaafbaarheid van de regelgeving, de waarborgen voor burgers (kenbaarheid, inspraak, rechtsbescherming) en de rechtmatigheid.'

Hierbij is een aantal uitgangspunten gehanteerd die zowel door hebben gewerkt in de keuze van onderwerpen van evaluatie als in de onderzoeksvragen en -aanpak. Die uitgangspunten waren:

- Een deel van de beschikbare onderzoektijd en onderzoeksbudget zou gericht worden op actuele beleidsvragen
- Waar relevant zou rekening gehouden worden met internationale en Europese regelgeving en beleid
- Bij de onderzoeken zou steeds ook aandacht worden besteed aan de doorwerking van onderzoeksresultaten naar praktijk en politiek.

De keuze van onderwerpen is bepaald door de agenda van de beleidsmakers en wetgever, maar óók door de uitgangspunten van de STEM-onderzoeksprogramma's. Dat gebeurde in samenspraak tussen de opdrachtgever, de onderzoekers en een Adviesraad, waarin alle doelgroepen van milieubeleid vertegenwoordigd waren. De beslissingen zijn genomen door een Programmaraad, waarin de onderzoeksinstellingen van STEM en het voormalige ministerie van VROM (thans IenM) gelijkelijk vertegenwoordigd waren. Opmerkelijk is dat in de loop van het programma wellicht als gevolg van een integratie van afdelingen binnen het ministerie van VROM en gelijk opgaand met de totstandkoming van de Wabo, steeds vaker onderwerpen werden gekozen waarin over de grens van het milieu ook naar aspecten van ruimtelijke ordening, water en natuur werd gekeken.
De onderzoeken werden – in verschillende fasen van uitvoering – twee keer per jaar met de Adviesraad, enkele keren in combinatie met de Werkgroep Juridisch Instrumentarium van het ministerie, besproken. Deze bijeenkomsten leidden tot ideeën en inzichten inzake de onderzoeksthema's die zeer waardevol waren.

Sinds 2005 zijn bij STEM 35 studies opgeleverd, die zijn te downloaden op www.evaluatiemilieuwetgeving.nl. Behalve om ex post evaluatieonderzoeken gaat het ook om ex ante studies.

Lessen uit de praktijk

Een aantal ervaren onderzoekers (Uylenburg et al., 2011) blikt terug op het STEM-programma, zowel op de keuze van onderwerpen als op de inhoud en aanpak van de resulterende rapportages, en trekt hier de volgende lessen uit.

(1) Een koppeling tussen de wet en empirisch onderzoek ontbreekt vaak: doeltreffendheid in juridische zin is iets anders dan in de empirische betekenis;

De Boer komt terugblikkend tot de conclusie dat de functie van het evaluatieonderzoek sterker gericht was op het geven van inzicht in het onderwerp, dan strikt op het bepalen van de effecten van een regeling. Uylenburg constateert dat lang niet altijd is ingegaan op de relatie tussen de doeltreffendheid van de wet in de juridische betekenis en het daadwerkelijk bereiken van milieudoelen – en mist daarmee een empirische component. 'Onderzocht is of de regelingen de noodzakelijke voorzieningen bevatten, en de wijze van uitvoering van die voorzieningen op een zodanige wijze plaatsvindt, dat het doel kán worden bereikt.'

(2) Het blijkt lastig om bij een wetsevaluatie aan alle criteria uit de Ar (tegelijk) recht te doen[115]

Oosterhuis merkt in dit verband op: '(....) [E]en standaard-set van te hanteren [evaluatie]criteria is in een zo divers beleidsterrein een illusie en ook niet zinvol. Daarentegen kan het wel de moeite waard zijn om een checklist van potentieel relevante criteria te hebben om na te gaan welke daarvan in een concrete evaluatie relevant zijn. Dat vermindert de kans dat bepaalde aspecten over het hoofd worden gezien.'

(3) Een meer systematische en multidisciplinaire aanpak had gevolgd kunnen worden; niet alleen gericht op actuele vraagstukken.

Oosterhuis bepleit een – ook ten opzichte van STEM – meer structurele benadering van de evaluatie van wetgeving. Evaluaties zouden bijvoorbeeld niet beperkt moeten worden tot die onderwerpen waarover belangengroepen of politici zich het meest opwinden; bij een systematische evaluatiebenadering kan het leereffect van de evaluatieonderzoeken worden vergroot. Een multidisciplinair onderzoeksteam past bij een meer systematische aanpak. Daarbij geldt, evalueer met mate: '(...) Een hoge frequentie van evaluaties en daaruit voortvloeiende aanpassingen staat welhaast garant voor verminderde doeltreffendheid op langere termijn. Het Nederlandse stimuleringsbeleid voor duurzame energie is een bekend voorbeeld.'

(4) Ten einde een effectieve doorwerking te bereiken is 'aansluiting bij de agenda van de wetgevers van het ministerie' de belangrijkste factor voor succes.

Met betrekking tot het evaluatie-gebruik in het STEM-programma is Uylenburg van mening dat 'de doorwerking van de onderzoeken zeer zou verbeteren wanneer het (collectieve) geheugen van de wetgevende ambtenaren groter zou zijn. Veel van de uitgevoerde onderzoeken behouden immers hun waarde.'

Bron: Uylenburg et al. (2011).

Behalve het STEM-programma zijn ook de opeenvolgende evaluaties van de Algemene wet bestuursrecht (Awb) een voorbeeld van een rond één (majeure) wet georganiseerd (tijdelijk) evaluatievermogen (zie Box 6). Dat komt voort uit de tot 1 januari 2013 in artikel 11:1 lid 2 Awb opgenomen verplichting die de ministers van Justitie en BZK opdroeg elke vijf jaar 'een verslag over de wijze waarop zij is toegepast' naar de Staten-Generaal te zenden (Michiels 2010, p. 42).

---

[115] Rechtmatigheid, doeltreffendheid en doelmatigheid, subsidiariteit en evenredigheid, uitvoerbaarheid en handhaafbaarheid, en de eenvoud, duidelijkheid en toegankelijkheid van de regeling.

Doeltreffendheid of doelbereiking maken niet expliciet onderdeel uit van de evaluatiebepaling.

Box 6 De evaluatie van de Awb

De Awb is inmiddels vier keer geëvalueerd naar aanleiding van de evaluatiebepaling, waarbij de eerste drie keer steeds een evaluatiecommissie is ingesteld. De vragen in de uitgevoerde onderzoeken en de door de drie opeenvolgende evaluatiecommissies uitgevoerde adviezen betroffen specifieke onderdelen van de Awb, verkaveld naar min of meer afgebakende onderdelen van de wet. Daarnaast zijn een paar algemene invalshoeken aan de orde geweest, zoals de manier waarop de burger de Awb ervaart en gebruikt. In een kritische terugblik bespreekt Michiels een aantal procedurele en inhoudelijke aspecten van drie rondes Awb-evaluaties (1996, 2001 en 2006) die in de praktijk hebben plaatsgevonden en samen 3.000 pagina's aan onderzoeksrapporten hebben opgeleverd, en 250 pagina's aam evaluatiecommissierapporten.

Evaluatieproces
Per evaluatieronde is bij de eerste drie evaluaties steeds een evaluatiecommissie ingesteld, bestaande uit personen afkomstig uit het openbaar bestuur, de rechterlijke macht, de advocatuur en de wetenschap. Het secretariaat was in handen van enkele ambtenaren van de ministeries van Justitie en BZK (Michiels, 2010; interview 11 augustus). De departementale inbreng in de bepaling van evaluatieonderwerpen was groot. In de derde evaluatie stonden drie onderwerpen bij voorbaat al vast, zijn er twee kort voor aanvang gekozen en zijn daarnaast twee onderwerpen op basis van suggesties van praktijk en wetenschap meegenomen (ibid., Tweede Kamer, 2009–2010, 29 279, nr. 111, p.2).
Vervolgens is voor elk onderwerp aan enkele universiteiten (groepen wetenschappers) gevraagd een offerte uit te brengen, waarna de gunning volgde. Behalve inhoudelijke kwaliteiten is daarbij ook naar de spreiding over universiteiten gekeken. Elke (deel)evaluatie is begeleid door een BC. De resulterende rapporten werden ingediend bij een evaluatiecommissie. De evaluatiecommissie fungeerde als tussenschakel, filter en doorgeefluik, niet als initiator, coördinator, of aanjager. Ze vatte de bevindingen samen, gaf commentaar, nam sommige aanbevelingen wel over en andere niet en adviseerde de wetgever waar nodig om de Awb aan te passen (Michiels, 2010, p. 45). Uiteindelijk diende de evaluatiecommissie te beslissen welke aanbevelingen ze deed, waarbij ze, anders dan de onderzoekers die één stukje Awb evalueerden, het overzicht over alle resultaten had, wat ertoe kon leiden dat een aanbeveling die op het eerste gezicht verstandig leek toch niet werd overgenomen (p. 49). Ook trokken de evaluatiecommissies vaak een eigen lijn en bleken niet alle aanbevelingen terug te voeren op de onderliggende rapporten (p.49-51); Michiels acht in die gevallen nadere onderbouwing door de evaluatiecommissie extra relevant.
Vergelijkbare tijdelijke constructies worden vaker opgetuigd rond grotere, vaak politiek gevoelige wetsevaluaties (interview 8 juni), zoals de recent opgeleverde evaluaties van de nieuwe staatkundige structuur Caribisch Nederland en van de ook in 2010 in werking getreden justitiële consensus-rijkswetten (Evaluatiecommissies Spies en Brouwer, beide 2015).[116] Er bestaat volgens ingewijden geen overkoepelende standaard of richtlijn voor de inrichting en rol van dergelijke evaluatiecommissies (interview 8 juni).

Evaluatieproces: lessen

[116] Zie www.evaluatiecaribischnederland.nl/commissie/. En www.gracao.com/index.php/ouder/12943-werkzaamheden-commissie-evaluatie-justitiele-consensusrijkswetten-op-schema . In de aanloop naar deze evaluaties heeft de Raad van State op verzoek van de Minister van Veiligheid en Justitie een advies uitgebracht met voorstellen voor evaluatiecriteria en -vragen (2014). Advies of bijstand verlenen bij wetsevaluaties behoort echter niet tot het standaard takenpakket van de RvS (communicatie d.d. 23 november 2015).

Anders dan in geval van de ECWM en het latere STEM (programmaraad) hebben vooral de verantwoordelijke ministeries de topics van de Awb-evaluatie bepaald. De vraag is of dit tot voldoende objectieve keuzes heeft geleid; Michiels stelt de vraag of de sterke rol van het ministerie tot het juridische karakter en de sterke focus op rechtsbescherming heeft geleid.

- Er is lang gewacht met het voorbereiden van de eerstvolgende evaluatieronde, waardoor de evaluaties een (te) korte doorlooptijd hebben gehad (vgl. interview 11 augustus). Ook de afweging ten aanzien van de vaststelling van evaluatieonderwerpen is volgens Michiels (2010, p. 45-46) niet voldoende voldragen geweest. De deelonderzoeken, hoe adequaat ook, zijn volgens hem hierdoor afzonderlijke partjes binnen het grote geheel geworden, met weinig onderlinge dwarsverbanden. Soms bestaat zelfs overlap in onderwerpskeuzes die door verschillende onderzoeksgroepen zijn behandeld.
- Voor een betere communicatie tussen onderzoeksgroepen en evaluatiecommissie – want er is vanuit de onderzoeksgroepen wel kritiek geleverd op de evaluatiecommissies – acht Michiels het nuttig als eerst een conceptadvies wordt besproken.

Inhoud: lessen

- De praktijkervaringen met de eerste drie evaluaties van de Awb wijzen op het belang van een multidisciplinaire, 'op maat'-benadering – net als de eerder besproken ervaringen met het programma STEM. Niet-juridische invalshoeken, zoals de rechtseconomische, zijn tot dusverre nauwelijks aan bod gekomen; het betrof beleidsgerichte, juridische wetsevaluatie (Michiels, 2010, p. 44; vgl. Eijlander 2003 en interview, 11 augustus). Zo schreef Eijlander naar aanleiding van de tweede Awb-evaluatie: 'De beleidstheorie en de vooronderstellingen die ten grondslag liggen aan de Awb zijn niet ter discussie gesteld. De evaluatie heeft zich gericht op de toepassing en de werking van de Awb in de praktijk en de mogelijkheden om daar eventueel verandering in te brengen. (….) Er heeft ook geen expliciet onderzoek plaatsgevonden naar de vraag of de doelstellingen van de Awb in de praktijk daadwerkelijk zijn of worden gerealiseerd (…) bijvoorbeeld, of met de Awb de bestuurswetgeving inderdaad uniformer en eenvoudiger is geworden. (…) Nu is dat ook complex (…) Een zg. *goal free evaluation* is dan beter uitvoerbaar.'
- Een verwant punt betreft de *maatstaven* waaraan met de evaluatie wordt getoetst. Systematische evaluatie van wetgeving houdt in 'het ex post beoordelen van de in empirisch onderzoek waargenomen inhoud, uitvoering en effecten van een wet aan de hand van bepaalde criteria' (Winter, 1996). Evaluatiecriteria of – maatstaven zijn echter in geval van de Awb-evaluaties bij voorbaat niet expliciet gehanteerd. Volgens Michiels zijn gezaghebbende evaluatiemaatstaven nodig, die hetzij door wetgever zelf, hetzij door de regering in overeenstemming met het parlement, zijn geformuleerd (p.48). *Wie* de maatstaven definieert en welke dat dan zijn, is een lastig vraagstuk – en ook veel bediscussieerd in de literatuur over de Awb. Zo is in de literatuur wel als maatstaf genoemd: 'bestuursproces van voldoende kwaliteit' – maar de vraag is hoe men kwaliteit dan weer definieert. Hier raakt men aan het verder liggende doel dat de Awb zou dienen: de uitgangspunten van de rechtsstaat? Primair codificatie?, of richting geven aan nieuwe bestuursrechtelijke ontwikkelingen?
- Het accent lag tot nu toe vooral op rechtsbescherming (bezwaar) in plaats van op de primaire besluitvorming. Michiels acht het laatste een minstens zo belangrijk evaluatieonderwerp.

Na de derde evaluatie van de Awb volgde nog een vierde (2007-2011), echter zonder evaluatiecommissie; deze ging over het extern klachtrecht (titel 9.2 Awb) en de bezwaarprocedure (titel 7.2 Awb). Beide onderzoeken zijn in 2011 aan de TK gezonden (Bex et al., 2010; De Waard, red. 2011). Met ingang van 2013 is, met de inwerkingtreding van de Wet aanpassing bestuursprocesrecht (Wab), het artikel 11:1 over de verplichte

vijfjaarlijkse evaluatie van de Awb afgeschaft. Niettemin is een aantal 'kleinere' evaluaties afgerond of lopend, zoals naar een drietal versnellingsinstrumenten[117] (Schueler et al., 2013) en de evaluatie van de Awb-bevoegdheid om gebreken te passeren en het relativiteitsvereiste (Wab).[118] Ook zijn de geldschuldenregeling en de per 2010 ingevoerde 'bestuurlijke lus' inmiddels geëvalueerd (Den Ouden et al., 2013; Backes et al., 2014).

Een overkoepelende les die volgt uit deze evaluatieprogramma's rond wetgeving is het belang van een multidisciplinaire aanpak voor inhoud en kwaliteit. Sociaalwetenschappers besteden vaak niet of nauwelijks aandacht aan het juridische kader (vgl. Klein Haarhuis en Niemeijer 2008). Wat juristen dan weer vaak nalaten, is het expliciteren van de (gedrags)assumpties waarop de wet rust (interview 9 juni). Ook het sociaalwetenschappelijk jargon en de methoden en technieken van sociaalwetenschappelijk onderzoek, zoals werken met steekproeven staan vaak ver van juristen af. Zo staat bijvoorbeeld in veel wettelijke evaluatiebepalingen dat de evaluatie moet concluderen of de interventie de beoogde *effecten* heeft bereikt, maar volgens de EWB-projectbegeleiders (WODC) zijn niet alle juristen vertrouwd met de sociaalwetenschappelijke lading van doeltreffendheid en effecten. 'Juristen stellen de vraag of de wet *geldt,* beleidsdirecties en sociaalwetenschappers onderzoek willen weten of de wet *werkt*, hoe en onder welke omstandigheden' (interview 8 juni). Het voldoen aan de juridische randvoorwaarden voor doeltreffendheid betekent nog niet dat de wet in de praktijk (de empirie) ook doel zal treffen. Alleen in het PER is een multidisciplinaire aanpak van wetsevaluaties voorgeschreven, zoals blijkt uit de aangehaalde zeven criteria.
Naast voornoemde drie (en de Bopz) zijn er, voor zover ons bekend, op dit moment geen andere programma's specifiek ingericht voor wetsevaluatieonderzoek. Over de behoefte daaraan en de noodzaak ervan hebben respondenten verschillende meningen bijvoorbeeld ten aanzien van de toepasbaarheid van experimentele designs (interview 8 juni).[119] Volgens sommigen zijn deze niet geschikt om wetsevaluaties aan te toetsen, omdat er bij wetsevaluaties meestal niet met controlegroep kan worden gewerkt; het onttrekken van een deel van de doelgroep aan de werking van de wet zou het gelijkheidsbeginsel schenden. Anderen delen dit gezichtspunt niet. In dit verband wijzen we op mogelijkheden van experimenteerwetgeving die het – zij het voor specifieke typen wetgeving – mogelijk maakt om niet ex post maar juist in een vroege of ex ante-fase te bezien hoe doelgroepen reageren op de (belangrijkste) wettelijke 'interventies' – afgezet tegen vergelijkbare controlegroepen.
Een respondent van DWJZ bij OCW moedigt aan dat initiatieven rondom specifiek *wets*evaluaties weer nieuw leven wordt ingeblazen (interview 14 juli).

3.5.2 Evaluatieprogramma's algemeen

Zicht op effectiviteit (OCW)
De directie Kennis van het ministerie van OCW runt sinds oktober 2010 het programma *Zicht op effectiviteit*. Het doel van dit programma is 'de kwaliteit van

[117] Het betreft de dwangsom bij niet tijdig beslissen (paragraaf 4.1.3.2 Awb), beroep bij niet tijdig beslissen (afdeling 8.2.4a Awb), en de positieve fictieve beschikking bij niet tijdig beslissen; ook wel lex silencio positivo genoemd (paragraaf 4.1.3.3 Awb).
[118] Awb, artikelen 6:22 en 8:69a. www.wodc.nl/onderzoeksdatabase/2422-evaluatie-artikel-622-awb-bevoegdheid-om-gebreken-te-passeren.aspx?nav=ra&l=staats_en_bestuursrecht&l=administratief_recht_procesrecht. De evaluatie van het nieuwe art. 6:22 Awb is mondeling toegezegd in de Eerste Kamer (Eerste Kamerstukken, Handelingen I 2012/13, nr 12, item 20, 17 december 2012) en die van art. 8:69a wordt voorgeschreven in art. 1a van deel C van de Wab.
[119] Meer precies verwijst men naar de Campbell-standards of *Maryland Scientific Methods Scale* (vgl. Farrington & Welsh, 2005) waarin oplopend vijf niveaus van experimenteel onderzoek worden onderscheiden.

evaluatieonderzoek, dat in opdracht van het ministerie OCW wordt verricht, te verbeteren en de methodologische opzet van dit type onderzoek voor zover mogelijk te harmoniseren' (OCW, 2015a; vgl. interview 1 juli).
De directie Kennis zet zich in voor complementariteit tussen experimenteel onderzoek en verklarende evaluaties (ibid.). In geval van potentieel geëigende beleidsinterventies laat Kennis een onderzoeksdesign opstellen door CPB of SCP. De toon hiervoor is gezet in 2011, toen op verzoek van OCW onderzoeksdesigns voor een negental (clusters van) nieuwe beleidsinterventies op onderwijs- en wetenschapsgebied zijn opgesteld (CPB, 2011). Het betrof zowel verklarende als experimenteel georiënteerde effectevaluaties.

> 'Op verzoek van het ministerie zijn voor alle interventies zowel ontwerpen gemaakt voor een verklarende evaluatie als voor een effectevaluatie. De verklarende evaluatie richt zich op de factoren die het succes of het ontbreken van het succes van de interventie bepalen (zie bijlage voor een nadere toelichting op de verklarende evaluatie). Anders gezegd, wat gebeurt er precies binnen scholen als gevolg van de interventie. Wat is de *Theory of Change* en welke intermediaire uitkomsten bieden hiervoor inzicht? De ontwerpen voor de effectevaluatie richten zich op het zo overtuigend mogelijk vaststellen van de effecten van de nieuwe beleidsinterventies.' (CPB, 2011, p.4).

Beleidsdirecties kunnen gebruik maken van het betreffende onderzoeksdesign bij het opstellen van offerteaanvragen voor evaluaties – maar zijn daartoe niet verplicht (OCW, 2015a). Bijlage 4 bevat een uitwerkingsvoorbeeld (centrale toetsing zwakke scholen) van een gecombineerd verklarende – toetsende opzet van een evaluatie met een juridische component. Een ander mogelijk leerzaam voorbeeld van een doeltreffendheidsevaluatie betreft het ontwerp van de evaluatie van de Lerarenbeurs – ingesteld in 2008 om leraren in staat te stellen een Bachelor- of Masteropleiding te volgen. In het evaluatiedesign is gebruik gemaakt van een discontinuïteit in de kans op toekenning van de beurzen, op basis van volgorde van binnenkomst. 'Rondom het moment waarop het budget uitgeput raakte, zijn in principe gelijke kandidaten de ene keer wel, en de andere keer niet in aanmerking gekomen voor een beurs. Vergelijking van de uitkomsten voor kandidaten geeft vervolgens een effectmeting.' (CPB, 2015, p.5). Dit strookt met internationale tips voor effectevaluaties (Hoofdstuk 2).
Per 2013 is door de Kennisdirectie het in 2010 gestarte programma Zicht op effectiviteit voortgezet, zij het alleen voor beleid dat aan drie criteria voldoet: (1) het biedt een goede kans op een robuuste effectevaluatie, gekoppeld aan een goede *verklarende* evaluatie; (2) het moet tot de kern van het OCW-beleid horen; (3) beleidsdirecties moeten serieuze belangstelling hebben voor deze aanpak (OCW, 2015a). Dit heeft de directie Kennis zo besloten om het specifieke karakter van het kwaliteitsprogramma te onderstrepen. De Directie Kennis is betrokken bij al het quasi-experimentele effectonderzoek; bij de overige evaluaties bepaalt de verantwoordelijke beleidsdirectie of zij directie Kennis betrekt bij het denken over de aanbesteding van offertes en het evaluatiedesign, of als lid van de BC. In en na 2013 zijn door CPB en SCP alweer nieuwe onderzoekdesigns voor OCW opgesteld.[120]
In de praktijk gaan de aanvragers – die beslissen – nog relatief weinig over tot de toepassing van quasi-experimentele designs (interviews 1 juli en 26 augustus). De indruk bij de directie Kennis is dat het programma *Zicht op effectiviteit*, bedoeld voor de stimulering van doeltreffendheidsevaluaties, 'in een omgeving landt waar niet gepland wordt nagedacht over het onderzoek dat gedaan moet worden' (interview 1 juli; vgl. 26 augustus).

[120] www.cpb.nl/publicatie/ontwerpen-voor-effectevaluatie; www.scp.nl/Publicaties/Alle_publicaties/Publicaties_2014/Op_zoek_naar_bewijs_II.

Op het moment van interviewen bleken ongeveer tien doeltreffendheidsevaluaties volgens het vooraf bepaalde design afgerond[121] en tien andere lopend, op verschillende deelgebieden van onderwijs en onderzoek (OCW, 2015a; interview, 1 juli). Deze betreffen niet of nauwelijks evaluaties van wetten in formele zin.[122]
Voor het opzetten van het complementaire *verklarende onderzoek* kunnen beleidsdirecties gebruik maken van een voor de directie Kennis ontwikkelde methodiek van het Kohnstamm-instituut (Pater et al., 2012). 'Dergelijk verklarend evaluatieonderzoek wordt in de evaluatiepraktijk van OCW nog weinig uitgevoerd. Voor het uitvoeren van een effectevaluatie [experiment of quasi-experiment, red.] zijn er duidelijke kwaliteitsstandaarden voor het gekozen onderzoeksdesign en de manier waarop de studie wordt uitgevoerd. Voor een verklarende evaluatie zijn er nog nauwelijks goed hanteerbare kwaliteitsstandaarden. Ook ontbreekt een ruim toepasbare methode, en is er geen wetenschappelijke consensus over hoe goed verklarend onderzoek uitgevoerd moet worden' (Pater et al., 2012, p.1, citaat uit tekst offerteverzoek).
Bij verklarende evaluatie staat niet de vaststelling van een causaal verband, maar het begrijpen dan wel het plausibel maken ervan. Dit is een complexe exercitie (Pater et al., 2012, p.15). Causale verbanden ontstaan doordat de interventie aansluit bij de wensen en mogelijkheden van de doelgroep van het beleid, in omstandigheden die bevorderlijk zijn voor het slagen van de interventie. Het is dus niet de interventie op zichzelf die 'werkt', maar: '*the people co-operating and choosing to make them work'* (vgl. Pawson & Tilley, 1997, p. 36). Het *hoe* is geen variabele, maar een mechanisme en dit mechanisme is bepalend voor het ontstaan van de relatie tussen interventie en uitkomst (Pater et al., 2012, p.13).
Ook de *implementatiefase* maakt volgens het Kohnstamm-instituut onderdeel uit van verklarende evaluatie: hoe uitvoerders en doelgroepen handelen.
Pater et al. dragen onderzoekers een aantal stappen aan (zie Bijlage 4) om tot verklarende evaluatie te komen, en passen deze vervolgens toe op twee cases rond voortijdig schoolverlaten.[123]

Begrotingsevaluatieprogramma sinds 2013 (IenM)
Het kader voor het opstellen van beleidsdoorlichtingen bij IenM is de rijksbreed (door IenM mee-) ontwikkelde handreiking voor beleidsdoorlichtingen en de RPE. De hoofddirectie FMC heeft onder meer met de Hoofddirectie Bestuurlijke en Juridische Zaken (HBJZ) kortgesloten dat in de toekomstige beleidsdoorlichtingen zoveel als mogelijk gebruik wordt gemaakt van de resultaten van het IAK, dat verplicht tot vermelding van de resultaten van ex ante onderzoek naar doeltreffendheid en doelmatigheid van wetgeving en beleid. FMC regisseert de opstelling van de beleidsdoorlichtingenagenda in de begroting. De beleidsdirecties zijn verantwoordelijk voor de (onderliggende) evaluatieprogrammering. Daarnaast stuurt FMC op de verdere ontwikkeling van evaluatieonderzoek via toezeggingen aan de Tweede Kamer (bijvoorbeeld klimaat) en naar aanleiding van het in hoofdstuk 2 aangehaalde Rekenkamerrapport Doeltreffendheidsonderzoek bij de rijksoverheid (AR, 2012, 2013). IenM beoogt, naast het beter programmeren van ex post evaluatieonderzoek, aandacht te trekken voor de noodzaak om begrippen zoals doelmatigheid en doeltreffendheid te expliciteren, liefst al in de ontwerpfase van beleid en regelgeving. Als de doelen van de wet niet expliciet worden omschreven in bijvoorbeeld de Memorie van Toelichting, is het lastiger om achteraf bij het uitvoeren van de ex post evaluatie tot een maatstaf voor de evaluatie van doeltreffendheid te komen. Idealiter worden ex ante, in de Maatschappelijke Kosten-

[121] Enkele zijn: Weekendschool (TIER), Wijkschool, Vernieuwingsimpuls onderzoek en Mbo Taalexperiment.
[122] Centrale toetsing is uiteindelijk niet in de praktijk geëvalueerd; wel andere, zoals de evaluatie van de Lerarenbeurs (Heijnen et al., 2015). Dit betreft echter geen wetsevaluatie.
[123] Deze cases (VM2 en De Wijkschool) zijn in Bijlage 4 niet weergegeven.

batenanalyse (MKBA) en in de MvT de aannames opgenomen ten aanzien van de verwachte bijdrage van het (wets)instrument aan het gestelde doel, binnen benoemde termijnen. In het kader van beleidsdoorlichtingen participeert FMC in de Interdepartementale Begeleidingsgroep Prestatiegegevens en Beleidsevaluatie (IBP), een netwerk van directies FEZ en het Ministerie van Financiën (interviews 16 juli en 1 september; intern document 2013).

3.5.3 Handreikingen, richtlijnen, standaarden

Naast, of in het kader van, genoemde programma's zijn – vooral door de genoemde actieve units bij departementen zoals FEZ en kennisdirecties – handreikingen, richtlijnen en standaarden over evalueren ontwikkeld voor beleidsmedewerkers, zowel wat betreft inhoud als proces.

Handreiking beleidsdoorlichtingen
De Handreiking Beleidsdoorlichting van het ministerie van Financiën (2016) is bedoeld als een departement-overstijgend hulpmiddel bij het opstellen van dit verantwoordingsdocument. De Handreiking gaat zowel in op methodologisch-inhoudelijke als op procesmatige aspecten. Omdat beleidsdoorlichtingen een synthese-document zijn van het onderliggend evaluatieonderzoek, bespreken we deze hier verder niet. Niettemin zijn er veel overeenkomsten met methodologie en proces van 'regulier' evaluatieonderzoek, dus de Handreiking kan ook hiervoor een hulpmiddel zijn.
Een deel van de input voor deze Handreiking is afkomstig van wat individuele departementen eerder ontwikkelden – via het IBP, het interdepartementaal 'FEZ'-overleg. Zo beschikt(e) SZW al over een reeks van eigen gedetailleerde handreikingen en modelbrieven en – formulieren met betrekking tot beleidsdoorlichtingen (interview 15 juli en ontvangen intranet-stukken).
Bij BZK worden de RPE en de binnen het departement ontwikkelde algemene handreiking 'Evaluatie als fundering voor beleid' (Bakker, 2012) meegegeven aan uitvoerders van evaluaties, naast de inmiddels verschenen rijksbrede handreiking voor beleidsdoorlichtingen (interview 13 oktober).

Handreikingen EZ
Het ministerie van EZ heeft in 2012-2013 twee gidsen voor het doen van evaluatieonderzoek opgesteld. De eerste, 'Handreikingenboekje beleidskwaliteit en beleidsevaluatie', dekt de gehele beleidscyclus en is bedoeld – uiteindelijk – voor departement-breed gebruik (BEC-secretariaat, z.j.).

> 'Met enige regelmaat wordt het secretariaat van de BEC benaderd door beleidsmakers met een verzoek om hulp bij het uitdenken van de onderbouwing van voorgenomen beleid. Het gaat dan om de analyse van het beleidsprobleem en de voorgenomen beleidsinterventie en het systeem van nulmetingen, monitoring en evaluaties. Deze gesprekken maken duidelijk dat beleidsmakers op zoek zijn naar basiskennis over beleidskwaliteit en beleidsevaluatie. Het gaat dan vaak om aan de hand van een aantal concrete stappen en met praktische tips op weg te worden geholpen. Dit boekje met handreikingen voorziet in deze behoefte.' (p.1).

Het boekje bevat concrete suggesties voor evaluatieopstellers in elke fase van 'de cyclus van beleidskwaliteit en beleidsevaluatie': probleemanalyse, instrumentkeuze, gevolgenbeoordeling, politieke beslissing, uitvoering en ex post evaluatie. Voor elke fase bevat het een zelfstandig leesbare handreiking. Bij ex post evaluaties gaat het volgens respondenten om basisprincipes (i.c. van de experimentele benadering) en 'tips & tricks' voor aanvragers om deze in staat te stellen het gesprek met de (kwantitatief) onderzoeker aan te gaan en kritische vragen te stellen (9 juli). Behalve over ex post evaluatie biedt het boekje ook aanwijzingen en tips bij het opstellen van (a) vroege beleidsexperimenten en (b) het opzetten van een

monitoringsysteem, met links naar EZ-voorbeelden (Monitor Bedrijvenbeleid en Monitor Green Deals) – inclusief nulmeting en de monitoring van het implementatieproces. Deze elementen staan eveneens in verband met de kwaliteit van latere evaluaties.
Het voorbeeld van de Wet bevordering speur- en ontwikkelingswerk (WBSO) laat zien wat het oplevert als men zich met een wetsevaluatie alleen richt op de belangrijkste 'interventies', bijvoorbeeld in financieel of principieel opzicht. De opeenvolgende WBSO-evaluaties (o.a. EIM, 2012) hebben zich toegespitst op het fiscale instrumentarium, wat gefocuste en geavanceerde evaluatiedesigns mogelijk heeft gemaakt (zie bijlage 4). Uiteraard moet de structuur van de wet in kwestie zich hiervoor lenen.
De tweede handreiking, van FEZ-EZ, behandelt vooral het *proces* van aanbesteding van doeltreffendheidsevaluaties vanuit het perspectief van de aanvrager. Op hoofdlijnen gaat het echter ook over het evaluatiedesign, met enkele verwijzingen naar *good practices*.[124] De handreiking staat in de sleutel van verantwoording.[125]
Voor de opeenvolgende stappen in het evaluatieproces zijn checklists ontwikkeld voor de aanvrager: 1) de evaluatieopzet – voornamelijk gericht op experimentele toetsing van doeltreffendheid; 2) een format voor de offerte-aanvraag; 3) een beoordelingsmatrix voor de vergelijking van offertes; 4) een checklist voor de beoordeling van het uiteindelijke evaluatierapport; 5) een formulier waarmee het evaluatieproject kan worden geëvalueerd. De eerste en de vierde checklist zijn afgedrukt in Bijlage 4.

Europese sturing op evaluatie
Tot slot komt een deel van het evaluatievermogen 'van bovenaf'. De evaluatie van het ESF, primair gericht op het ondersteunen van het scheppen van meer en betere banen, is formeel in handen van het Agentschap SZW.[126] Het evaluatieplan ESF 2014-2020 (EC, 2015d) beschrijft de plannen van het Agentschap inzake monitoring en evaluaties.[127] Zo wordt elke evaluatie bezien door het toezichtcomité vanuit de EC en wordt altijd een selectiecommissie en BC ingesteld. Het evaluatieplan laat verder zien hoe programmering, procesinrichting en kwaliteitsmanagement vorm krijgen. Dat laatste gebeurt bijvoorbeeld via de selectie van uitvoerders op vooraf bepaalde criteria en het gebruik van geaccrediteerde CBS-data.

Methoden en tips van externe onderzoekers voor departementen
De Vereniging voor Beleidsonderzoek (VBO) ontwikkelde 12 concrete tips voor beleidsmedewerkers met betrekking tot de aanbesteding van evaluatie- en ander onderzoek. Er zijn voorts twee met de VBO verbonden publicaties verschenen met uiteenlopende voorbeelden uit de praktijk voor beleidsonderzoekers, deels ook relevant voor evaluatieaanpakken (o.a. Van Hoesel et al., 2015).

3.5.4 Samenvattend over evaluatiekader

---

[124] Het betreft o.a. de opzet van de evaluatie subsidieregeling jonge landbouwers (Bureau Bartels, 2011) bij wijze van denkrichting. Hier zijn drie groepen geënquêteerd die qua samenstelling op elkaar lijken, maar een experimentele en twee natuurlijke 'controlegroepen' vormen: respectievelijk subsidieontvangers, afgewezenen en niet-inners.
[125] Zo zijn de aanwijzingen onder 'programmering' uiteindelijk gericht op een dekkende beleidsdoorlichting (vgl. par. 3.3).
[126] De Structuurfondsen (naast ESF het EFRO en het Cohesiefonds) zijn herverdelende financiële instrumenten die de sociale samenhang en economische welvaart binnen Europa ondersteunen door uitgaven te concentreren op minder ontwikkelde regio. De lidstaten hebben evaluatieverplichtingen met betrekking tot de fondsen (EC No. 1083/2006 artikel 48[3]). De evaluaties kunnen uitmonden in wijzigingen van de operationele programma's (art. 33[1]).
[127] 'Het plan voldoet aan de Verordeningen van de EC en volgt in haar structuur en inhoud de *guidelines* van de EC. Het is opgesteld op basis van de beschreven richtlijnen in het *Guidance Document on Evaluation Plans'* (EC, 2015a).

Recapitulatie van deze paragraaf is niet eenvoudig; voor een overzicht van programma's en handreikingen of richtlijnen verwijzen we nog eens naar Tabel 8 en voor concrete hulpbronnen naar Bijlage 4. Wel zijn een paar grove lijnen te trekken. De inhoudelijke evaluatiefocus verschilt per departement en past vaak bij de rol van de op het evaluatievlak meest actieve organisatieonderdelen. Is FEZ zeer actief, dan weerspiegelt zich dat in een meer summatieve, op verantwoording gerichte aanpak (bijv. bij EZ en SZW). Bij EZ werkt men bijvoorbeeld met uitleg en checklists voor evaluatieopzetten (procesmatig en inhoudelijk), offertes en offertevergelijking en rapporten. De ervaring met de WBSO (EZ) leert bijvoorbeeld hoe een (gedeeltelijke) *wets*evaluatie ook kwantitatief kan worden opgezet; het fiscale instrumentarium keert als element in meerdere wiskundige formules terug. Is een directie Kennis of een onafhankelijk onderzoeksinstituut relatief actief, dan uit dat zich vaak in verschillende soorten voorgestane aanpakken, zowel experimenteel of black box als verklarend (o.a. VenJ, OCW). Het onderscheid is dus niet zwartwit; ook respondenten bij financiële directies (o.a. IenM en EZ) en bij het ministerie van Financiën hameren steeds vaker op het belang van toetsing van aannames in de beleidstheorie.
Bij EZ zijn evaluatie-handvatten geïntegreerd in een handreiking voor *beleids*kwaliteit – het overkoepelende doel is effectiever beleid te bewerkstelligen, en evalueren is één van de middelen. Deze handreiking beslaat alle fasen van de beleidscyclus, van probleemanalyse en instrumentkeuze en ex ante gevolgenbeoordeling, tot en met politieke beslissing, uitvoering en – uiteindelijk – ex post evaluatie. Daarnaast bevat de gids aanwijzingen en tips bij het opstellen van (a) vroege beleidsexperimenten en (b) het opzetten van een monitoringsysteem, met concrete voorbeelden.
Het IAK wordt nauwelijks genoemd in het verband van ex post evaluatie; bij IenM probeert men bij ex post beleidsdoorlichtingen terug te grijpen op de resultaten van het ex ante te gebruiken IAK voor beleid en wetgeving.
Veel van de genoemde programma's en richtlijnen of handreikingen zijn nog pril. Er is beperkt zicht op hoe vaak inzichten worden toegepast en dus ook op wat er vervolgens in de praktijk mee is bereikt. We kunnen dus niet zeggen of met de geschetste initiatieven de beleidsdoorlichtingen bijvoorbeeld minder lacunes zijn gaan bevatten, of dat de kwaliteit van evaluatieonderzoek is verstevigd.
Juist ten aanzien van wetgeving zijn enkele programma's besproken die tastbare ervaringslessen hebben opgeleverd. Wat de inhoud betreft leert de praktijk van zowel de tweede en derde Awb-evaluatie, het STEM- en het Programma evaluatie regelgeving (PER) bij ZonMw dat multidisciplinariteit in de aanpak onontbeerlijk is: een combinatie tussen juridisch onderzoek naar gelding en empirisch onderzoek naar o.a. doelbereiking en doeltreffendheid. Bij ZonMw is men tot zeven algemene aanpakcriteria gekomen – die in de praktijk ook al langere tijd worden toegepast.

3.6 Borging van evaluatiekennis en -ervaring

Eerder is al besproken hoe, via de verschillende organisatieonderdelen, procesmatige en methodologische evaluatiekennis de aanvrager zou moeten bereiken. Evaluatievermogen gaat daarnaast ook over: de mate waarin en wijze waarop het evaluatiekader binnen de organisatie wordt gecommuniceerd, hoe men voor een concreet evaluatieproject de benodigde kennis bijeen krijgt, en hoe eenmaal opgedane evaluatiekennis en –ervaring binnen de departementale organisatie wordt verankerd en gedeeld.[128]

[128] We beperken ons hier tot de inspanningen van de (departementale) *organisaties.* De veelal externe onderzoekers komen alleen aan bod waar het gaat om de manier waarop zij door aanvragers voor evaluatieprojecten worden geselecteerd.

3.6.1 Communicatie van het evaluatiekader en cursussen

Een evaluatieprogramma of hulpmiddel bij evalueren, hoe adequaat van opzet ook, zal niet of nauwelijks worden opgepikt als het niet naar aanvragers en betrokken onderzoekers wordt gecommuniceerd.

Cursussen
Het ministerie van EZ heeft in samenwerking met de speciaal aangestelde evaluatie-hoogleraar (VU) een evaluatietraining voor beleidsmedewerkers laten ontwikkelen (Box 7). Die is tot op heden drie keer verzorgd op het departement en is in het najaar van 2015 in aangepaste vorm overgeheveld naar de Rijksacademie voor Financiën, Economie en Bedrijfsvoering ('Beleid in perspectief'- deel Evaluatie).[129]
De cursus biedt de beleidsmedewerkers een eerste kennismaking met kwalitatieve maar ook kwantitatieve evaluatiemethodes, omdat 'voor negen van de tien gewone beleidsmedewerkers de evaluatiemethoden toch lastig zijn' (interview 9 juli).

De training bij EZ kende de volgende doelen[130]

- Kennis over het bestaansrecht van evaluatie en het groeiend maatschappelijk belang ervan
- Inzicht in de samenhang tussen het evaluatieproces en de verschillende fasen van de beleidscyclus (zodat beleidsmedewerkers opgedane kennis kunnen doen gelden in verschillende stadia van beleidsontwikkeling)
- Kennis over methoden en technieken voor beleidsevaluaties (opbouwen van methodologische kennis voor tegenwicht bij uitbesteden aan bureau)
- Kennis over rollen van betrokkenen (beleidsdirecties, FEZ, onderzoeksbureau)
- Kennis over het kwalitatief goed uitvoeren van een evaluatie (praktisch, best practices, oefenen):
    - Hoe richt je beleid zo in dat een beleidsevaluatie straks mogelijk is?
    - Hoe start je een beleidsevaluatie op (Plan van Aanpak, tijdpad)?
    - Hoe zet je een beleidsevaluatie op en waar let je op bij de uitvoering en bij uitbesteding?
    - Hoe beoordeel je de uitkomsten en hoe ga je er mee aan de slag?

- Praktische ervaring opdoen met het uitvoeren van belangrijke onderdelen van evaluaties

Bij EZ zijn daarnaast andere manieren ingezet om de kennis van beleidsambtenaren op peil te houden. Zo zet de voormalige Regiegroep M&E prikkels in zoals een 'quick response team' en een nieuwsbrief waarmee mensen geattendeerd worden op nieuwe *tools* en publicaties. Men is zelf van mening dat dit soort initiatieven geïntensiveerd zouden kunnen worden (interview 9 juli)
Bij OCW biedt het *Top Institute for Evidence Based Education Research* (TIER, 2011) beleidsmedewerkers een Masterprogramma *Evidence based policy and evaluation* aan – met name bedoeld voor onderzoekscoördinatoren.[131] Het programma bestaat uit vier modules; beleidsmedewerkers niet zijnde coördinator volgen meestal één module. De directie Kennis speelt ook een rol bij de opleiding van

---

[129] www.rijksacademie.nl/opleiding/bip/. De cursus is hiermee toegankelijk geworden voor beleidsmedewerkers van ook andere departementen dan EZ.
[130] In de nieuwe vorm bij de Rijksacademie zijn deze leerdoelen gereduceerd en in de sleutel van 'beleidscontrol' gezet: het grondig kunnen toetsen van een evaluatie en kunnen toepassen ervan in de eigen organisatie.
[131] www.tierweb.nl/dtpa/opleidingsinformatie/master-evidence-based-policy-and-evaluation.html

beleidsmedewerkers, door collega's te stimuleren aan het Masterprogramma mee te doen (interview 1 juli).
Volgens respondenten bij de Algemene Rekenkamer zijn naar aanleiding van het kritische Rekenkamer-onderzoek naar subsidie-evaluaties (2011) clubjes beleidsambtenaren (in totaal 100) getraind in de planning en het design van evaluatieonderzoek – waarmee men al rekening dient te houden bij het design van de subsidies zelf (interview 1 september). Het betreft wederom de cursus 'Beleid in perspectief' bij de Rijksacademie voor Financiën, Economie en Bedrijfsvoering, waarvan de Rekenkamer het deel beleidscontrol voor haar rekening neemt.
Bij VenJ liepen in 2015 zeker twee cursussen met een evaluatiecomponent, bij de Beleidsacademie en via het Centrum voor Criminaliteitspreventie en Veiligheid (CCV). Verder is er het cursusaanbod bij de Academie voor wetgeving en overheidsjuristen, met daarbinnen eveneens aandacht voor wetsevaluaties.[132]
Op basis van internationale ervaringen met evaluatievermogen constateerden we in het tweede hoofdstuk dat het bouwen hieraan een zaak van de lange adem is. Bij de communicatie en bestendigheid van kennisverruimende initiatieven binnen departementen vallen wel nog enkele vraagtekens te plaatsen. Zo bleek enkele keren dat bestaande hulpmiddelen niet bekend waren bij onderzoekscoördinatoren of DWJZ (interviews 15 en 16 juli, 26 augustus). Wetgevingsjuristen bleken bijvoorbeeld niet bekend met het instrument beleidsdoorlichting.

3.6.2 Selectie van onderzoekers

Een andere belangrijke manier om de kwaliteit van evaluaties bij ministeries te borgen is via gerichte selectie van onderzoeksteams. Wat de opleiding en kennis en expertise van de uitvoerende onderzoekers van evaluaties en wetsevaluaties betreft, stellen beleidsdirecties geen strenge of vaste eisen. Competitie tussen kandidaat-uitvoerders voor evaluatieprojecten boven de aanbestedingsgrens fungeert als belangrijk selectiemechanisme.
Bij VenJ (WODC-EWB) is een database opgezet met de specifieke (a) contactpersonen en (b) expertises van alle onderzoeksbureaus en onderzoeksgroepen (universitair, particulier of anderszins) die ooit een offerte hebben ingediend of een evaluatie hebben verricht. Dit helpt het overzicht te houden over welke teams op welk gebied expertise hebben (interviews 8 juni en 15 juli). Niettemin kost het zeer veel moeite om het bestand 'op peil' te houden in een dynamische personele omgeving.
Onder andere bij het ministerie van SZW wordt gebruik gemaakt van formulieren met betrekking tot de opleiding en ervaring van het uitvoerende onderzoeksteam. Steeds meer wordt binnen SZW gestimuleerd dat hoogleraren en/of ervaren onderzoekers betrokken zijn in het onderzoeksteam (interview 15 juli). Ook hier werkt men aan een bestand waarin de ervaringen met onderzoeksbureaus worden vastgelegd.
Bij EZ en OCW is met de initiatieven om doeltreffendheidsonderzoek te stimuleren, al enige ervaring opgedaan met de overdracht van inhoudelijke evaluatiestandaarden aan onderzoekers. Soms reageerden externe onderzoekers hierop terughoudend: niet iedereen beschikt over de expertise die voor kwantitatief/experimenteel onderzoek vereist is (interviews 1 en 9 juli). Bij EZ zou op dit punt een communicatiestrategie zijn ontwikkeld (gebaseerd op overtuiging en meer betrekken van onderzoeksbureaus).

Multidisciplinariteit bij wetsevaluaties
Alle departementen gaven aan de *multidisciplinariteit* van het onderzoeksteam heel sterk te waarderen. EWB-projectbegeleiders bij het WODC zien een duidelijk verschil

[132] http://academievoorwetgeving.nl/landingpage/cursussen.

tussen wetsevaluaties waarbij goeddeels of uitsluitend juristen zijn betrokken en wetsevaluaties die door een multidisciplinair team worden gedaan (interview 8 juni). Dit is bij wetsevaluaties, een combinatie van juridische en sociaalwetenschappelijke aspecten, extra relevant (vgl. de ervaringen in par. 3.5).
Bewustzijn van het belang van multidisciplinariteit bestaat echter niet overal. Medewerkers werkzaam bij FEZ (SZW) en projectbegeleiders van WODC-EWB merken op dat wetsevaluaties aangevraagd door DWJZ vaak eenzijdig op de juridische kant focussen (interviews 8 juni; vgl. 16 juli en 11 augustus). De eenvoudige reden is volgens een jurist betrokken bij de drie evaluaties van de Awb dat medewerkers van de directie voornamelijk putten uit hun netwerk van juristen en academici, omdat zij onderzoekers uit andere disciplines niet kennen. Ook speelt tijdsdruk mee. In het onderzoek naar de versnellingsinstrumenten Awb uit 2014, is het sociaalwetenschappelijk perspectief stevig in de methode vertegenwoordigd (comm. december 2015). Omgekeerd ontberen veel aan externe bureaus uitbestede evaluaties een juridische component (Klein Haarhuis en Niemeijer, 2008).
Bij WODC-EWB zijn interne, informele afspraken gemaakt rond de samenstelling van onderzoeksteams. Ook de CER van ZonMw stuurt via het reeds beschreven PER aan op multidisciplinaire onderzoeksteams bij evaluaties van gezondheidswet- en regelgeving voor VWS. Twee referenten beoordelen onderzoeksvoorstellen niet alleen inhoudelijk, maar ook in termen van de (wetenschappelijke) kundigheid van de onderzoeksgroep.

3.6.3 Begeleidingscommissies

Om de kwaliteit van (wets)evaluaties te borgen, wordt verder veel gewerkt met begeleidingscommissies (BC's). Dit is bij veel departementen niet standaard, maar wel vaak het geval. Bij WODC- en CER-evaluatieonderzoek is altijd een BC betrokken. De BC monitort het verloop en de voortgang van het evaluatieonderzoek en komt gedurende het evaluatietraject enkele malen bijeen, meestal om achtereenvolgens tenminste opzet/design, voortgangs- en concept-eindrapportage te bespreken (vgl. interview 8 april). Door respondenten genoemde functies van BC's zijn kwaliteitsbewaking, bijvoorbeeld door het stellen van kritische vragen en het tijdig ingrijpen en bijsturen als het onderzoek niet naar wens verloopt (o.a. interview 8 april, 8 juni, 16 juli). Ook het creëren van draagvlak voor het onderzoek wordt genoemd; in die gevallen wil men bijvoorbeeld vertegenwoordigers van betrokken regio's, waterschappen etc. aan tafel halen (interviews 16 juli en 26 augustus).
Bij beleidsdoorlichtingen neemt de directie FEZ vaak zitting in de BC of klankbordgroep (meerdere interviews).
Bij de meeste departementen bestaan bij ex post evaluatieonderzoek geen vaste richtlijnen voor de samenstelling van BC's, men acht dit situatieafhankelijk en een kwestie van gezond verstand (o.a. interview 16 juli). Aan tafel zitten meestal een vertegenwoordiger van de aanvrager of opdrachtgever, van de uitvoeringspraktijk of andere belanghebbenden en/of inhoudelijke experts. Bij EZ heeft de vertegenwoordiger van de beleidsdirectie die ook in de Regiegroep zit, vaak ook zitting in BC's van ex post evaluatieonderzoek (als die is ingesteld, wat niet standaard het geval is; interview 9 juli).[133] Bij SZW-wetten met een groot financieel belang, zoals de modernisering van de Ziektewet en bij beleidsdoorlichtingen neemt FEZ vaak zitting in de BC (interview 15 juli).
Bij WODC-EWB en de CER bij ZonMw zijn de kaders voor de samenstelling van de BC nog het meest afgebakend. De BC die door WODC-EWB wordt samengesteld, is primair bedoeld als 'wetenschappelijke toets' en helpt de kwaliteit en de

[133] De in voorgaande paragraaf besproken gids van het BEC-secretariaat (z.j.) bevat aanwijzingen voor de samenstelling van de begeleidingscommissie.

onafhankelijke uitvoering van het onderzoek borgen (interview 8 juni). De formeel omschreven functie van de 'BC' is een inhoudelijke bijdrage leveren aan de uitvoering en de kwaliteit van het onderzoek. Het betrekken van veldorganisaties en het creëren van draagvlak is geen afdoende argument voor deelname aan de BC (WODC, 2012, p. 15). De leden nemen op basis van persoonlijke deskundigheid deel, niet als vertegenwoordiger van de organisatie waar zij werken. De voorzitter is vaak een hoogleraar of UHD met (diepgaande) kennis op het terrein van het evaluatieonderwerp. Naast de voorzitter, de projectleider van WODC-EWB en een vertegenwoordiger van de aanvrager, bevat de BC vaak een methodoloog met expertise over de te gebruiken waarnemings- experimenteer- of analysetechnieken. Ten slotte bevat de BC vaak één of twee externe, vaak wetenschappelijke deskundige(n) op het te onderzoeken terrein. Binnen het genoemde kader bestaan geen nadere regels over de samenstelling van BC's; het is maatwerk en van veel factoren afhankelijk (WODC, 2012; 8 juni).
Projectbegeleiders geven aan dat het vaak lastig is om mensen voor BC's te vinden die zowel methodologische als domein-specifieke kennis hebben. Bij specifieke onderwerpen, waaronder *wetsevaluaties* op het snijvlak van recht en empirie, is de groep die aan deze twee voorwaarden voldoet vaak klein. Bovendien hebben wetenschappers met kennis van sociaalwetenschappelijke methoden over het algemeen weinig belangstelling voor het begeleiden van juridisch georiënteerde evaluaties.
De CER bij ZonMw stelt bij elke uit te voeren wets- of thematische evaluatie een BC in. Dit is volgens de CER (interview 8 april) één van de belangrijkste factoren in de borging van evaluatiekwaliteit. De voorzitter en eventueel ook de vicevoorzitter van de BC zijn lid van de CER. Verder worden de voor de betreffende wet relevante disciplines afgevaardigd in de commissie, zoals rechten, ethiek, gezondheidszorg, psychologie. In de regel woont een waarnemer van het (of de) betrokken ministerie(s) de vergaderingen van de BC bij (ZonMw, 2013, p.8).
'

3.6.4 Delen en verankeren van ervaringen met evaluatie (aanvragers)

Dat eerdere ervaringen met evaluatieprojecten worden benut voor de aanvragende organisatie kan op verschillende manieren worden bereikt, onder meer via de aanstelling van medewerkers, via het uitwisselen van ervaringen, binnenshuis of met andere departementen, of met inzet van de vaste constructies zoals FEZ en Kennisdirecties. De praktijkervaringen die op deze punten zijn opgedaan bespreken we in deze deelparagraaf.

Selectie van beleidsmedewerkers en andere betrokkenen
Eerdere ervaring met het uitbesteden en begeleiden van onderzoeks- en evaluatietrajecten is een criterium in de werving van de projectbegeleiders bij WODC-EWB. Dat is niet het geval bij FEZ-SZW, maar wel beschouwt men wetenschappelijke ervaring, zoals een promotie in de sociale wetenschappen, als een pre (interviews 15 en 16 juli).

Overdracht en specialisatie
Het is een complexe opgave, ten minste volgens respondenten bij SZW en VenJ, om ervaringskennis en expertise aanwezig in de hoofden van dossierhouders vast te houden en te benutten. Personen binnen de departementen weten veel meer van evaluatieonderzoek en het uitbestedingsproces, dan wat geregistreerd wordt – en redelijkerwijs kan worden (interviews 8 juni en 15 juli). Veel ervaringskennis zit bij

personen en op het moment dat die vertrekken, is de kennis weg. Het kost dan moeite om deze ervaring over te dragen.[134]
Bij WODC-EWB bestrijken de projectbegeleiders van onder meer *wets*evaluaties diverse deelgebieden, wat hen de ruimte biedt om inhoudelijke expertise te ontwikkelen. Dat lukt echter alleen als de afdeling een aanzienlijk aantal evaluaties en wetsevaluaties uitbesteedt.

Informeel delen
Naast (of in plaats van) het gebruik van voornoemde structuren en evaluatie-tools vindt raadpleging en *face-to-face* overleg tussen betrokken beleidsmedewerkers, onderzoekscoördinatoren en andere betrokken units plaats – al dan niet in het periodiek afdelingsoverleg (o.a. interviews 8 juni, 9 juli en 26 augustus). Men leert van elkaar door ervaringen met de evaluatieaanpak uit te wisselen. Dit gebeurt bijvoorbeeld na afloop van ieder evaluatieonderzoek voor EZ. Bij BZK worden de projectleiders van lopende en aanstaande beleidsdoorlichtingen door FEZ aan elkaar gekoppeld voor kennisuitwisseling (interview 13 oktober). Ook FEZ-SZW probeert samen met de CSO en zijn werkgroep te voorkomen dat de kennis binnen directies verkokert (interview 15 juli).
Bij WODC-EWB worden kennis en ervaring vooral informeel gedeeld tussen projectbegeleiders: via regelmatig afdelingsoverleg met collega's, feedback vragen bij het afdelingshoofd en het MT en het aangereikt krijgen van (nieuwe) wetenschappelijke inzichten door de directeur (interview 8 juni). Op deze manier worden ervaringen met evaluaties informeel doorgegeven; hoewel men zich de vraag kan stellen of dit toereikend is voor het behoud – het borgen – van opgedane kennis/ervaringen en voor evaluatiekwaliteit op de langere termijn (interview 8 juni).
Binnen het IBP wisselen FEZ-vertegenwoordigers werkzame aanpakken uit van evaluaties en beleidsdoorlichtingen, wat door menigeen als nuttig wordt ervaren (interviews 16 juli, 15 september en 13 oktober).

Verankering van kennis en ervaring via de organisatie
Omdat het evaluatieonderzoek slechts een klein onderdeel is van al het onderzoek dat voor het ministerie wordt gedaan, komt het vaak voor dat een beleidsmedewerker voor het eerst verantwoordelijk is voor een dergelijk (wets)evaluatiedossier (interview 26 augustus). De onderzoekscoördinatoren bij veel departementen hebben relatief veel kennis van en ervaring met evaluatie- en ander onderzoek.[135] Bij OCW biedt de coördinator bijvoorbeeld ondersteuning bij: de formulering van de vraag- en probleemstelling, de opstelling van startnotities en keuzes met betrekking tot onderzoeksdesign en –methoden (interviews 1 juli, 26 augustus). Bij een aantal departementen is voor DG's en directies is, zoals al aangegeven, extra inhoudelijke ondersteuning georganiseerd. De genoemde CSO bij SZW en de Kennisdirectie bij OCW staan beleidsmedewerkers en onderzoekscoördinatoren bij met kennis en feedback over vraag- en probleemstelling, onderzoeksdesign en -methoden (interviews 1 en 15 juli).
De Kennisdirectie van OCW levert feitelijk vooral in de beginfase ondersteuning, zoals bij het opstellen van startnotities (interview 26 augustus). De inbreng van de directie Kennis, bijvoorbeeld in het kader van Zicht op effectiviteit, wordt door beleidsmedewerkers soms wat technisch-methodologisch gevonden, wat vertaling naar de praktijk dan lastig maakt. Keuzes in het onderzoeksdesign worden in de

[134] Bij SZW stimuleert de werkgroep van de CSO dat opgedane kennis gedeeld wordt. 'Trekkers' van een onderwerp presenteren resultaten in het MT SZW. Ook wordt een overzicht gemaakt van de onderzoeksresultaten.
[135] Hoewel we geen duidelijk aangegeven ondergrenzen aantroffen.

praktijk vaak aan het (externe) onderzoekersteam overgelaten (interviews 1 juli en 26 augustus).
Een vergelijkbare rol als voornoemde vervult de vertegenwoordiger van de beleidsdirectie of *linking pin* in de Regiegroep Monitoring en Effectmeting bij EZ: die geeft desgewenst tips over methoden, geschikte onderzoeksbureaus en mogelijke BC-leden (BEC-secretariaat, z.j., p. 26). Achterin het BEC-boekje is per DG de betreffende 'hulplijn' (contactpersoon) te vinden.
Bij het WODC is de rol van de aanvrager deels 'uitbesteed' aan de EWB-projectbegeleiders, die het proces na de inventarisatie van de onderzoekwensen bij de aanvrager goeddeels overnemen.
Het bestendigen van evaluatiekennis- en ervaring is een uitdaging gebleken in dynamische personele omgevingen. Zo is bij BZK een van de kennisdirecties na 2011 praktisch geheel ontmanteld (interview, 13 oktober).[136] 'Verloop' van kundige beleidsmedewerkers is fnuikend voor het behoud van expertise (interviews 24 en 26 augustus). Rond wetsevaluatie is wel geklaagd over verloop van verantwoordelijke wetgevingsjuristen: de betrokkene bij de totstandkoming van een wet en evaluatiebepaling is vaak niet meer betrokken zodra de wetsevaluatie op handen is. Als het gaat om nieuwe werkwijzen en handreikingen in het kader van evaluatievermogen, lijken de beoogde gebruikers hier bovendien niet altijd 'klaar voor' (interviews 1 juli en 26 augustus). Hierdoor kunnen initiatieven in een weinig vruchtbare omgeving landen.[137] Ten slotte dreigen ontwikkelingen op aanpalende terreinen de opgebouwde structuren rondom evalueren soms uit te hollen. De centralisering van inkooppunten was bedoeld om het offerte- en aanbestedingstraject, ook rond evaluaties, door standaardisering transparanter te maken (Rijksoverheid, z.j.). Maar deze transparantieslag heeft een uithollende werking op aanvrager-ondersteunende formats die specifiek voor evaluaties zijn ontwikkeld (interviews 1 en 9 juli). De uitdaging volgens respondenten bij EZ is dat de nieuwe inkoopcentra[138] de op evaluatiegebied ontwikkelde formats en kaders inhoudelijk volgen.

3.6.5 Samenvattend over borging

In lijn met wat we in hoofdstuk 2 internationaal signaleerden, wordt de evaluatie-aanvrager ook in ons land steeds vaker gezien als een belangrijke schakel in procesverloop en kwaliteit. In deze paragraaf is beschreven, op welke manieren bij beleidsdepartementen is geborgd dat aanvragers kennis hebben van het evaluatieproces en -inhoud om deze in goede banen te kunnen leiden. Bij de meeste DG's en directies vervullen de onderzoekscoördinatoren hierin een spilfunctie; zij hebben relatief veel onderzoeksexpertise en - ervaring en fungeren als vraagbaak voor collega's. Daarbuiten verschilt de ondersteuning bij departementen sterk, afhankelijk van de betrokken organisatieonderdelen. Zo heeft bij SZW de *Chief Science Officer* (CSO) inhoudelijk een ondersteunende rol; bij OCW is dat de directie Kennis en bij EZ de Regiegroep M&E. Bij VenJ is de rol van de aanvrager voor een

[136] Bij het DG Wonen en Rijksdienst van BZK bestaat nog wel de directie Kennis & Strategie.
[137] Eerder stierven initiatieven ter borging van evaluatiekennis een stille dood; waaronder een rijksbreed initiatief aan het begin van deze eeuw om een evaluatie-overzicht (EOR-evaluatieoverzicht Rijk) op te zetten en steeds te actualiseren, met daarin alle door de departementen opgeleverde evaluaties waaronder ook wetsevaluaties. Het EOR hield werd niet systematisch bijgehouden en bevatte veel dubbelingen.
[138] Sinds 1 januari 2014 zijn er 20 nieuw ingestelde 'inkoopuitvoeringscentra' (IUC's) verspreid over het land. Deze moeten het inkoopproces voor alle partijen meer doelmatig maken. Sinds 2008 is sprake van categoriemanagement; dit betekent voor het Rijk dat departementen niet alleen voor zichzelf bepaalde producten of diensten inkopen, maar gecoördineerd per productgroep of categorie voor elkaar, ter bevordering van specialisatie. Uit onderzoek van Gfk (2015) bleek dat driekwart van de ambtenaren niet bekend is met het nieuwe inkoopstelsel.

deel ‘uitbesteed’ aan de projectbegeleiders bij het WODC, die na inventarisatie van de wensen bij de aanvrager het proces goeddeels overnemen.
Meerdere departementen en rijksacademies bieden bovendien cursussen of trainingen aan voor beleidsmedewerkers; die deels of geheel over evaluatie gaan. Omdat de beschreven constructies en cursussen nogal ‘vers’ zijn, is het lastig na te gaan in hoeverre ze daadwerkelijk hun vruchten hebben afgeworpen in termen van evaluatieproces en –kwaliteit.
Bij de selectie van onderzoekers wordt vooral gebruik gemaakt van de competitie tussen onderzoeksgroepen in het aanbestedingsproces. Daarbovenop zijn een beperkt aantal inspanningen benoemd (bij WODC, ZonMw, OCW, SZW en EZ) zoals het gerichter werken met selectiecriteria en het bijhouden van databases met onderzoeksgroepen. In een dynamische personele omgeving is het een uitdaging gebleken om gegevens over de capaciteiten van onderzoekspartijen up-to-date te houden. Juist bij wetsevaluaties is multidisciplinariteit in het onderzoeksteam van belang: een combinatie van juridische en empirische (sociaalwetenschappelijke) expertise. In de praktijk is die combinatie lang niet altijd waargemaakt, maar er zijn ook andere voorbeelden, zoals de vaste procedure bij ZonMw en de latere Awb-evaluatie van versnellingsinstrumenten.
Ook BC’s worden vaak ingezet ter bewaking van voortgang, inhoud of kwaliteit. Bij het WODC (VenJ) is het standaardpraktijk en tevens bij de CER, voor de daar door VWS uitgezette wetsevaluaties. Bij WODC-evaluatieonderzoek wordt altijd een wetenschapper als voorzitter aangewezen; bij ZonMw is het voorzitterschap in handen van de CER zelf.
Het uitwisselen van ervaringen op het vlak van evaluatie gebeurt binnen departementen meestal op informele manieren of in afdelingsoverleggen. Interdepartementaal gebeurt het vooral tussen FEZ-directies en Financiën in het IBP.
Initiatieven ter versterking van het evaluatievermogen lijken kwetsbaar want dreigen soms in de knel te komen: beleidsmedewerkers zijn er niet klaar voor, of er zijn reorganisaties (zoals bij inkoop), of snel personeelsverloop. Dit bevestigt een observatie uit de internationale praktijk dat evaluatievermogen een kwestie van een lange adem is.

3.7 Bruikbaarheid en gebruik van (wets)evaluaties

In deze laatste paragraaf gaan we in op het gebruik van evaluaties en wetsevaluaties. Gebruik vormt een belangrijk sluitstuk van evaluatievermogen: vindt geen (systematische) terugkoppeling plaats naar de beleids- en verantwoordingscyclus, dan is de vraag wat voor nut het organiseren en uitvoeren van evaluaties nog heeft.
Gebruik begint met bruikbaarheid: de communicatie van bevindingen in het evaluatierapport zelf en de verspreiding ervan, via de publicatie van het rapport en van documenten en artikelen die over de resultaten worden geproduceerd en verspreid. Ook de timing van afronding en openbaarmaking is relevant. Deze aspecten beschrijven we eerst. Daarna gaan we in op de vraag in hoeverre bij departementen procedures of andere voorzieningen bestaan ter bevordering van evaluatiegebruik en het leren van evaluaties voor beleid.

3.7.1. Communicatie tussen aanvrager en onderzoeker

Johnson et al. (2009) signaleerden in een metastudie dat de betrokkenheid van belangrijke stakeholders bij een evaluatie voor het gebruik ervan cruciaal is; meer nog dan de evaluatiekwaliteit *an sich*. Eén vorm daarvan is contact tussen aanvrager of andere stakeholders enerzijds en evaluatieonderzoekers anderzijds in

vergaderingen van de BC. Ook de timing en tijdigheid van oplevering en de politieke of ambtelijke besluitvormingscontext zijn doorslaggevend gebleken.
Dit beeld wordt door de meerderheid van gesproken respondenten bevestigd. Contact en overleg tussen aanvrager en onderzoekers gedurende het evaluatietraject – al dan niet via BC-vergaderingen – is volgens meerdere respondenten van aanvragerszijde een belangrijke manier om de kans op evaluatiegebruik te vergroten. Om het contact in te kaderen, geldt bij PER-wetsevaluaties: 'Beide partijen spreken hun wederzijdse bedoelingen en verwachtingen helder en nadrukkelijk uit en komen gemaakte afspraken na voor effectieve uitvoering en realisatie van de opdracht. De aard van de relatie is gelijkwaardig en niet vrijblijvend' (ZonMw, 2013, p.8). De onderzoekers zijn eindverantwoordelijk voor de evaluatieresultaten. ZonMw kan alleen besluiten een rapport niet vast te stellen of kanttekeningen maken in de begeleidende brief (interview 8 april).
In het kader van 'timing' in relatie tot beleidsontwikkeling of besluitvorming vinden meerdere respondenten dat sommige wetsevaluaties *te vroeg* worden verricht voor het kunnen trekken van adequate conclusies (zie ook hoofdstuk 2 en par. 3.3). Anderzijds vindt men een evaluatietraject soms te laat afgerond om nog tot doorwerking te kunnen leiden.

3.7.2 Wijze van presentatie, aanbevelingen

Wetsevaluaties bevatten niet altijd geschikte aanbevelingen. Soms behoeft de wettekst zelf geen aanpassing, maar omliggende structuren of flankerend beleid (interview 24 augustus; vgl. Michiels, 2010). Een wetswijziging naar aanleiding van een evaluatie vergt zeer veel energie en tijd; bovendien dient een wijziging te passen binnen het grotere systeem van wet- en regelgeving. De baten van een wetswijziging zouden tegen deze kosten moeten opwegen.
Een aanvrager bij VenJ stelt dat aanbevelingen in evaluatieonderzoek hetzij vaak ontbreken, hetzij niet adequaat zijn. Hij schrijft dit toe aan het gegeven dat de beleids- en de onderzoekerswereld verschillend in elkaar zitten (interview 15 juni).
De aanvragers van WODC-evaluaties en andere onderzoeken geven de bruikbaarheid van aanbevelingen en adviezen gemiddeld een ruime voldoende. De score voor een 'goede' bruikbaarheid is in een substantieel deel van de gevisiteerde periode gedaald en heeft zich in 2012 weer iets hersteld (WODC, 2014, p. 17-18). Naar aanleiding van de laatste visitatie is de bruikbaarheid voor beleid van WODC-onderzoek, inclusief evaluatieonderzoek, benoemd als blijvend aandachtspunt voor de toekomst – onder andere via scherpe vraagarticulatie en methodenkeuzen, het 'nazorgtraject' van onderzoek en het betrekken van stakeholders van binnen en buiten de departementen van VenJ en BZK.

3.7.3 Gebruik

Een belangrijke vraag is, uiteindelijk, in hoeverre evaluatieresultaten hun weg terug weten te vinden naar het proces van ambtelijke besluitvorming of de politieke arena. Als evaluatieresultaten niet worden teruggekoppeld, kan men zich immers afvragen wat de zin van de hele exercitie is geweest. In hoofdstuk 1 gebruikten we in dit verband het begrip *'regulatory cycle'*: is de beleids- of wetgevingscyclus rond? Alle departementen volgen op hoofdlijnen dezelfde uitwendige procedure na de afronding van (wets)evaluaties. De opdrachtgever of aanvrager, zoals de beleids- of wetgevingsdirectie, moet binnen drie maanden na de oplevering van het evaluatierapport een beleidsreactie formuleren en eventueel – bijv. in geval van een toezegging – naar de Tweede Kamer sturen. Daarin moet worden gemeld wat de minister van plan is met het onderzoek te gaan doen. Het rapport en de kabinetsreacties worden in de meeste gevallen gepubliceerd en kunnen anders op

aanvraag worden geraadpleegd. Zodra de beleidsreactie is gepubliceerd, is de betrokkenheid van eventuele ondersteunende afdelingen (Kennis, FEZ, CER, WODC-EWB etc.) formeel voorbij en kunnen de ontwikkelingen alleen nog van een afstand worden gevolgd.
Bij EZ had men aanvankelijk de intentie dat de BEC en de voormalige Regiegroep M&E (nu BAT) de conclusies van enkele evaluaties zouden bespreken met deskundigen en uitvoerders, mede om van de aanpak te leren. Dit is echter niet of nauwelijks in de praktijk gebracht. Wel heeft men een enkele keer met een onderzoeksbureau op de uitvoering van een beleidsdoorlichting teruggeblikt.
Voor de CER blijft begeleiding belangrijk tot na de afrondingsfase: na afronding van een wetsevaluatie stelt de CER het rapport vast. Dat gebeurt alleen als de wetsevaluatie van voldoende kwaliteit is. De CER kan ook een aantal kanttekeningen formuleren; ZonMw neemt deze op in de begeleidende brief bij het aanbieden van het rapport aan de opdrachtgever (ZonMw, 2013, p.8).

Verspreiding
Een andere manier om de evaluatiebevindingen bekend te maken en te verspreiden, is door korte presentaties en toelichtingen op de werkvloer te geven. Voor EZ schrijven onderzoeksbureaus zelf soms brieven en notities over hun evaluatierapport (interview, 9 juli). Bij het WODC stellen de onderzoekers of de EWB-projectbegeleiders de oplegnotitie voor 'het beleid' op.
Daarnaast vervaardigen onderzoekers vaak extra vak- of wetenschappelijke publicaties met daarin bijvoorbeeld een samenvatting of beschouwing van bevindingen van het onderzoek.[139] Vaak gebeurt dit op persoonlijke titel.
Al deze vormen van verspreiding kunnen bijdragen aan het gebruik van evaluaties door zowel beleidsmakers als door anderen, zoals professionals of de wetenschap. Ook ZonMw zet zich in om de uitkomsten van wetsevaluatieonderzoek bekend te maken. Door de CER vastgestelde wetsevaluaties worden opgenomen in de rapportenreeks *Evaluatie Regelgeving*. Daarnaast moedigt de CER onderzoekers aan, voordrachten en publicaties over hun project te verzorgen of een congres te organiseren. Zo stimuleert ZonMw kennisoverdracht aan relevante actoren in het veld, zoals beroepsbeoefenaren, ziekenhuizen, overige gezondheidszorginstellingen en patiënten (ZonMw, 2013, p.10). Implementatie van de resultaten van een wetsevaluatie is echter in eerste instantie voorbehouden aan de wetgever zelf. Dit neemt niet weg dat ZonMw steeds zal bezien of ze bij de implementatie van de resultaten van evaluatieonderzoek een verdere rol kan spelen (ibid.; interview 8 april).

Doorwerking
Het WODC en FEZ-BZ doen aan follow-up van gebruik van evaluaties (interviews 9 juni en 2 september; Box 8). Het WODC polst de benutting in het kader van de interne projectevaluaties met een korte vragenlijst ter invulling door de aanvrager, hooguit enkele maanden na de oplevering. Ook is bibliografisch onderzoek naar benutting van onder meer SCP-, WODC- en IOB-studies verricht, mede op basis van citatiescores (o.a. Prins, 2012a,b).[140] Uit de zelfevaluatie over de periode 2006-2012 blijkt dat aanvragers van bij het WODC uitgezette onderzoeken aangeven de resultaten in ongeveer de helft van de gevallen te gebruiken ten behoeve van *beleid en regelgeving* en in een kwart van de gevallen (ook) ten behoeve van beleids*uitvoering* (Van der Laan et al., 2013; vgl. WODC 2013). Een kleine 20% van

[139] Overigens kan ook een onderzoeksbureau bijv. op haar eigen website een korte samenvatting van het rapport opnemen, met een link naar het rapport en een evt. brief aan de Kamer.
[140] Dit laat zien dat in wetenschappelijke bronnen minder en in aan het beleid gerelateerde bronnen meer wordt verwezen naar WODC-onderzoek, onder meer in vergelijking met SCP-onderzoek.

de WODC onderzoeken wordt voor overleggen met Tweede en Eerste Kamer gebruikt.

Box 8 Onderzoek naar gebruik bij BZ (Van Gils en Leeuw, 2010)

Rond 2010 is onderzoek verricht naar het gebruik van 30 evaluaties bij het ministerie van BZ. Er zijn veel typen evaluatiegebruik. Evaluaties zijn gebruikt ter verantwoording van beleid onder meer via toezending van evaluaties van de IOB, voorzien van een beleidsreactie, aan de Tweede Kamer, door publicatie van rapporten op de website van het ministerie en door gebruik als informatiebron voor bijvoorbeeld het jaarverslag van het ministerie.
Andere gevonden vormen van gebruik zijn: ter ondersteuning van overdracht bij functiewisselingen (op individueel initiatief), in de externe communicatie: (a) ter legitimering, als basis voor overleg in het kader van fondsenverstrekking en (b) ter promotie, d.w.z. de beleidsprestaties op de kaart zetten.
Tussen DG's bestaat verschil in het directe, lerende gebruik van evaluaties: '[v]eel evaluaties binnen (...) Internationale Samenwerking (DGIS) gaan gepaard met een herziening van het geëvalueerde beleid.' Causaal en lineair is het verband tussen evaluatieresultaten en beleidsaanpassingen volgens de onderzoekers niet; evaluaties blijken zeker niet de enige bron. Wel is duidelijk geworden dat een BC 'de best denkbare leeromgeving is om iets van een evaluatie op te steken'. Hierdoor heeft direct gebruik (in beleid) vaak al voor de officiële publicatie plaats gevonden.

Hoewel verplichte evaluaties vaak wel afgerond worden, komt het voor dat ze na verloop van tijd als minder politiek relevant worden gezien (interview 2 september); dan blijft gebruik vaak beperkt tot kennisname. Niettemin percipiëren enkele respondenten dat de Tweede Kamer toenemende belangstelling laat zien in ex post evaluaties, met name in het licht van de verantwoordingsfunctie en financiële lasten (interviews 15 en 16 juli).[141] Naar aanleiding van kabinetsreacties op evaluatierapporten en beleidsdoorlichtingen stellen Tweede Kamerleden vaker vragen over wat de minister van plan is met de bevindingen en eventuele aanbevelingen *te doen*.
Over het gebruik van de derde evaluatie van de Awb gaat box 9.

Box 9 Gebruik van de derde Awb-evaluatie

De aanbevelingen van de derde evaluatie werden onderverdeeld in aanbevelingen aan de wetgever, de regering, bestuursorganen, rechters en onderzoekers. Voor groepen zoals rechters moge duidelijk zijn dat de aanbevelingen nauwelijks een verplichtend karakter hebben (Michiels, 2010). Lang niet alle aanbevelingen waren gericht op aanpassing van de Awb.
Naar aanleiding van het verschijnen van het verslag van de evaluatiecommissie is op 8 maart 2007 een symposium gehouden voor alle bij de evaluatie betrokkenen (onderzoekers, leden van de BC's, etc). Voordien was het rapport voorgelegd aan een groot aantal betrokken partijen, waarop vervolgens een drietal reacties zijn ontvangen. De latere kabinetsreactie heeft rekening gehouden met de uitkomsten hiervan en met commentaren die in de juridische vakliteratuur zijn verschenen (Tweede Kamer, vergaderjaar 2009–2010, 29 279, nr. 111, p.3). Uit de kabinetsreactie op de derde Awb-evaluatie valt op te maken dat aan een groot aantal overige aanbevelingen reeds geheel of gedeeltelijk gevolg is gegeven (Tweede Kamer, vergaderjaar 2009–2010, 29 279, nr. 111). Niettemin constateerde Michiels (2010) dat het al beperkte aantal tot de wetgever gerichte aanbevelingen slechts beperkt heeft doorgewerkt in

[141] Deze perceptie wordt echter niet alom gedeeld, getuige het verslag van een bijeenkomst van VIDE over het centrale evaluatiestelsel en de beleidsdoorlichtingen, november 2015: 'De behandeling van beleidsdoorlichtingen in de TK is veelal summier en is (te) vaak opgenomen in een volle AO-agenda.' www.videnet.nl/download/?id=17773737.

wetsaanpassingen. De auteur acht dit op zichzelf niet nadelig, omdat (a) evaluaties ook een positief of bevestigend oordeel in kunnen houden en (b) mede dankzij de evaluaties er toch wel sprake is van een voortdurend proces van ontwikkeling van de Awb. De evaluaties zouden bovendien het bestuursrecht in het algemeen en de wetenschap van het bestuursrecht verder brengen.

Horizonbepaling
In geval van subsidieregelingen – veel gebruikt instrument bij o.a. EZ – is er sinds enkele jaren een horizonbepaling die stelt dat de regeling vijf jaar na inwerkingtreding automatisch vervalt (interviews 9 juli, 1 september). Tegelijkertijd geldt op grond van de Awb een periodieke evaluatieverplichting voor eveneens vijf jaar (Algemene Rekenkamer, 2011). Een regeling kan gemotiveerd worden voortgezet, bijvoorbeeld op grond van een positieve doeltreffendheidsevaluatie. De subsidieregeling voor innovatie in de scheepsbouw is ingetrokken als gevolg van een negatieve evaluatie.

3.7.4 Leeropbrengsten

En aantal respondenten betreurt de nadruk die in het centrale evaluatiestelsel en de beleidsdoorlichtingen op verantwoording is komen te liggen (interviews 15 juli, 2 september). Van plan- en procesevaluatie – en vele andere soorten beleidsonderzoek – kan bijvoorbeeld veel geleerd worden, ondanks dat de 'hamvraag' van doeltreffendheid en doelmatigheid buiten beeld blijft. Zo hebben plan- en procesevaluaties in de OER-studies van de Algemene Rekenkamer (2011-2013) geen aandeel in het percentage dat uitdrukt welk deel van een begrotingsartikel door doeltreffendheidsonderzoek is afgedekt.
Ex post evaluaties zijn primair bedoeld voor terugblikken op uitrol en doorwerking van beleid. Om de leerfunctie van evaluatie een belangrijker plaats te geven, zouden ex post evaluaties volgens sommige respondenten (ook) op de toekomst gericht moeten zijn. In plaats van alleen de vraag te beantwoorden of het beleid of de wet de gewenste effecten wist te bereiken, zou moeten worden besproken of het beleid of wet nog *zinvol* is binnen het huidige stelsel (interviews 15 juli en 2 september). Respondenten stellen dat zij veel hebben geleerd van evaluaties die in staat zijn nieuwe discussies te starten over hoe een (maatschappelijk) probleem kan worden aangepakt. Als voorbeelden daarvan noemen respondenten de evaluaties van het bedrijvenbeleid, de agroregeling, de Pensioenwet, de Algemene nabestaandenwet (Anw), de vroegere Politiewet en de Spoorwetgeving (V&W, 2008).
De Algemene Rekenkamer denkt ook steeds vaker mee aan de voorkant en doet aanbevelingen; o.a. over de formulering van meetbare doelstellingen (1 september). Daarbij wordt toenemende nadruk op thema's en casus gelegd.[142]
Een tweede leerfunctie van (wets)evaluaties is volgens respondenten dat ze de impliciete aannames waarop de wet of het beleid berust, boven water halen (interviews 8 juni, 2 september; vgl. Panel van Advies IOB). Evaluaties dragen bij aan de explicitering en verheldering van de beleidstheorie en aannames waarop beleid/wetgeving steunt. Een goede wetsevaluatie zou dus de vraag moeten stellen of de veronderstellingen van de beleidsmaker kloppen en via welke mechanismen bepaalde effecten zijn bereikt. Want het kan blijken dat de gewenste effecten wel zijn bereikt, maar via andere mechanismen dan in de beleidstheorie aangenomen.

Uitgelicht: leren als speerpunt (BZ)

142
http://verantwoordingsonderzoek.rekenkamer.nl/2014/rijksbreed/beleidsresultaten/beleidsdoorlichtingen-nog-te-weinig-zicht-op-doelmatigheid-en.

Van Gils en Leeuw (2010) vonden dat bij de beleidsdirecties van BZ geen sprake was van een echt leergierige cultuur: 'Het is niet gebruikelijk om actief kennis uit evaluaties te zoeken' (p.12). Volgens de auteurs heeft (nieuwe) kennis over hoe gedrag te beïnvloeden een grotere generaliseerbaarheid en aantrekkingskracht dan terugblikkende, op processen gerichte ex post evaluaties. Verder reikten de ambities met betrekking tot leren binnen het ministerie verder dan wat tot dan toe fragmentarisch was gerealiseerd; individueel leren is nog geen institutioneel leren. Eerste aanbeveling was dan ook: 'schep meer duidelijkheid over wat bij het ministerie onder leren verstaan wordt.. (....) en hoe organisatie en medewerkers geprikkeld kunnen worden tot leren' (p.13).
Om de leerfunctie van evaluaties te versterken, heeft de toenmalige Minister voor Ontwikkelingssamenwerking de Tweede Kamer op 19 februari 2009 toegezegd een Panel van Advies (PvA) van externe deskundigen[143] in het leven te roepen – met een mandaat dat uiteindelijk tot en met 2014 heeft geduurd. Het Panel had tot taak de bruikbaarheid en het gebruik van evaluatiestudies voor beleid en praktijk te bevorderen. Er zijn drie opeenvolgende adviezen aan de bewindspersoon uitgebracht, waarin verschillende vormen van gebruik van IOB rapporten onder de loep zijn genomen en in het kader van 'leren' ook een wetenschappelijk-inhoudelijke kwaliteitstoets is verricht – met nadruk op de beleidstheorie. Box 10 geeft enkele adviezen weer.

Box 10 Adviezen van het Panel van Advies inzake Bruikbaarheid van IOB-rapporten (BZ)

- Baken evaluatieobjecten en probleemstellingen zo precies mogelijk af;
- Maak meer gebruik van (internationale) wetenschappelijke literatuur over de verschijnselen en hun oorzaken die in de evaluaties centraal staan;
- Zorg voor een evaluatiedesign dat aansluit bij de vraagstelling (...) en dat geldt ook voor mointoring;
- Draag er aan bij dat er baselinedata beschikbaar zijn voordat evaluaties naar (nieuw) beleid worden aangevangen;
- Sluit zoveel mogelijk aan bij de *state of the art* op het gebied van de methodologie van synthesestudies: 'systematic reviews', meta-analyses en 'realist syntheses';
- Draag bij aan het via een 'portal', resp. 'repository' snel en efficiënt toegankelijk maken van resultaten uit evaluaties voor beleid, control en samenleving;
- Maak meer gebruik van de ervaringen die zijn opgedaan bij eerdere, te vergelijken, evaluaties, óók op andere beleidsterreinen of in andere landen, mits de ingezette *'tools of government'* en de daaraan ten grondslag liggende gedragsmechanismen (de interventietheorie) en de context(en) redelijk vergelijkbaar zijn;
- Draag bij aan een gearticuleerde theorie over wat BZ/ontwikkelingssamenwerking verstaat onder 'leren van evaluaties';
- Bepaal van te voren wie de gebruikers van de evaluatie kunnen zijn en hoe zij het beste geïnformeerd kunnen worden;
- Zorg dat de omvang van de rapporten het gebruik niet vermindert of belemmert;
- Zorg dat voor de gebruiker duidelijk is waar welke informatie in een rapport te vinden is; en
- Maak in het rapport duidelijk wat men van deze evaluatie kan leren en zo mogelijk hoe dit kan worden gedaan.

3.7.5 Samenvattend over bruikbaarheid en gebruik

[143] In het expertpanel zullen de volgende onderscheiden en complementaire expertises vertegenwoordigd zijn: 1. kennis van het functioneren van overheidsorganisaties en in het bijzonder de betekenis van de evaluatiefunctie binnen de overheid; 2. wetenschappelijke kennis op het terrein van ontwikkelingssamenwerking, relevante onderzoeksactiviteiten en methoden en technieken van toetsend onderzoek.

Evaluatiegebruik geeft zin aan alle overige behandelde aspecten van evaluatievermogen: zonder dat vindt geen terugkoppeling naar de beleids- of wetgevingscyclus plaats en rijst de vraag wat het nut van evalueren is geweest. We maakten onderscheid naar bruikbaarheid en wat bij departementen omtrent gebruik is geregeld.
Contact en overleg tussen aanvrager en onderzoekers gedurende het evaluatietraject – al dan niet in een BC – is volgens de literatuur en ook volgens de ondervraagde respondenten een belangrijke manier om de kans op gebruik te vergroten. Voorts is de timing en tijdigheid in relatie tot de beleidsvorming van belang gebleken.
Over het algemeen hanteert elk onderzocht departement een vergelijkbare formele wijze van communicatie, (korte-termijn) kennisname en gebruik van evaluatieonderzoek. In de regel worden rapporten openbaar gemaakt en vaak ook naar de Tweede Kamer gestuurd – met een beleidsreactie uiterlijk drie maanden na de oplevering door de onderzoekers. Door onderzoekers vindt daarnaast vaak nog verspreiding plaats in de vorm van academische en vakpublicaties. Een andere vorm van valorisatie is het geven van presentaties en toelichtingen op de departementale werkvloer.
Na publicatie en eventuele valorisatie is het aan beleidsmedewerkers en politiek wat er met bevindingen gebeurt. Er zijn bij departementen geen ‘verplichte’ constructies gevonden voor terugkoppeling naar de beleidscyclus. De voorbeelden van gebruik die respondenten aan hebben gedragen zijn casuïstisch. Bij subsidieregelingen (EZ, met name) is er een horizonbepaling, die stelt dat de regeling na vijf jaar automatisch vervalt; voortzetting kan mits voldoende gemotiveerd, bijvoorbeeld op basis van doeltreffendheidsonderzoek. Verder doen WODC en FEZ-BuZa aan monitoring van gebruik. Bij BZ is van ‘leren’ een speerpunt gemaakt en een Panel van Advies in het leven geroepen onder meer om de bruikbaarheid van evaluaties en de terugkoppeling naar beleid te vergroten.
Hoewel verplichte evaluaties vaak wel afgerond worden, komt het voor dat ze na verloop van tijd als minder politiek relevant worden gezien; dan blijft gebruik vaak beperkt tot kennisname. Niettemin percipiëren enkele respondenten dat de Tweede Kamer toenemende belangstelling laat zien in ex post evaluaties en in wat de minister daarmee wil gaan doen. Het betreft met name de verantwoordingsfunctie en financiële lasten.
Om de leerfunctie van evaluatie een belangrijker plaats te geven, zouden de evaluaties volgens sommige gebruikers meer actie- en toekomstgericht moeten zijn: in plaats van (alleen) de vraag te beantwoorden of het beleid of de wet de gewenste effecten wist te bereiken, zou bijvoorbeeld moeten worden besproken of het beleid of wet nog zinvol is binnen het huidige stelsel. Respondenten stelden dat zij waarde hechten aan evaluaties die in staat zijn nieuwe discussies te starten over hoe een bepaald maatschappelijk probleem kan worden aangepakt. Een tweede genoemde leerfunctie van (wets)evaluaties is dat ze de impliciete aannames waarop wet of beleid rust, boven water halen en toetsen aan de praktijk.

Hoofdstuk 4. Slotbeschouwing

4.1 Inleiding

Evalueren is iets waar men principieel bijna niet tegen kan zijn; het is een stap in het bereiken van verbetering van, of afleggen van verantwoording over, eerdere plannen en acties. Dat geldt ook voor het evalueren van beleid en wetgeving. Vragen zoals 'hoe is het precies uitgepakt', 'zijn de beoogde doelen bereikt'? kunnen moeilijk als irrelevant worden afgedaan, helemaal als men wil leren voor toekomstige initiatieven. Maar dergelijke vragen beantwoorden kost geld en tijd; de omwentelingssnelheid van evaluaties ligt meestal een stuk lager dan die van politiek en beleid.
Willen evaluaties daadwerkelijk een bijdrage kunnen leveren aan de kwaliteit van wetgeving, dan is het van belang om: (1) te zorgen dat ze worden gedaan; (2) ze zó op te zetten dat ze bruikbare en tevens valide en betrouwbare kennis en inzichten genereren; (3) dat ze beleidsmakers en eventueel wetgevingsjuristen bereiken, die ze vervolgens gebruiken in de terugkoppeling naar hun producten. Maar dit is in de praktijk niet vanzelfsprekend – en niet alleen vanwege de tijd en energie die evalueren kost. In de eerste plaats zeggen veel individuele aanvragers – van 'beleid' en 'wetgeving' – niet precies te weten wat hen te doen staat als er een wetsevaluatie voor de deur staat. Afgezien van de algemene begrippen in de Aanwijzingen voor de regelgeving – waaronder doeltreffendheid, doelmatigheid, uitvoerbaarheid, handhaafbaarheid, evenredigheid en duidelijkheid – is er geen centraal wetsevaluatiebeleid en bestaat niet zoiets als een IAK voor ex post evaluaties. In de tweede plaats is wetgeving een complex evaluatieobject. Ook al gaan beleid en wetgeving hand in hand, het evalueren van wetten of wetswijzigingen brengt een extra dimensie met zich mee. Wetten zijn doorgaans een mix van regels, normenkaders en beleidsinterventies. Behalve empirische en beleidsevaluatie-vaardigheden vereisen wetsevaluaties dus ook een juridische blik. Het ontbreken van een algemeen evaluatiekader, gecombineerd met de complexiteit van wetten als evaluatieonderwerp kan ertoe hebben geleid dat wetsvaluaties in de praktijk heel verschillend worden aangepakt: zowel qua inrichting van het aansturingsproces als methodologisch-inhoudelijk. Dat hoeft helemaal niet problematisch te zijn – de ene wet is immers de andere niet. Maar problematisch wordt de gevarieerde praktijk mogelijk wel, als de validiteit en betrouwbaarheid van wetsevaluaties sterk gaan variëren. Hier zijn in eerder meta-onderzoek aanwijzingen voor gevonden (Klein Haarhuis en Niemeijer, 2008; Veerman, 2013). De gevarieerde praktijk van wetsevaluatie houdt mogelijk verband met verschillen in *evaluatievermogen* bij departementen: het vermogen om wetsevaluaties te verrichten en ze bovendien te gebruiken (Bourgeois & Cousins, 2013). In hoeverre zijn de voorwaarden aanwezig om het doen en gebruiken van evaluaties mogelijk te maken?
Bij die voorwaarden kan zowel gedacht worden aan kenmerken van het evaluatieproces als aan de inhoud van evaluaties. Kenmerken van het proces betreffen bijvoorbeeld: of er een speciale organisatie-eenheid voor evaluatie is, of gewerkt wordt met evaluatieprogrammering en -budget, wat de beschikbare tijd en capaciteit is en tevens, wat geregeld is omtrent het gebruiken en leren van evaluaties. Randvoorwaarden voor de evaluatie-inhoud zijn voorschriften of handreikingen voor de aanpak (bijv. evaluatiedesign), en inhoudelijke criteria zoals kwaliteit, relevantie of toegankelijkheid van wetgeving of beleid voor potentiële gebruikers.
Het doel van dit onderzoek was om een *beeld te krijgen van bestaande departementale praktijken rond het verrichten en gebruiken van evaluaties – met name wetsevaluaties.* Hier aan ten grondslag ligt een streven bij de aanvrager van voorliggend onderzoek, de directie wetgevingskwaliteitsbeleid (WKB) van het

Kenniscentrum Wetgeving en Juridische Zaken (KCWJ), naar meer vergelijkbaarheid tussen wetsevaluaties. Dit om te zorgen dat ervaringen *met* wetsevaluaties en kennis *uit* wetsevaluaties worden geaccumuleerd, opdat bredere lessen kunnen worden getrokken.

Onderzoeksvragen

1. Welke inzichten bieden internationale (a) literatuur en (b) normen en handreikingen met betrekking tot *evaluatievermogen*? (hoofdstuk 1 en 2)
2. Welke praktijkervaringen met evaluatievermogen zijn opgedaan door beleidsmedewerkers en wetgevingsjuristen, betrokken bij (wets)evaluaties? (hoofdstuk 3)
3. Welke inhoudelijke en methodologische eigenschappen van *ex post*-evaluaties en -wetsevaluaties acht men daarbij van belang? (hoofdstuk 3)
4. Welke lessen bieden de antwoorden op voorgaande onderzoeksvragen voor verdere gedachtevorming over de ontwikkeling van evaluatievermogen – vooral met betrekking tot wetgeving? (hoofdstuk 4)[144]

De volgende aspecten ontleend aan literatuur over evaluatievermogen vormden de leidraad voor dit rapport:

- Structuur: doelstellingen en organisatieonderdelen voor evaluatie
- Programmering en evaluatieverplichtingen
- Middelen voor evaluatie: financiële en andere
- Evaluatiekader: evaluatieprogramma's; richtlijnen bij de aanpak (ten aanzien van zowel proces als inhoud)
- Borging van evaluatiekennis en –ervaring (meerdere kenmerken)
- Bevordering van gebruik

In deze structuur zijn eerst een aantal internationale aanpakken en ervaringen op een rij gezet. Vervolgens zijn, aan de hand van *desk study* en een twintigtal gesprekken met 35 sleutelinformanten in de periode juni-oktober 2015, gegevens verzameld bij negen Nederlandse beleidsdepartementen. Kernbevindingen uit beide exercities passeren in dit hoofdstuk de revue, gevolgd door een slotbeschouwing. Ook al zijn de genoemde aspecten van evaluatievermogen leidend geweest, onze bedoeling was *niet* om een oordeel te vellen over de mate waarin departementen voldoen aan de aspecten. De beschrijving van evaluatievermogen en de doorwerking van dat vermogen in de departementale praktijk lopen soms in elkaar over. Zo beschrijven respondenten soms hoe nieuwe evaluatieprogramma's zijn 'geland' bij beleidsdirecties. Het meenemen van praktijkervaringen met evaluatievermogen past bij het open, lerende doel van deze studie. Het onderscheid tussen randvoorwaarden en praktijk is echter niet steeds expliciet benoemd – en is soms lastig te maken.

Politieke context en doelen van (wets)evaluatie

Evaluaties en wetsevaluaties vinden niet in een vacuüm plaats, ze worden verricht vanuit verschillende motieven. Soms zijn ze verplicht uit oogpunt van verantwoording of een wettelijke evaluatiebepaling, een andere keer komen ze van binnenuit, bijvoorbeeld om te leren met het oog op verdere beleidsontwikkeling. Weer andere geven antwoord op actuele vragen vanuit de Tweede Kamer, of hebben een strategische functie zoals uitstel van een ingrijpende beslissing. De politieke context bepaalt dus mede de motieven voor evalueren, die kunnen

[144] Het conceptrapport heeft de inbreng gevormd voor een expertmeeting met als doel de discussie en gedachtenvorming over de (eventuele) verdere ontwikkeling van evaluatievermogen – in het bijzonder rond wetgeving – aan te jagen.

doorwerken in de inhoud en het gebruik van evaluaties (Van Aeken, 2007; Bovens et al., 2008).
In de literatuurbespreking in het eerste hoofdstuk lieten we zien dat er twee ideaaltypische evaluatiemotieven zijn, namelijk (a) verantwoorden en (b) motieven die we samen 'leren' noemen – met name kennisvergroting of *enlightenment* en beleidsaanpassing of *embetterment* (Scriven 1994; Vedung 1997). Het motief om te leren komt vaak tot uiting in *formatieve* evaluatie waarbij bruikbaarheid het belangrijkste beoordelingscriterium is. Zijn we op de goede weg? In *summatieve* evaluaties (ibid.) wordt een 'eindoordeel' over de doeltreffendheid van een beleid gegeven, passend bij het motief van verantwoorden. Niet bruikbaarheid, maar objectiviteit en transparantie staan daarbij voorop. Vaak staat de doeltreffendheid centraal, het kunnen toeschrijven van resultaten aan beleidsinspanningen (attributie).
Verantwoording en leren kunnen overlappen. Zo kan een evaluatie die aanvankelijk werd verricht met een verantwoordingsmotief alsnog worden gebruikt om lessen voor beleid te trekken, en omgekeerd.
Het bouwen aan evaluatievermogen is uit te leggen als tegenwicht bieden aan korte-termijn politieke belangen, door een schil om het evaluatieproces aan te brengen. Door te werken met bijvoorbeeld een (meerjarige) evaluatieprogrammering en standaarden voor de evaluatie-inhoud kan weerstand worden geboden aan – onder meer – zaken als tijdsdruk en de politieke 'waan van de dag'. Maar, zeker op langere termijn vormt het politieke klimaat óók een belangrijke randvoorwaarde voor de verdere ontwikkeling van evaluatievermogen (Cooksy, 2012; Nielsen et al., 2011).[145]

4.2 Internationale ervaringen met evaluatievermogen (onderzoeksvraag 1)

In hoofdstuk 2 zijn, gestructureerd naar de zojuist genoemde aspecten van het conceptuele raamwerk, mogelijk leerzame internationale ervaringen met het bouwen aan evaluatievermogen (*evaluation capacity building,* ECB) op een rij gezet, waaronder met name evaluatiegidsen en andere instructies voor aanvragers en onderzoekers. Het betreft documenten van grotere internationale organisaties (Europese Commissie, VN, Wereldbank) en een aantal grotere, 'westerse' landen (Verenigd Koninkrijk, VS). ECB betreft in de praktijk vooral beleid in het algemeen, hoewel her en der wel enige aandacht uitgaat naar *regulatory evaluation*. De academische belangstelling voor evaluatie van wet- en regelgeving lijkt nog kleiner. Bij sommige internationale organisaties ligt de oorsprong van evaluatievermogen in een sterke verantwoordingsbehoefte. Denk aan aanzienlijke uitgaven aan leningen en ontwikkelingsprogramma's door bijvoorbeeld de Wereldbank. Maar juist hier is al snel ook een sterke beleidsinhoudelijk gedreven functie naar voren gekomen. Ook in het centrale overheidsapparaat van de VS en Canada vindt ECB al meerdere decennia lang plaats, wat in de literatuur wel is verklaard uit het feit dat sociale wetenschappers hier al langere tijd onderdeel van uit hebben gemaakt. Dit zou contrasteren met Europa, waar juristen aanvankelijk domineerden en waar bestendiging van evaluatiepraktijken sinds eind jaren '80 op gang is gekomen (o.a. Stern, 2009).

Evaluatie als onderdeel van het beleidsproces
Internationaal is de notie doorgedrongen dat de kwaliteit en het gebruik van evaluaties niet alleen afhankelijk zijn van wat de onderzoekers opleveren, maar juist

[145] Dit vertoont analogie met het Mattheüs-effect in de sociaalwetenschappelijke literatuur: daar waar de (politieke) randvoorwaarden vooraf gunstig zijn, is vaak al een basisniveau van evaluatievermogen aanwezig en genereren nieuwe inspanningen bovendien meer opbrengsten dan waar gunstige randvoorwaarden niet vooraf aanwezig zijn.

ook van *de aanvragers* ervan – hier met name de beleidsmedewerkers en soms wetgevingsjuristen. Dit blijkt zowel uit handboeken te gebruiken bij opdrachtformulering, aanbesteding en benutting als uit normen en *ethical guidelines*, specifiek gericht op aanvragers (o.a. in de EU, de VN, de Wereldbank, Verenigd Koninkrijk en ook in ons land).
Aanvragers nemen voor evaluaties bepalende beslissingen, bijvoorbeeld over evaluatiebudget, -reikwijdte en -vragen en in de verdere begeleiding van het evaluatieproces.
Niet alleen als een evaluatie op handen is, maar ook in het beleidsproces eraan voorafgaand worden – doelbewust of niet – de randvoorwaarden voor evaluaties bepaald. Evaluatie wordt steeds vaker beschouwd als een geïntegreerd onderdeel van de beleidscyclus of *regulatory cycle* en minder als een daarvan los staande activiteit. De inhoud en de opzet van het te evalueren beleid en de onderbouwing ervan door de aanvrager, ziet men steeds vaker als randvoorwaarden voor evaluatiekwaliteit. Ontbreekt bijvoorbeeld de argumentatie voor beleidskeuzes (beleidstheorie) dan wordt het voor aanvragers en onderzoekers lastiger om focus aan te brengen in een latere evaluatie.

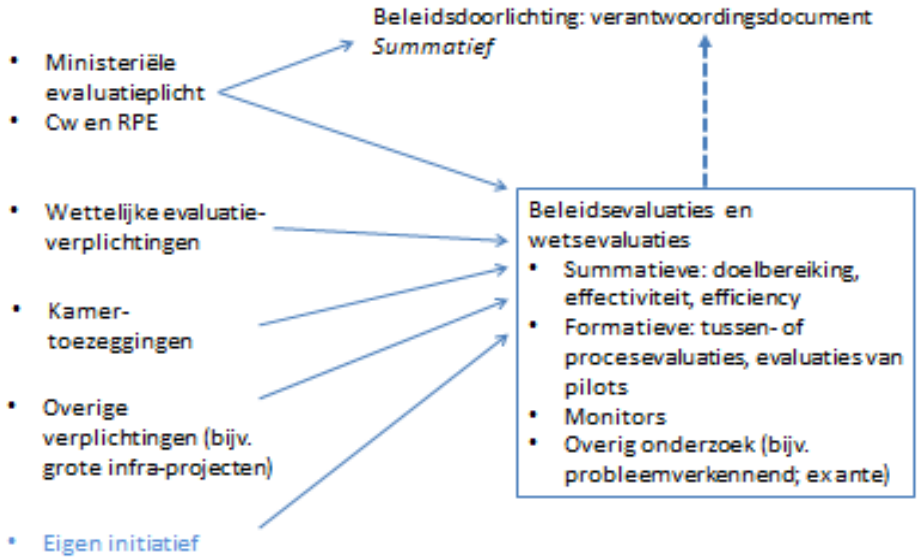
Evaluatiedesigns: stromingen en fit-for-purpose
In handreikingen voor de evaluatieaanpak zijn stromingen te onderscheiden; illustratief is het onderscheid tussen de 'attributieve' (*black box*) en de 'contributieve' (*clear box*) stroming. In de eerste staat het kunnen toeschrijven van resultaten aan beleid centraal: heeft het gewerkt en in welke mate? Dit past goed bij het evaluatiedoel van verantwoording. Nulmetingen en controlegroepen zijn onlosmakelijk verbonden met de attributieve benadering, omdat dit type evaluatiedesign zich toespitst op het uitschakelen van externe invloeden op de beleidsuitkomsten. Zo wordt de geïsoleerde, netto-invloed van beleid benaderd. De tweede stroming is gericht op het begrijpen hoe eventuele resultaten zijn geboekt. Wat werkt *hoe* en waarom/in welke contexten, of onder welke condities? Deze vragen worden vaker geassocieerd met formatieve evaluatie en kunnen op nagenoeg elk moment in de beleidscyclus worden toegepast. Vooral het afgelopen decennium zijn ook de evaluatiebenaderingen in deze 'familie' doorgedrongen tot evaluatiesystemen wereldwijd.
De stromingen sluiten elkaar niet uit, ze kunnen wederzijds aanvullend worden gebruikt. Een uitgangspunt dat stromingen overstijgt en steeds vaker opduikt, is 'fit for purpose'. De keuze voor het 'juiste' design is afhankelijk van kennisbehoefte en situatie, o.a. soort beleid, beleidsfase, evaluatieonderwerp, hoofdvragen en soort doelgroep. Verder stuitten we op relatief nieuwe benaderingen, zoals *real-time evaluation* op basis waarvan relatief snel kan worden bijgestuurd.
Meerdere internationale organisaties en landen vinden het een gemiste kans voor kwalitatief hoogwaardige evaluaties, als aan de kant van de aanvrager (1) niet tijdig wordt begonnen met nulmetingen, (2) het formeren van experimentele en controlegroepen niet of te laat gebeurt, of (3) als systemen van monitoring, gericht op proces- en uitkomstindicatoren achterwege blijven. Concrete handvatten volgen, hoe tot verbeteringen op deze punten te komen. Bij sommige typen beleid worden bijvoorbeeld soms groepen in- en uitgesloten op basis van een bepaalde 'drempelwaarde'. Door in de evaluatie twee groepen, bijvoorbeeld personen, bedrijven of gemeenten te betrekken die nét-wel en nét-niet voldoen aan de criteria voor bijvoorbeeld een milieusubsidie, zijn relatief eenvoudig een experimentele en een vergelijkbare controlegroep te vormen. Ook in Nederland is dit al gedaan.
Verder kunnen doelgroepen, zoals gemeenten, burgers en bedrijven een actieve rol vervullen in de registratie van output en eindresultaten en zo bijdragen aan monitoring en evaluatie.
De visie op evaluatie als integraal onderdeel van het beleids- en wetgevingsproces werkt door in die op het *gebruik* ervan; steeds vaker beschouwen grote instituties

evaluaties als méér dan alleen een eindrapport. Zo voorziet het *Smart Regulation* programma van de Europese Commissie op papier in terugkoppeling van evaluatieresultaten naar de regelgevingscyclus via het instrument van de verplichte ex ante toetsing. Bij de UNEG, de evaluatiegroep van de VN, is gebruik van evaluaties geprotocolliseerd tot en met de implementatie van resultaten in actieplannen en het aanleggen van evaluatie-'kennisbanken'. Tegelijkertijd blijkt uit de update van de *Evaluation Atlas* op basis van individuele OESO-landen dat systematisch evaluatiegebruik zeer lastig te bewerkstelligen is (Jacob et al., 2015).
Het sterkere accent op de aanvrager neemt niet weg dat de evaluatieonderzoeker nog steeds als belangrijke schakel wordt gezien in de bevordering van gebruik. Dit blijkt bijvoorbeeld uit leidraden voor onderzoekers om bevindingen bondig en toegankelijk te presenteren. Voorts blijkt het uit competentiebeschrijvingen: niet alleen moeten evaluatoren methodologisch-inhoudelijk competent zijn, maar ook interpersoonlijk en communicatief vaardig, flexibel en gevoelig voor de politieke en organisatorische context.
Behalve op aanvragers en onderzoekers wordt internationaal gewezen op het belang van 'aanjager'-instituties in de bevordering van evaluatievermogen van buitenaf, zoals evaluatiegemeenschappen, kennisinstituten en parlementen.

Werkt evaluatievermogen?
Hoewel de literatuur over evaluatievermogen omvangrijk is, is vooralsnog niet wetenschappelijk bewezen dat ECB tot betere beleidsresultaten leidt. Wel zijn uiteenlopende (positieve) case studies opgeleverd. Ook is gewaarschuwd voor neveneffecten, zoals dat het aanleggen en handhaven van standaarden opportunistisch gedrag of ritualisering in de hand kan werken waarbij het eigenlijke doel uit het oog wordt verloren.

4.3. Praktijken en ervaringen bij Nederlandse beleidsdepartementen (onderzoeksvragen 2 en 3)

In ons land kunnen twee niveaus van evaluatievermogen worden onderscheiden (figuur 2). Het eerste niveau betreft het begrotings- en verantwoordingsproces en vertegenwoordigt het centrale evaluatiestelsel.



*Bovenin de figuur zouden naast de beleidsdoorlichtingen ook de interdepartementale beleidsonderzoeken (IBO's) passen. Ook deze steunen vooral op bestaand onderzoek. IBO's vinden per jaar ongeveer vijf keer plaats en zijn erop gericht om met meerdere departementen samen na te denken over alternatieven voor huidig beleid, met als invalshoek bezuinigingen cq, het vergroten van *value for money* (zie ook de Box in Bijlage 4).

Figuur 2 Samenvattende schets van het centrale evaluatiestelsel*

In de jaarlijkse begroting wordt een departementale evaluatieplanning opgenomen. Veruit het belangrijkste verantwoordingsinstrument is de beleidsdoorlichting op begrotingsartikelniveau, die elke 4 tot maximaal 7 jaar dient plaats te vinden en aan de Kamer wordt gestuurd. Voor de inhoud van de beleidsdoorlichtingen – een synthesedocument – zijn de betreffende beleidsdirecties verantwoordelijk.
De beleidsdoorlichting, waar departementen op worden afgerekend, werkt door in de ambtelijke evaluatiepraktijken waar beleidsevaluaties plaatsvinden (het tweede niveau). Idealiter stapelt zich hier in de loop van elke 4 tot 7 jaar het basismateriaal op voor synthese in de uiteindelijke beleidsdoorlichting.
Figuur 2 laat zien dat verantwoorden – behalve begrotingstechnisch, ook politiek en naar aanleiding van wettelijke evaluatieverplichtingen[146]– een belangrijke drijfveer

[146] Wettelijke evaluatieverplichtingen zouden volgens een expert in naar schatting 20% van recentere wetten voorkomen (mondelinge comm. december 2013).

van evaluatieonderzoek is. Primair gericht op verantwoorden is het document van de beleidsdoorlichting zelf.
Standaarden hanteren bij (de beoordeling van) verantwoordingsevaluaties gebeurt binnen het centrale stelsel in toenemende mate, bijvoorbeeld door de Rekenkamer in 2011 en 2012 met meta-onderzoek naar doeltreffendheidsevaluaties van subsidies en vervolgens alle vindbare doeltreffendheidsevaluaties. Bovendien heeft het Ministerie van Financiën in 2016 een Handreiking beleidsdoorlichting geïntroduceerd.[147] Het synthesedocument van de doorlichting zelf moet aan de RPE-standaarden voldoen en tevens dient de doorlichting een kritische bespreking van het onderliggende evaluatiemateriaal te bevatten. Voor adequate verantwoording zou volgens Financiën strikt genomen moeten worden onderzocht of een wet of beleid de *oorzaak* is van resultaten in de praktijk. In de nieuwe Handreiking Beleidsdoorlichting klinkt echter niet alleen 'hard maken' van die oorzakelijkheid door maar tevens 'plausibel maken', bijv. met behulp van reconstructie van de aannames achter beleid (beleidstheorie).

4.3.1 Structuur: organisatieonderdelen voor evaluatie

Binnen beleidsdepartementen vervullen verschillende organisatieonderdelen een rol in het managen van evaluaties. Doorgaans ligt het initiatief voor ex post evaluaties - wetgevings- en beleidsevaluaties – bij de verantwoordelijke onderzoekscoördinatoren van beleidsdirecties of DG's. Bij de meeste departementen is er geen aparte evaluatieafdeling en zeker geen afdeling uitsluitend gericht op wetsevaluaties. Elk departement kent wel één of meerdere stafafdelingen units of directies die de evaluatiefunctie ondersteunen en stimuleren.
Een gemeenschappelijke is de directie Financiële en Economische Zaken (FEZ). Wat evalueren betreft, vullen directies FEZ hun rol per departement verschillend in – hoewel verantwoording uiteraard wel steeds het hoofddoel is. Zo heeft FEZ bij SZW een stimulerende en aanjagende rol in zowel de planning, de kwaliteitsborging als het gebruik van evaluatieonderzoek en beleidsdoorlichtingen, terwijl andere directies FEZ dichter op hun formele controletaken blijven. Binnen EZ is er sinds 2012-2013 een netwerk van organisatieonderdelen (waaronder BEC en de voorm. Regiegroep M&E) inclusief FEZ, dat zich specifiek richt op de kwaliteit van doeltreffendheidsevaluaties. Doel is het bevorderen van effectief beleid. Op het vlak van ex post evaluaties heeft FEZ bij IenM[148], naar aanleiding van een kritische inventarisatie van doeltreffendheidsevaluaties bij beleidsdepartementen (Algemene Rekenkamer, 2012, 2013), de regie meer in handen genomen waar het de planning van het evaluatieonderzoek betreft, vooral in relatie tot beleidsdoorlichtingen.
Soms is de stimulans en ondersteuning voor evaluatie- (en overig) onderzoek behalve bij FEZ ook bij andere onderdelen belegd, bijvoorbeeld in de 'kennishoek'.[149] Bij zowel EZ als bij SZW (*Chief Science Officer*) is in aanvulling op al genoemde organisatieonderdelen een hoogleraar aangetrokken voor extra inhoudelijke ondersteuning, bij OCW vervult de directie Kennis die rol. Een gezamenlijk doel is om de kwaliteit van onderzoek naar doeltreffendheid te stimuleren. Bij OCW is dat behalve *black box* nadrukkelijk ook *clear box* of

[147] Een ander, nieuw aspect van de RPE van 2015 en de nieuwe Handreiking is dat departementen veel aandacht dienen te besteden aan de hypothetische situatie dat *20% korting* op het beleidsbudget plaatsvindt. Doelmatigheid krijgt dus meer nadruk.
[148] Daar FMC geheten: *Financiën, Management en Control.*
[149] Bij BZK is een van de directie Kennis sinds 2011 nagenoeg ontmanteld. Hierdoor is sinds 2014 het zwaartepunt bij FEZ komen te liggen wat het stimuleren van de evaluatieprogrammering ten behoeve van de beleidsdoorlichtingen betreft. Bij IenM ligt het accent vooralsnog sterk op ex ante onderzoek. Beleidsmedewerkers worden bijgestaan door onder meer het Kennisinstituut voor Mobiliteitsbeleid (KiM) en het *Behavioural Insights Team* (BIT-IenM). Overigens heeft ook EZ een BIT.

verklarend onderzoek. Directies beslissen uiteindelijk of dergelijke designs daadwerkelijk worden ingezet.
Soms lopen evaluaties zo goed als standaard via een departementaal onderzoeksinstituut, zoals bij VenJ via het WODC. Dat besteedt een belangrijk deel van de aangevraagde evaluaties verder extern uit en begeleidt deze, terwijl het andere deel door eigen WODC-onderzoekers wordt verricht.
Behalve directies FEZ en instituten als WODC gaat het om relatief jonge constructies. Of en hoe deze doorwerken in planning, inhoud en gebruik van evaluaties (zie de volgende topics) valt hierdoor moeilijk vast te stellen.

Externe evaluaties
Lerende evaluaties worden (internationaal) vaker geassocieerd met intern onderzoek, terwijl men externe uitvoerders beter vindt passen bij evaluaties met een verantwoordingsdoel – onder meer uit oogpunt van objectiviteit (Bourgeois et al., 2011).
Evaluaties worden bij de Nederlandse beleidsdepartementen voor het overgrote merendeel uitbesteed aan externe partijen, zoals academische onderzoeksteams en onderzoeksbureaus. Een verklaring hiervoor is dat evalueren slechts één van de vele taken van beleidsmedewerkers is. Gebrek aan tijd en formatiecapaciteit bij beleidsdirecties maakt dat men het eenvoudiger en flexibeler vindt om financiële middelen voor evaluaties vrij te maken. Een andere reden is dat evalueren over het algemeen specialistische kennis en vaardigheden vereist, waar men binnenshuis niet altijd over beschikt. Ook de objectiviteit van extern onderzoek wordt genoemd. Bij uitbesteding wordt wel eens gewerkt met mantelcontracten, zoals bij OCW.

4.3.2 Evaluatieprogrammering

Een evaluatieprogrammering legt vast wanneer welke evaluaties staan gepland – bij voorkeur over een langere periode. Dit vermindert de kans dat evaluaties vergeten, of uit- of afgesteld worden – bijvoorbeeld, als gevolg van politieke of budgettaire prioriteiten. Idealiter helpt een programmering ook om tijdig te beginnen met de monitoring van begin- en tussenresultaten.
Elk departement is verplicht om jaarlijks (tenminste) één planning van alle soorten onderzoek vast te stellen en in te dienen bij Financiën. Deze wordt als bijlage bij de Rijksbegroting opgenomen. Het onderzoek is ingedeeld in drie categorieën: (1) beleidsdoorlichtingen (2) ander (evaluatie-)onderzoek naar doeltreffendheid en doelmatigheid en (3) overig onderzoek. Er is geen apart kopje voor wetsevaluaties.
Een belangrijk doel van evalueren is – zoals al genoemd – verantwoording van overheidsuitgaven in het kader van de begrotingscyclus. Idealiter is de evaluatieprogrammering zo opgebouwd, dat deze leidt tot evaluatierapporten waarin voldoende informatie staat om onderbouwde uitspraken over doeltreffendheid en doelmatigheid mogelijk te maken (Financiën, 2014, 2016). Voor veel betrokkenen blijkt dit in de praktijk een worsteling – wat zich volgens vertegenwoordigers van het centrale evaluatiestelsel weerspiegelt in de inhoud en kwaliteit van de beleidsdoorlichtingen. Niettemin stellen meerdere respondenten dat ook voortgang is geboekt. Er zijn de afgelopen paar jaar door meerdere directies FEZ en eventuele gerelateerde departementale organisatieonderdelen initiatieven genomen om kennis- en evaluatieprogrammeringen meer proactief te maken – vooral in het licht van de periodieke beleidsdoorlichtingen. We hebben niet kunnen vaststellen wat hier tot nu toe de vruchten van zijn.
De meerderheid van de respondenten gaf aan dat niet al het onderzoek dat in de evaluatie- en onderzoeksprogrammering van de Rijksbegroting staat, ook daadwerkelijk bruikbaar is bij de invulling van de beleidsdoorlichtingen. Bekeken vanuit het perspectief van kennisvergroting en tussentijds ‘leren’ voor beleid (formatieve evaluatie) is dit wellicht eerder een deugd dan een probleem.

Bij verschillende departementen en directies bestaat behalve de reguliere evaluatieprogrammering ook een kennisagenda. Ex post evaluatie speelt in de geraadpleegde kennisagenda's echter een verwaarloosbare rol.

De politieke context werkt in de praktijk door in de evaluatieprogrammering. Genoemde oorzaken van bijstellingen van evaluatieprogrammeringen zijn veranderingen in kabinetssteun voor een bepaald beleid, of nieuwe maatschappelijke ontwikkelingen. Sommige respondenten relativeren het feitelijke belang van een evaluatieprogrammering, door aan te geven dat evaluaties vooral worden gedaan als het bewindspersonen goed uitkomt of als er om wordt gevraagd. Gechargeerd spreken sommigen van evaluaties als politiek instrument (vgl. Bovens et al., 2008).
Ook het al of niet opnemen van een evaluatiebepaling in wetten (artikel 64 Ar) is onderhevig aan politieke onderhandelingen – ook internationaal is dit waargenomen. In hoeverre evaluatiebepalingen worden nageleefd hebben we niet nader onderzocht; hierover bestaan verschillende percepties. Bij OCW heeft DWJZ een kalender met daarin de evaluatietermijnen van wetten naar aanleiding van de evaluatiebepalingen.

Om werking en doeltreffendheid van een (wettelijke) interventie goed in kaart te brengen moet men idealiter beginnen met de gegevensregistratie en -verzameling al vóór invoering van de wet of maatregel, of meteen daarna. Respondenten stellen dat veel dossierhouders pas over de vraag- en probleemstelling van een evaluatieonderzoek beginnen na te denken op het moment dat de evaluatie in beeld komt – zoals bij veel wetsevaluaties tegen het aflopen van de in de evaluatiebepaling gestelde termijn. Op dat moment is het vaak te laat om tot een onderbouwde doeltreffendheidsstudie te komen. Het ontbreken van de hiervoor benodigde nulmeting en een eventueel monitoringsysteem wordt vaak geweten aan tijds- en capaciteitsrestricties – die op hun beurt kunnen samenhangen met ontwikkelingen in de politieke context. Volgens sommigen verklaart dit waarom veel ex post evaluaties *de facto* geen doeltreffendheids- maar proces- en doelbereikingsevaluaties zijn (vgl. Klein Haarhuis en Niemeijer, 2008 ten aanzien van wetsevaluaties). Omdat attributie aan beleid of wetgeving niet meer te realiseren is, beperkt men zich tot reconstructie van de implementatie ervan.

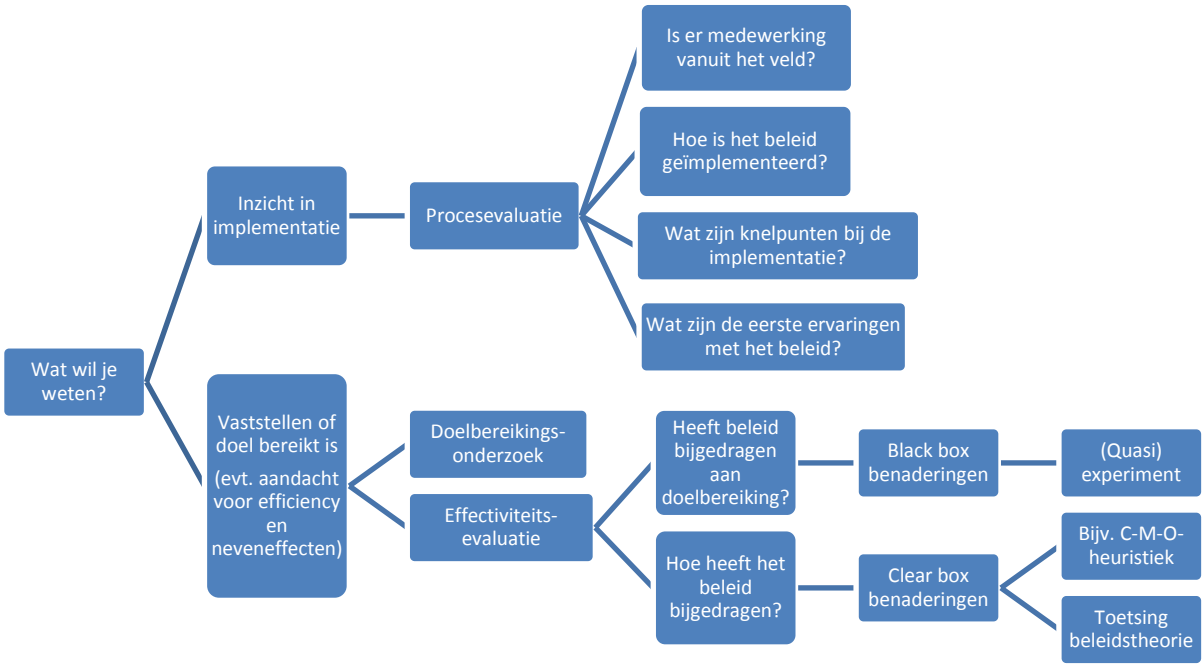
4.3.3 Middelen

Adequaat evaluatieonderzoek vergt voorts voldoende middelen. De meeste departementen stellen eerst een integraal budget voor onderzoek vast en daaronder vallen ook de evaluaties. Alle uitgevoerde evaluaties moeten binnen dat financiële kader passen. In de praktijk gaat men vaak uit van de bestedingsplannen, voortkomend uit de vastgestelde budgetten. Bij het uitzetten van evaluaties voor extern onderzoek bepaalt het departement het budget waarbinnen de offertes moeten blijven. Rijksonderzoeksinstellingen beheren vaak hun eigen onderzoeksbudget. We hebben niet onderzocht of middelen toereikend zijn of worden gevonden.
Wat de niet-financiële middelen betreft, melden veel respondenten dat binnen beleidsdirecties weinig *tijd* en personele capaciteit beschikbaar is voor evaluaties en onderzoek. Zo wordt de capaciteit voor de onderzoekscoördinatie door meerdere respondenten gering bevonden. Bij een zeker departement betreft die op papier 1 fte op een directie van 60 personen, maar komt het in de praktijk op minder dan 0,5 fte neer. Dit kan onder meer doorwerken in een laat tijdstip waarop evaluaties worden uitgezet, wat (zoals al aangegeven) doorwerkt in hun opzet en validiteit.

4.3.4 Evaluatiekader: inhoud en proces

Inhoudelijke keuzes, zoals omtrent het evaluatiedesign, zijn in Nederland vaak aan de externe 'aannemers' van evaluaties overgelaten. De laatste paar jaar zien we echter een ontwikkeling waarbij aanvragers (beleidsmedewerkers) vaker instrumenten krijgen aangereikt om evaluaties gerichter aan te sturen (o.a. bij EZ, SZW en OCW).[150] Deze aansturing betreft zowel het proces van evaluatie als de inhoud.
Dit past bij de internationaal gesignaleerde internationale trend dat de aanvrager een groter gewicht krijgt toebedeeld in de totstandkoming en de aanpak van evaluaties. Denkt de aanvrager vooruit, dan draagt dit bij aan de organisatie van robuuste evaluatiedesigns, met gebruikmaking van bijvoorbeeld nulmetingen, monitoring en eventuele controlegroepen.
Afhankelijk van de evaluatievraag resulteren verschillende typen evaluaties (Figuur 3).

Figuur 3 Evaluatievragen en bijpassende benaderingen



Bron: Klein Haarhuis en Oosterwaal (Pleio, 2014)

We vatten hier niet de evaluatiekaders van verschillende departementen samen, maar beperken ons tot enkele observaties en voorbeelden.[151] De meeste van de in hoofdstuk 3 bijeengebrachte inzichten betreffen doeltreffendheidsevaluatie – onderzijde figuur – hetzij via de eerder beschreven *black-box* (experimentele)

[150] De projectbegeleiders van EWB bij het WODC, dat onderzoek verricht voor het ministerie van VenJ nemen na goedkeuring van de (evaluatie)opdracht, de begeleiding goeddeels van de aanvrager over. Laatstgenoemde is wel vertegenwoordigd in de BC.
[151] Voor een *overzicht van programma's en handreikingen of richtlijnen verwijzen we naar* Tabel 8 in het derde hoofdstuk. Bijlage 4 bevat enkele uitwerkingen.

aanpakken, hetzij via *clear-box* benaderingen waarbij het begrijpen en verklaren van gebeurtenissen centraal staat.
De inhoudelijke evaluatiefocus verschilt per departement en sluit vaak aan bij de rol van de op het evaluatievlak meest actieve organisatieonderdelen. Zijn FEZ – en eventuele gerelateerde organisatieonderdelen – zeer actief, dan weerspiegelt zich dat bijvoorbeeld in een meer summatieve aanpak – heeft het gewerkt?. Bij EZ werkt men bijvoorbeeld met uitleg en checklists ten behoeve van evaluatieopzetten (procesmatig en inhoudelijk), offertes en offertevergelijking en rapporten. Zijn andere organisatieonderdelen actief, dan blijkt dat uit een sterkere aandacht voor verklarende evaluatiebenaderingen. Bij Kennis-OCW ziet men bijvoorbeeld de meeste kracht in een combinatie van black-box- en clear-box-aanpakken, men beschouwt deze als complementair.
Veel van de genoemde programma's en richtlijnen of handreikingen zijn nog vers, waardoor weinig zicht bestaat op de resultaten ervan. We weten dus niet of de beleidsdoorlichtingen minder lacunes zijn gaan bevatten (verantwoording), of dat de programmering en kwaliteit van evaluatieonderzoek in het algemeen is verbeterd. Niettemin hebben, juist ten aanzien van wetgeving, de beschreven programma's tastbare ervaringslessen opgeleverd. Wat de inhoud betreft leren de praktijken van zowel de tweede en derde Awb-evaluatie, het programma STEM als het programma evaluatie regelgeving (PER, ZonMw) dat *multidisciplinariteit* in de aanpak onontbeerlijk is. Dat vergt een combinatie tussen juridisch onderzoek naar gelding en empirisch onderzoek naar doelbereiking en doeltreffendheid. Heel vernieuwend is dit niet maar toch een bevestiging. ZonMw is het gelukt om tot zeven algemene aanpakcriteria bij wetsevaluaties te komen – die in de PER-praktijk al langere tijd worden toegepast.

(1) Wetsevaluaties moeten door een multidisciplinaire groep worden uitgevoerd met aandacht voor zowel het empirische als voor het juridische deel;
(2) kennisnemen van de beginsituatie – zo mogelijk een nulmeting – moet onderdeel zijn van de evaluatie. Hieronder schaart men ook de beleidstheorie;
(3) wetsevaluatie beperkt zich niet tot de te evalueren wet, maar houdt ook rekening met de internationale en Europese context;
(4) wetsevaluaties kijken zowel naar gewenste als ook ongewenste effecten;
(5) wetsevaluaties moeten aandacht schenken aan de relatie tussen wetgeving en zelfregulering – een criterium verbonden aan het subsidiariteitsbeginsel;
(6) elke nieuwe evaluatie moet naast nieuw onderzoek ook gebruik maken van bestaande kennis, ter voorkoming van verspilling van tijd en geld; en tot slot
(7) wetsevaluaties besteden aandacht aan diversiteit en aan de toegankelijkheid van alle bevolkingsgroepen tot de wet.

Enkele in het derde hoofdstuk beschreven initiatieven sluiten naadloos aan bij de internationaal toegenomen nadruk op de verantwoordelijkheid van de aanvrager voor adequaat evalueren, niet alleen in het evaluatieproces zelf maar reeds bij de onderbouwing van het beleidsinstrumentarium en met de tijdige aanleg van nulmeting- en monitoringsystemen. Bij EZ zijn evaluatie-handvatten geïntegreerd in een gids voor beleidskwaliteit – het overkoepelende doel is: effectiever beleid bewerkstelligen. Deze gids beslaat alle fasen van de beleidscyclus, van probleemanalyse en instrumentkeuze en ex ante gevolgenbeoordeling, tot en met politieke beslissing, uitvoering en – uiteindelijk – ex post evaluatie. Zo bevat de gids aanwijzingen en tips bij het opstellen van (a) vroege beleidsexperimenten en (b) het opzetten van een monitoringsysteem, met concrete voorbeelden. Dit benadert sterk het internationale 'ideaal' – en dat van de aanvrager van onderhavige studie – dat de beleids- of wetgevingscyclus via evaluatie en terugkoppeling wordt rondgemaakt. Het in 2011 gelanceerde IAK wordt (nog) niet vaak genoemd. Bij IenM probeert men sinds kort bij beleidsdoorlichtingen terug te grijpen op resultaten van het ex

ante gebruikte IAK voor beleid en wetgeving. Ook dit kan een signaal zijn van een meer 'integrale' kijk op beleid en evaluatie.

4.3.5 Borgen en verankeren van evaluatiekennis en –ervaring

Individuele kennis bij de aanvrager van wetsevaluaties is nog geen collectieve kennis. Eerder beschreven we hoe, via de verschillende organisatieonderdelen, procesmatige en methodologische evaluatiekennis bij de aanvrager terecht zou moeten komen. Evaluatievermogen gaat daarnaast ook over de mate waarin en wijze waarop: (1) het evaluatiekader binnen de organisatie wordt gecommuniceerd, (2) hoe men zorgt dat men de benodigde kennis voor een evaluatieproject bijeen krijgt, en (3) hoe eenmaal opgedane kennis en ervaring wordt verankerd en gedeeld. Beleidsdepartementen en –directies lijken sterk te verschillen in deze opzichten.

Communicatie van het evaluatiekader
De communicatie van bijvoorbeeld handreikingen of hulpmiddelen richting aanvragers hebben we niet uitgebreid onderzocht. Enkele malen signaleerden we dat bestaande handreikingen evaluatieonderzoek niet bekend waren bij een coördinator onderzoek of de wetgevingsdirectie. Bij EZ heeft men een nieuwsbrief en een *quick response team* dat het (betrekkelijk nieuwe) evaluatiekader helpt uitdragen. Verder zijn een klein aantal departementale of departementsoverstijgende cursussen ontwikkeld die geheel of deels over evaluatie gaan en beschikbaar zijn voor beleidsmedewerkers en wetgevings- en overheidsjuristen.[152] Net als in de eerder genoemde gids maakt ook hier evaluatie soms onderdeel uit van een breder curriculum met betrekking tot (evidence-based) beleid maken.
Ook richting de onderzoekers is communicatie van het evaluatiekader vereist, zo erkennen enkele respondenten. Dit is des te meer het geval als op 'harde' designs of methoden wordt ingezet waarvoor specialistische kennis en vaardigheden vereist zijn.

Selectie van onderzoekers
Het borgen van de kwaliteit van evaluaties gebeurt voornamelijk via de selectie van onderzoeksteams. Wat de opleiding en kennis van de uitvoerende onderzoekers van wetsevaluaties betreft, stellen beleidsdirecties over het algemeen geen strenge of vaste eisen. Competitie tussen offrerende partijen bij evaluatieprojecten boven de aanbestedingsgrens fungeert als een belangrijk selectiemechanisme.
Bij alle bezochte departementen zei men veel waarde te hechten aan de *multidisciplinariteit* van het onderzoeksteam. Dit is bij wetsevaluaties, idealiter een combinatie van juridische en sociaalwetenschappelijke aspecten, extra relevant – zoals duidelijk gebleken is uit de lessen die aanvragers en onderzoekers hebben getrokken uit praktijkervaringen met STEM, Awb en ZonMw. Projectbegeleiders van WODC-EWB werken met interne afspraken over de samenstelling van onderzoeksteams; zo is bij de vierde evaluatie van de Awb een multidisciplinair team betrokken. Bij ZonMw is het de eerste van de zeven vereisten dat wetsevaluaties door een multidisciplinair team worden gedaan.
WODC-EWB (VenJ) heeft een database opgezet met contactpersonen en expertise van alle onderzoeksbureaus en –groepen, universitair, particulier of anderszins, die ooit een offerte hebben ingediend of een evaluatie hebben gedaan. Het kost wel

[152] Voorbeelden zijn een masterprogramma voor OCW-onderzoekscoördinatoren van TIER, het instituut voor *evidence-based* onderwijsonderzoek, een cursus beleidsevaluatie bij EZ en inmiddels Rijksacademie, en het cursusprogramma van de Academie voor Wetgevings- en overheidsjuristen. Verder, aan 'de ex ante kant', bestaan rijksbrede cursussen over het IAK en over gedrag & beleid, zoals bij VenJ – dat deels ook over evaluatie gaat.

veel moeite om de database in een dynamische personele omgeving up-to-date te houden. Bij SZW is steeds vaker het streven dat hoogleraren en/of ervaren onderzoekers in het onderzoeksteam betrokken zijn.

Begeleidingscommissies
Om de kwaliteit van (wets)evaluaties verder te bevorderen, wordt veel gewerkt met begeleidingscommissies. Deze monitoren het verloop en de voortgang van het evaluatieonderzoek en komen een paar keer tijdens het traject bijeen. Door respondenten genoemde functies van begeleidingscommissies zijn kwaliteits- en voortgangsbewaking door het stellen van kritische vragen, het tijdig ingrijpen en bijsturen als het onderzoek niet goed verloopt. Ook het creëren van draagvlak voor het onderzoek wordt genoemd. Aan tafel zit meestal iemand namens de opdrachtgever, een vertegenwoordiger van de uitvoeringspraktijk en inhoudelijke experts.
Bij de meeste departementen bestaan bij ex post evaluatieonderzoek geen vaste richtlijnen voor de samenstelling van begeleidingscommissies, met uitzondering van WODC-EWB, waar een senior-wetenschapper standaard voorzitter is, en ZonMw waar dit een CER-lid is. Bij beleidsdoorlichtingen neemt de directie FEZ meestal zitting in de begeleidingscommissie of klankbordgroep.

Ervaringen verankeren
Via de verschillende organisatieonderdelen dient procesmatige en methodologische evaluatiekennis bij de aanvrager terecht te komen. Dit loopt veelal via de onderzoekscoördinatoren bij DG's of directies, die – bij sommige onderzochte departementen – extra inhoudelijke ondersteuning kunnen vragen, bijvoorbeeld van: de CSO bij SZW of de Regiegroep M&E (inmiddels BAT) die een vertegenwoordiger in elk van de B&I- beleidsdirecties heeft – en de directie Kennis van OCW.
In een klein aantal gevallen worden beleids- en andere medewerkers of projectbegeleiders voorafgaand aan hun aanstelling *geselecteerd* op ervaring met het uitbesteden en begeleiden van evaluatieonderzoek; dit is bijvoorbeeld het geval bij de projectbegeleiders van WODC-EWB en bij FEZ-SZW. Meestal is dit echter niet het geval, omdat evaluatie niet als *core business* wordt beschouwd.
Ter vervanging of ter compensatie van bestaande structuren vindt vaak informeel overleg plaats – bijvoorbeeld tussen junior-dossierhouders en onderzoekscoördinatoren of met dan wel binnen kennisdirecties en -instituten. Op deze manier worden ervaringen met evaluaties alsnog doorgegeven. Toch weten medewerkers binnen departementen vaak meer van evaluatietrajecten en de praktische en concrete uitdagingen in het uitbestedingsproces, dan wat geregistreerd en benut wordt. Het is een uitdaging gebleken om evaluatiekennis en -ervaring aanwezig in de hoofden van individuele dossierhouders in het collectieve geheugen vast te leggen en voor andere projecten te benutten. Met het vertrek van personen kan die bijvoorbeeld uit het departement verdwijnen. Voor de langere termijn kan men zich daarom de vraag stellen of informeel advies inwinnen en ervaring delen toereikend is – hoe flexibel en nuttig vaak ook.
Een van de verklaringen voor dit geringe institutionele geheugen – en leren – is de beperkte tijd en capaciteit voor bijvoorbeeld coördinatie van onderzoek. Het ontwikkelen van specialisaties is bovendien lastig bij een lage evaluatie-omzet; alleen bij een op projectbegeleiding gerichte afdeling, zoals EWB bij het WODC lukt dit nog aardig.
De Interdepartementale Begeleidingsgroep Prestatiegegevens en Beleidsevaluatie (IBP), onder coördinatie van het ministerie van Financiën, is een forum waarin FEZ-vertegenwoordigers van elk departement samen met Financiën werkzame ervaringen met en aanpakken van evaluaties uitwisselen. Meerdere (FEZ) respondenten bestempelden dit als nuttig.

4.3.6 Bruikbaarheid en gebruik

Men kan nog zoveel doen om randvoorwaarden voor evalueren te scheppen, als de resulterende evaluaties niet, of alleen incidenteel worden benut voor kennisvergroting of beleidsaanpassing, zal menigeen de inspanningen als zinloos afdoen.
Hoewel evaluatiegebruik in de praktijk een rommelig proces is (o.a. Hertin et al., 2007) maken we onderscheid tussen bruikbaarheid en daadwerkelijk gebruik.

Bruikbaarheid
Contact en overleg tussen aanvrager en onderzoekers gedurende het evaluatietraject – al dan niet via de begeleidingscommissie – is niet alleen volgens de literatuur maar ook volgens respondenten een belangrijke manier om de bruikbaarheid en de kans op gebruik van onderzoek te vergroten, maar ook volgens respondenten. Een enkele keer vonden voorafgaand aan publicatie zelfs al beleidsaanpassingen plaats. Net als bij grote instituties en landen ziet men gebruik kennelijk als een resultaat van het hele evaluatieproces en niet van alleen een eindrapport.
Verder is de timing van de oplevering van evaluaties – in relatie tot besluitvorming in departement of parlement – natuurlijk van groot belang voor hun bruikbaarheid.

Gebruik
Over het algemeen hanteert elk onderzocht departement formeel een vergelijkbare wijze van communicatie en (korte-termijn) kennisname en gebruik van evaluatieonderzoek. In de regel worden rapporten openbaar gemaakt en vaak ook naar de Tweede Kamer gestuurd – in veel gevallen met een beleidsreactie. Sommige departementen doen iets extra's, zoals monitoring van gebruik (WODC, FEZ-BZ) en het nemen van acties gericht op het vergroten ervan (BZ-IOB).
Door onderzoekers vindt daarnaast vaak nog verspreiding plaats in de vorm van academische en vakpublicaties. Een andere vorm van valorisatie is het geven van presentaties en toelichtingen op de departementale werkvloer.
Na publicatie en eventuele valorisatie is het aan beleidsmedewerkers en politiek wat er verder met de bevindingen gebeurt. Bij departementen zijn geen vaste of voorgeschreven constructies gevonden voor het gebruiken van evaluaties, laat staan voor terugkoppeling naar de beleidscyclus. In geval van subsidieregelingen is er een horizonbepaling die stelt dat de regeling na vijf jaar vervalt. Voortzetting kan, mits voldoende gemotiveerd, bijvoorbeeld op basis van doeltreffendheidsonderzoek.
Hoewel verplichte evaluaties vaak wel afgerond worden, komt het voor dat ze na verloop van tijd als minder politiek relevant worden gezien. Niettemin percipiëren enkele respondenten toenemende belangstelling bij Tweede Kamerleden in ex post evaluaties en in wat de minister ermee wil gaan doen, met name in het licht van de verantwoordingsfunctie en financiële lasten.

Leeropbrengsten
Om de leerfunctie van evaluatie een belangrijker plaats te geven, zouden ex post-evaluaties volgens sommige gebruikers meer toekomstgericht moeten zijn. In plaats van alleen de vraag te beantwoorden in hoeverre beleid of wet de gewenste resultaten hielp te bereiken, zou daarnaast moeten worden besproken of het *nog zinvol is* binnen het stelsel. Respondenten stelden dat zij waarde hechten aan evaluaties die in staat zijn nieuwe discussies te starten over hoe een bepaald maatschappelijk probleem kan worden aangepakt. Met hun retrospectief, met inbegrip van reconstructies van bijvoorbeeld de implementatie en het bereiken van doelgroepen, bieden ex post evaluaties volgens sommigen slechts beperkte handvatten voor toekomstige beleidsaanpassingen (vgl. Van Gils en Leeuw, 2010).

‘Nieuwe’ kennis over hoe gedrag te beïnvloeden heeft in dat opzicht meer te bieden, ook omdat die vaak beter te generaliseren is.
Een tweede leerfunctie van (wets)evaluaties is volgens respondenten (niet zijnde aanvragers) dat ze de impliciete aannames waarop wetgeving of beleid rust, boven water halen en toetsen aan de praktijk.
‘Lerend’ gebruik van evaluaties valt om uiteenlopende redenen niet af te dwingen, maar blijft nastrevenswaardig zolang niet-gebruik toe te schrijven is aan de bruikbaarheid van studies, of aan gebrekkige communicatie met aanvragers. Kennis uit evaluaties zinkt vaak nog weg, terwijl ze later weer nodig kan worden. Dit maakt het raadzaam om te investeren in (thematische) bundeling van evaluatieresultaten in meta-onderzoeken en resultaatoverzichten.[153] Bij VenJ/WODC is onlangs bijvoorbeeld de interactieve Kennisbank justitiële interventies in gebruik genomen. Ook de evaluaties zelf kunnen thematisch worden ingericht, mede om ritualisering, maar ook overbelasting van de beleids- en wetgevingscyclus, tegen te gaan (vgl. de thematische wetsevaluaties van ZonMw).

Box 11 Globale vergelijking internationale en nationale observaties

Zowel internationaal als in ‘Den Haag’ is een accentverschuiving zichtbaar van de onderzoeker naar de verantwoordelijkheid die de evaluatie-aanvrager heeft in het tot stand brengen en gebruiken van evaluaties. Althans, in de rolomschrijvingen op papier.
Dat de helderheid en de onderbouwing van beleid en wetgeving de latere kwaliteit van evaluaties beïnvloeden, komt in de Nederlandse praktijk vooralsnog wat minder uit de verf. Zo ver als de Europese Commissie, die stelt dat het formuleren van minder doelstellingen van beleid bijdraagt aan gefocuste en budgettair proportionele evaluaties, gaat men in Den Haag bijvoorbeeld (nog) niet. Maar recente ‘integrale’ hulpmiddelen zoals een EZ-gids en enkele cursussen voor aanvragers, zijn wel aanwijzingen in die richting. Hierin wordt evaluatie geïntegreerd behandeld met de andere onderdelen van de beleidscyclus, met effectiever beleid als einddoel. Ook is door sommige respondenten een relatie gelegd tussen evaluaties en beleidsdoorlichtingen en de aandachtspunten van het ex ante IAK.

4.4 Evaluatievermogen in context

4.4.1 Politiek en verantwoording

De politieke realiteit van alledag werkt volgens de literatuur door in de behandelde aspecten van evaluatievermogen. Zo kan de politieke dynamiek strategisch of legitimerend evalueren in de hand werken. Is bij de politieke of ambtelijke leiding bijvoorbeeld weinig ruimte voor tegenspraak – of wordt dit alleen al zo gepercipieerd door ambtenaren – dan zullen evaluaties gericht op kennisvergroting en beleidsverbetering geen voet aan de grond krijgen.[154] Op de langere termijn kan dit gevolgen hebben voor het evaluatievermogen, zoals in termen van organisatiecapaciteit, budget en tijd.
Dat aanhoudende politieke druk op langere termijn een remmende werking kan hebben op de ontwikkeling van het evaluatievermogen, hebben we met dit onderzoek niet hard kunnen toetsen, maar op basis van de interviews zijn er aanwijzingen voor. De kracht ervan hangt samen met de evaluatieonderwerpen en hoe politiek of maatschappelijk gevoelig die liggen. Sommige beleidsdirecties

[153] In lijn met de in Hoofdstuk 2 aangehaalde internationale initiatieven van bijvoorbeeld EPPI, UNEG en Campbell, en met adviezen van het Panel van Advies bij BZ-IOB (2011, 2013, 2014).
[154] Dit ligt wel genuanceerd. Er zijn meerdere lagen: de politieke leiding legt weer verantwoording af aan de Kamer. De Kamer kan soms hameren op lerende evaluaties, maar juist ook rappe koerswijzigingen in beleid aanjagen, waarbij de stap van evaluatie wordt overgeslagen.

worden notoir 'politiek' genoemd en weinig vatbaar voor empirische evidentie; bij andere beleidsdirecties ligt het anders.
Volgens sommigen werkt de politiek door in de evaluatieprogrammering, met name in aanpassingen en in de naleving ervan. Niettemin kan de politieke dynamiek óók aanleiding geven tot *intensievere* monitoring en evaluatie. Bij het innovatiebeleid (topsectoren) zijn rond 2010 expliciete beslissingen genomen om het evaluatievermogen bij EZ voor de langere termijn te verstevigen – resulterend in onder meer het beschreven evaluatievermogen daar.[155]
Een randvoorwaarde voor 'lerend evaluatievermogen' lijkt dat initiatieven ter vermeerdering van beleidsrelevante kennis niet worden afgestraft, maar ondersteund door de politiek en door de instrumenten in het centrale evaluatiestelsel. En aantal respondenten betreurt de nadruk die met het centrale evaluatiestelsel en de beleidsdoorlichtingen op verantwoording is komen te liggen. Van plan- en procesevaluatie – en vele andere soorten beleidsonderzoek – kan bijvoorbeeld veel geleerd worden, ondanks dat de 'hamvraag' van doeltreffendheid en doelmatigheid buiten beeld blijft.
Nastrevenswaardig is als leren en verantwoorden in zekere mate in elkaars verlengde liggen – en '*controlitis*' of evaluatie-ritualisering beperkt blijven. Ondanks dat het leerelement de laatste jaren wel meer wordt benadrukt, ook in de nieuwe handreiking beleidsdoorlichtingen (Ministerie van Financiën, 2016), lijkt het instrument van de beleidsdoorlichtingen voor departementen op dit moment niet de belangrijkste prikkel voor werkzaamheden op het vlak van *enlightenment* en *embetterment*.[156] Perverse effecten liggen op de loer: de huidige verantwoordingssystematiek kan op inhoudelijk gedreven evaluatiewerk een uithollende uitwerking hebben (vgl. Schwartz & Mayne, 2005). Ook kan het naleven van verantwoordingsplichten veel tijd vergen, tijd die anders mogelijk aan andere evaluatievormen zou zijn besteed.
Veel beleidsmedewerkers staan niet te springen om kwetsbaarheden in de implementatie of werking van beleid aan de Kamer te tonen. Op dit punt is wel gesuggereerd dat Kamerleden zelf ook anders naar evaluaties zouden kunnen kijken. Behalve erop afgerekend te worden, zouden opstellers van evaluaties en beleidsdoorlichtingen parlementaire waardering moeten oogsten als in het lange termijn-streven naar resultaten kwetsbaarheden en verbeterpunten kenbaar worden gemaakt. Een dergelijke 'verschuiving' in de politieke context zou meer openheid bij beleidsmakers teweeg kunnen brengen en meer ruimte laten voor lerend evalueren.

4.4.2 Evaluatiebepalingen in wetten

Net als verantwoordingsplichten kunnen evaluatiebepalingen in wetten leren soms stimuleren, maar ook leiden tot ritualisering. Tegenwicht kan bieden dat wetgevingsjuristen de evaluatiebepaling bij het maken van wet al meer in de richting van lerend evalueren formuleren. Dit in plaats van standaard 'doeltreffendheid/ efficiency' als het enige evaluatiedoel in de bepaling te kopiëren. Voorts zijn realistische evaluatietermijnen noodzakelijk om recht te doen aan de ontwikkeling van het beleid en het uitkristalliseren van doeltreffendheid en neveneffecten. Zo duurt het zeker twee jaar na uitrol voordat een betrouwbare recidivemeting kan worden uitgevoerd. Bij evaluaties van de implementatie en de

[155] De website *Volg innovatie* van de Rijksdienst voor Ondernemend Nederland (RVO) dient als invulling van de motie Koppejan c.s. (Tweede Kamer 2011-2012, 33 000 XIII nr. 25). Hier kan iedereen zien welk bedrijf innovatiegelden en -kredieten hebben ontvangen. Dit naar Amerikaans en Brits model.
[156] Naar de stand van de implementatie wordt bijvoorbeeld niet meer expliciet gevraagd (voorheen wel). Verder kan men zich de vraag stellen of de begrotingsdoelstelling wel het juiste richtpunt is voor verantwoorden; mogelijk zou dit thematischer kunnen.

voorlopige werking van de wet zou de termijn mogelijk wel korter kunnen dan bij summatieve evaluaties.
Verder draagt een adequate toelichting op de wet in de Memorie van Toelichting bij aan de werking van de evaluatiebepaling: wat is de bedoeling van de wetgever; wat zijn de aannames, welke resultaten van de wet worden op basis hiervan verwacht? Op welke punten kan een wetsevaluatie hier over x jaar aan bijdragen? WKB heeft inmiddels een schrijfwijzer voor Memories van Toelichting voltooid, die mede is bedoeld om de onderbouwing van wetgeving en beleid (beleidstheorie) nader te verhelderen.[157]

4.4.3 Organisatiecontext

Zoals al beschreven is het verankeren van inhoudelijke en evaluatie-gerelateerde kennis lastig gebleken, alleen al vanwege het verloop van ervaren beleidsmedewerkers. Dit is volgens enkele respondenten fnuikend voor het behoud van evaluatiekennis en -ervaring. Hier komt bij dat het evaluatiekader, in de zin van handreikingen om bepaalde werkwijzen tot stand te brengen, niet altijd goed 'landt' bij beleidsdirecties – en ook externe onderzoekers. Men is bijvoorbeeld nog onvoldoende vaardig of overtuigd om er gebruik van te maken. Hierdoor landen die initiatieven in een weinig vruchtbare omgeving. Nieuwe, geüniformeerde inkoopprocedures zijn ook een potentiële bedreiging voor de reeds ontwikkelde initiatieven.

4.5 Wetsevaluaties: een vak apart?

Is het nodig om wetsevaluaties een aparte plek te geven in de organisatie van evaluatievermogen? Wetten zijn sterk verweven met beleid. Dat de wet niet alleen empirisch of sociaalwetenschappelijk maar ook juridisch-georiënteerd onderzoek vergt, is toegelicht in het eerste hoofdstuk. Beleids- en FEZ-medewerkers zien wetsevaluatie echter niet als een afzonderlijke tak van sport, zo komt naar voren uit dit onderzoek. Dit geldt eveneens voor enkele respondenten bij wetgevingsdirecties. We hebben geen handreikingen aangetroffen specifiek voor wetsevaluaties. Ook onder academici en internationale organisaties is maar beperkte belangstelling voor evaluatie van wet- en regelgeving in het bijzonder.[158] Over nut en noodzaak hiervan lijken respondenten van mening te verschillen. Beleidsdirecties zijn in het huidige evaluatiesysteem *in the lead* en bepalend voor wat, wanneer en hoe geëvalueerd wordt – en tevens voor het evaluatiegebruik. Maar ook daar ontbreekt het – op zeker een aantal plaatsen – aan evaluatiekennis en -ervaring. Het lijkt ook om die reden zinvol om initiatieven in het kader van evaluatievermogen breder te trekken dan wetsevaluatie alleen.
Het bijzondere karakter van de wet als evaluatie-onderwerp is echter wel een belangrijk punt van aandacht. Wetgevingsjuristen en andere juridisch specialisten zijn van waarde voor evaluaties met een juridische component; zuiver sociaalwetenschappelijke aanpakken zijn dan te arm. Eerder genoemde praktijkervaringen met STEM, Awb en ZonMw onderstrepen het belang van een multidisciplinaire – juridische én sociaalwetenschappelijke – benadering van wetsevaluaties, blijkend uit in een gemengde samenstelling van het onderzoeksteam en een meerledig onderzoeksdesign.
Directies Wetgeving en/of juridische zaken lijken op grond van dit onderzoek echter niet of nauwelijks aangehaakt bij ex post beleidsevaluaties. Meerdere medewerkers

[157] www.kcwj.nl/kennisbank/schrijfwijzer-memorie-van-toelichting .
[158] Men kan zich afvragen: is het complex van regels en interventies dat 'wet' heet wel een geschikte analyse-eenheid voor stapeling van kennis? Bij enkelvoudige beleidsinterventies lijkt de 'vergelijkbaarheid' die de directie WKB nastreeft, eenvoudiger te realiseren.

van directies Wetgeving en/of juridische zaken herkenden het begrip beleidsdoorlichting bijvoorbeeld niet. Zij werken in een context die los lijkt te staan van het geschetste centrale evaluatiestelsel. Aan de voorkant, in het kader van ex ante onderzoek en IAK, zijn wetgevingsdirecties nadrukkelijker betrokken.
Met het programma *Versterking juridische functie Rijk* (2010) wordt aan diezelfde voorkant meer samenwerking tussen beleid en wetgeving beoogd. De enkele keer dat een wetgevingsdirectie een ex post wetsevaluatie coördineert heeft deze een primair juridische en niet sociaalwetenschappelijke inslag – uitzonderingen daargelaten. Vindt een wetsevaluatie vanuit beleidsdirecties plaats, dan ontbreekt juist vaak de juridische component (vgl. Klein Haarhuis en Niemeijer, 2008).
Gelet op het voorgaande zouden wetgevingsmedewerkers eerst meer betrokken kunnen raken bij wat beleidsdirecties en de onderzoekscoördinatoren doen op het vlak van evalueren, voordat wordt ingezet op een aparte lijn van 'wetsevaluatievermogen'. Dit is geen nieuwe conclusie, want een vergelijkbare werd ook in 1993-1994 getrokken naar aanleiding van het in hoofdstuk 1 genoemde Rekenkamerrapport: 'Wil wetsevaluatief onderzoek vaker een structureel onderdeel van het wetgevingsproces vormen, dan is het wenselijk de taakafbakening tussen wetgevingsafdelingen en beleidssectoren ook op dit gebied te expliciteren en aan de uitvoering van het onderzoek voldoende middelen toe te kennen. In de tussentijd mag van wetgevingsafdelingen worden verwacht dat zij meer activiteiten op dit terrein ontwikkelen, zoals het deelnemen aan structureel overleg op het gebied van evaluatie.'[159]

Het design van wetsevaluaties
Zoals aangegeven biedt de literatuur weinig handvatten die specifiek gericht zijn op de inhoudelijke aanpak van wetsevaluaties. Het lijkt daarom in elk geval raadzaam dat aanvragers zich laten inspireren door soortgelijke, al voltooide juridisch-empirische wetsevaluaties, zoals die bijvoorbeeld via de procedure van ZonMw zijn opgeleverd en gepubliceerd. De internationale en Nederlandse praktijk leerden dat een goed evaluatiedesign een design op maat is: toegesneden op, bijvoorbeeld, het type beleidsinstrumentarium, de fase van het beleid en de doelgroep(en).
Toch zijn een paar richtinggevende aanwijzingen uit het bijeengebrachte te halen. Zo geeft de Amerikaanse federale Rekenkamer (GAO, 2012) designtips om het maximale te halen uit de evaluatie van complexe federale programma's. Net als de wetgeving van eigen bodem zijn ook die programma's meestal geen 'cleane' interventies, en kan het lastig blijken om resultaten te isoleren van hun context om ze aan het beleid te kunnen toeschrijven. Met de GAO en ook Bussmann (2010) pleiten we hier voor wijze benaderingen in het evalueren van de complexen van regels en interventies, die wetten zijn. Juist gelet op die complexiteit is het raadzaam zo analytisch mogelijk te werk te gaan, om tot een maximaal begrip van wat een wet in de werkelijkheid teweeg brengt, te kunnen komen.

- Het verdient in het algemeen vaak de voorkeur om plausbiliteitsanalyse te verrichten, in plaats van proberen de (mate van) doeltreffendheid van een wet als geheel te bewijzen. Dit komt neer op het bijeen puzzelen van evidentie uit nieuw en bestaand onderzoek tot de *chain of effects*;
- Ook zou men, als de structuur van de wet zich hiervoor leent, zich kunnen richten op de belangrijkste met de wet gemoeide interventies, bijvoorbeeld in begrotings- of principieel opzicht. Dit was het geval in de WBSO-evaluaties, waarin men zich beperkt heeft tot het fiscale instrumentarium. Een dergelijke inperking tot saillante aspecten van de wet maakt ook experimenteren of quasi-experimentele designs mogelijk – passend in de *black box* stroming.

[159] TK 1993-1994, 23710, nrs. 1-2, p.30.

- Is het moment gepasseerd om nog een controlegroep te vormen dan kan men, onder voorwaarde van een monitoringsysteem, terugvallen op kwantitatieve tijdreeksanalyses om te toetsen of beleid of wet tot 'trendbreuken' heeft geleid. Bij een dergelijke aanpak dient zoveel mogelijk gecontroleerd te worden voor andere mogelijke verklaringen.
- Ten slotte kan men variatie tussen regio's, gemeenten of specifieke doelgroepen benutten, om beter te begrijpen hoe beleid in de praktijk is uitgepakt. Dit geïnspireerd op de methodiek van *multiple case study* onderzoek: de onderzoeker maakt gebruik van variatie in kenmerken tussen groepen (cases) die onder dezelfde wet of beleid vallen, om zo tot een beter begrip van werking en resultaten te komen. Wat zijn onderscheidende kenmerken van verschillende gemeentes of doelgroepen en hoe werkt dit door in de resultaten? Bij welke kenmerken / onder welke condities worden de meeste resultaten behaald en wat zijn plausibele verklaringen hiervoor? Dit past bij de eerder onderscheiden *white box*-stroming in evaluatiedesigns.

Kennisvergroting en geïnformeerde beleidsaanpassing komen uiteraard niet alleen via ex post evaluaties, de focus van dit onderzoek, tot stand. Leren kan deels ook worden bereikt met behulp van ex ante evaluaties en met vroege beleidsexperimenten of pilots. In geval van wetten kan experimenteerwetgeving ethische en andere knelpunten in het werken met experimentele en controlegroepen wegnemen. In een latere beleidsfase kunnen dergelijke 'vroege' studies bovendien de verantwoording ondersteunen.

4.6 Slotconclusies (onderzoeksvraag 4)

Het evaluatievermogen is – los van de basistaken van directies FEZ en de rol van onderzoekscoördinatoren – niet centraal geregeld, maar heeft per departement verschillend vorm gekregen. Dit betreft zowel actieve organisatieonderdelen als procesmatige en inhoudelijke evaluatie-initiatieven en -propgramma's. Veel van de in dit rapport beschreven constructies, procedures en aanpakken zijn nog vers. Hierdoor konden we niet nagaan in hoeverre ze structurele verbeteringen hebben opgeleverd in termen van leren en verantwoorden; laat staan of ze – het doel bij EZ aanhalend – uiteindelijk tot effectiever beleid hebben geleid. Uit ervaringen elders is opgemaakt dat evaluatievermogen ook doorzettingsvermogen is. Dit maakt het interessant om de verdere ontwikkelingen te volgen.

Tabel 9 Belangrijkste conclusies in een notendop

Motieven (politieke context)	Aspecten van evaluatievermogen en hun kenmerken
	Structuur: organisatieonderdelen gericht op evaluatie en hun rol • Het beeld is zeer wisselend. FEZ is in elk departement standaard betrokken voor controletaken, maar de reikwijdte van haar rol wisselt. Daarnaast zijn er meestal coördinatoren onderzoek en voor een aantal departementen aanvullende, per departement verschillende ondersteunende (kennis)directies en organen. Vaste instituten bij het departement alleen bij VenJ, BZ en IenM • De aanvrager cq. beleidsmedewerker krijgt in internationale en ook Nederlandse leidraden een steeds belangrijker rol toebedeeld in (a) evaluatieproces en –inhoud en (b) het zorgen voor evalueerbaar (helder) beleid. • **Evaluatieprogrammering en –verplichtingen**

(A) Leren voor beleid	• De politieke context werkt door in het al of niet doen van evaluaties; ook, of in een wet een evaluatiebepaling wordt opgenomen. Soms doorkruist dit de bestaande programmering • Internationale instituties en grotere landen zien evaluatie als onderdeel van het beleidsproces; vanaf de conceptie van beleid zou evaluatief denken ingebakken moeten zijn. Zo kan bijv. worden bezien of beleid wel evalueerbaar (helder genoeg) is en voorts, of een nulmeting of monitoringsysteem moet worden aangelegd of een experimentele en controlegroep gevormd. • De evaluatieprogrammering in de rijksbegroting is leidend. Kennisagenda's zijn relatief evaluatie-arm. Er zijn initiatieven om evaluatieprogrammeringen pro-actiever te maken, maar dan vooral in anticipatie op de periodieke beleidsdoorlichtingen (verantwoording) • Binnen de programmering beginnen aanvragers vaak te laat met het opstarten van de geplande evaluaties, waardoor men 'meetkansen' laat liggen met gevolgen voor evaluatiedesign en -kwaliteit
(B) Verantwoorden Begrotingsuitgaven Politiek of strategisch: Kamervraag of -toezegging Evaluatiebepaling Anders	• **Evaluatiebudget** • Volgt uit het algemene onderzoeksbudget in de begroting en de resulterende bestedingsplannen; soms is het budget instituut- of programma-gebonden • Of middelen toereikend zijn is niet beoordeeld, wel bestaan aanwijzingen voor tijdgebrek, met name voor coördinatie van onderzoek. • Doorgaans bestaat weinig interne capaciteit voor evaluaties, het meeste onderzoek wordt uitgezet (beweegredenen: relatief snel te organiseren en objectief) • **Evaluatiekader: procesmatig en inhoudelijk** • Internationaal zeer veel handreikingen evaluatieonderzoek, steeds vaker gericht op de aanvrager; ook landelijk zijn er een heel aantal • Wat de inhoud betreft is *Fit-for-purpose* het belangrijkste devies: evaluaties zijn toegesneden op (o.a.) type instrumentarium (in wet), de beleidsfase en het type doelgroep(en) • Zowel internationaal als in Nederland groeit het aantal handreikingen en programma's voor aanvragers met betrekking tot de inrichting, opzet, begeleiding en beoordeling van evaluaties • Een enkele keer biedt men de evaluatiehandreiking geïntegreerd aan in het verband van *evidence-based* beleid maken. Raakt evaluatie meer 'ingebakken' in het beleidsproces? • **Verankering van evaluatiekennis en – ervaring in departement, DG of directie** • Ter communicatie van het evaluatiekader is een aantal cursussen ontwikkeld. Voorts bestaan her en der vaste 'hulplijnen' • Beleidsmedewerkers worden doorgaans niet op hun onderzoeks- en evaluatievaardigheden geselecteerd, met uitzondering van enkele onderzoekscoördinatoren en FEZ-medewerkers • Kwaliteitsborging gebeurt vooral via competitie in de aanbestedings- of subsidieprocedure; en via de inzet en samenstelling van begeleidingscommissies • Een multidisciplinair onderzoeksteam is volgens ervaringsdeskundigen van belang, juist bij wetsevaluaties: juridisch en sociaalwetenschappelijk • Het collectieve geheugen rond evalueren bij aanvragers lijkt gaten te

	bevatten. Verloop van medewerkers en wisselingen op dossiers werken negatief uit. Wel vindt informeel overdracht van leerervaringen plaats en soms via geijkte organisatieonderdelen, of interdepartementaal (IBP) • **Bevordering van bruikbaarheid en gebruik** • Er zijn – buiten horizonbepalingen in subsidieregelingen om – geen institutionele garanties voor lerend gebruik, nationaal noch internationaal. Wel zijn bruikbaarheidsinitiatieven genomen en vindt veel verspreiding in de vorm van primaire en secundaire publicaties plaats, ook richting de Tweede Kamer • De tamelijk veel gehanteerde begeleidingscommissies vervullen een belangrijke rol in termen van bruikbaarheid en feitelijk gebruik, soms al tijdens de evaluatie • De politieke context is cruciaal voor gebruik. Een sterke focus op verantwoording en legitimering staat mogelijk aan leren in de weg.

Dit rapport en zijn conclusies zijn aanleiding geweest voor een expertmeeting, gehouden op 6 april 2016. Deze had als doel om de discussie en de gedachtenvorming – in het bijzonder over het evaluatievermogen rond *wetgeving* – een stap verder te brengen.[160]

[160] Prof. dr. P. Van der Knaap verzorgde daar een referaat over het onderhavige rapport. Het verslag van deze expertmeeting is te vinden op de website van het KCWJ.

Summary

In the strive for regulatory quality, ex post evaluation research has an important role to play in the Netherlands. To actually fulfil this role, requires that (1) evaluations of laws and regulations (further: evaluations of laws) are planned and carried out; (2) these evaluations are designed in such a way as to generate useful, valid and reliable insights and (3) they are being used by policy officials and designers of laws to inform new or revised laws and policies. These three requirements are not automatically reached in practice. Laws and legal amendments are often complex units of analysis: a mix of norms, rules and manifestations of policy interventions. Apart from social-scientific skills, evaluations of laws also require a legal perspective. What is more, individual policy officials and jurists (designers of laws) indicated that know-how is lacking as to the commissioning and management of these evaluations. These factors may have caused evaluations of laws to vary quite strongly in terms of process, scope and methodological design. In the Netherlands there is no central regulatory evaluation policy. This should not have to be problematic at all. However, significant variation in terms of validity and reliability of evaluations of laws was found earlier.

This widely varying practice of evaluations of legislation has bearing on issues of *evaluation capacity*: the capacity to conduct evaluations, and, to use them in the regulatory cycle. Which conditions are present in policy departments to facilitate the production and utilization of evaluations of laws, in view of the evaluation process as well as content?

The aim of this study is to *gain insight into existing departmental practices as regards the conduct and use of ex post evaluations – particularly, evaluations of laws*. The study was commissioned by the Knowledge Centre for legislation and legal affairs of the Ministry of Security and Justice (Legislative quality division). The Knowledge Centre strives for more comparability between evaluations of laws, in terms of both process and content, to be able to better accumulate knowledge and draw lessons from experiences with these evaluations.

Research questions and approach

1. *Which general insights about core aspects of evaluation capacity are offered by the international (a) evaluation literature and (b) norms and guides of international institutions and larger countries?*
2. *Which experiences as regards evaluation capacity have been accumulated by policy departments, by policy officials and jurists?*
3. *What research designs are deemed relevant for methodological quality and content of evaluations of laws?*
4. *Which lessons can be drawn from the answers to the above questions for the (further) development of evaluation capacity – particularly in relation to legislation?*

This study was structured along with the following aspects of evaluation capacity and their characteristics, based on evaluation literature (research question 1). These constituted the conceptual backbone of the report:

- Evaluation structure: goals and organizational (evaluation) parts or units
- Evaluation programming and obligations
- Financial and other means for evaluation
- Evaluation frameworks & programs (process and content)
- Safeguarding evaluation knowledge and experience
- The promotion of evaluation utilization

We used this conceptual frame to, first, report a non-exhaustive number of approaches and experiences by leading international organisations and larger (western) countries. Second, desk study combined with 35 interviews with key informants were conducted to collect information from departmental practice as regards the six aspects.
Evaluations, including those of laws, do not take place in a vacuum. They are conducted from one or more of a variety of possible motives, e.g.: political accountability, for example, addressing a parliamentary question; financial accountability, to complete the budget cycle; to learn, i.e. to close the policy or regulatory cycle and use evaluation knowledge to feedback into new or adjusted polices. Evaluations may also serve a strategic goal. Two ideal-typical evaluation motives are discerned in the literature: (a) accountability and (b) learning, as in enlightenment and/or policy embetterment. The political context influences which of these motives is dominant. In turn, the evaluation motive affects how evaluations are dealt with and used. What is more, there is a correlation between motive and evaluation type. Summative evaluations, focused on the question 'did it work', more often suit the accountability motive, whereas formative evaluations, focused on understanding (intermediate) results from policies or laws, particularly fit the learning motive.

International experiences with evaluation capacity (research question 1)

A variety of international handbooks, guides and related materials offered by (among others) the UN, European Commission (EC), US, UK and World Bank demonstrate that policy officials – often the commissioners – have been assigned a significantly larger role in the commissioning and management of evaluations over the past decennium.
First, commissioners make crucial decisions about the evaluation process. They co-determine evaluation content and quality – as well as use. For example, they have a say or determine the overall evaluation plan / program and budget as well as the scope, focus and main evaluation questions. A multitude of criteria and guides circulate on formulating Terms of Reference (ToRs), the process of commissioning, supervision, judgement and utilization of evaluations (see Annex 3 to this report). These guides often also include information on suitable evaluation designs and methodologies, particularly for evaluations of effectiveness. Here, two strands can be discerned: *black-box* and *clear-box* evaluation approaches. Black-box approaches address the attribution problem: to what extent did the policy or law under evaluation actually trigger the results aimed for – rather than factors external to it? *How* this happened is outside the scope of study. In this strand an emphasis is put on experimental designs – including randomized controlled trials (RCTs) or quasi-experimental ones, where confounding factors are filtered out as much as possible. Clear-box approaches address the question: how did a policy or law work out, and why or under which circumstances? The two strands are not mutually exclusive; they may complement one another. In this regard, fit-for-purpose is an emerging mantra; 'the right' evaluation design hardly exists. Whether a design is adequate depends on, for example, type of policy, policy stage, topic, main questions to be answered and target group type(s). For example, the US General Audit Office (GAO) published a guide for evaluating highly complex programs: how to make the most out of an effectiveness evaluation when a program can hardly be disentangled from its context?
By making policy officials more aware of the relative strengths and limitations of different evaluation designs, they are enabled to make better informed decisions in relation to evaluation – so it is thought.
Second, apart from key players in the evaluation process, policy officials also help create evaluation preconditions by shaping their policies in the process of policy

development. It is they, not evaluators who determine the extent to which policies are clear and based on an articulated policy theory. Explicit assumptions about the intended chain of effects helps to focus an evaluation study at a later stage, adding to evaluation quality. A sharper focus could also enhance the cost-efficiency of the evaluation process, according to an EC-guide.
Apart from the policy theory, policy officials can help forward evaluation quality at an early stage by setting up a monitoring system from the inception of a new or revised policy or law. It is viewed as a missed opportunity if baseline and follow-up measurements of process, intermediate and end result indicators are not made from the start. Target groups like communities, citizens and companies can contribute to these monitoring processes.
Third and related, evaluation *use* is increasingly viewed as part of the policy-making process, rather than as the trickling down of an end report. The EC 'Smart Regulation' program prescribes a feedback loop from ex post evaluation findings to the first stage of policy formation or adjustment, through the obligatory ex ante Regulatory Impact Assessments (RIA). In these broad assessments ex post evaluation findings need to be included. At the UN Evaluation Group (UNEG), use is also prescribed and also meta research and knowledge repositories are promoted to keep the oversight of evaluation knowledge. At the country level, however, it has been hard to stimulate the systematic use of evaluations, according to a recent international comparative study. Nevertheless, evaluation societies and parliaments do (or can) fulfil a stimulating role in drawing lessons from evaluation.
Evaluators remain a key factor in evaluation quality as well as their use, judging, for example, from guides to write and present findings in a transparent manner. Their importance is also demonstrated by overviews of evaluator competencies, to be used by, among others, commissioners in the selection process. Not only methodological skills are judged important, also interpersonal and communication skills, flexibility and an antenna for the political and policy context.
Although internationally, much has been published regarding the promotion of evaluation capacity, the overall effectiveness of such initiatives is unknown as of yet. For example, adverse effects, for which some have warned, could be at work, like opportunistic behavior or evaluation ritualization.

Experiences by Dutch policy departments (research questions 2 and 3)

Policy departments very differently shaped evaluation capacity, as regards the involved organizational structure(s) as well as evaluation processes and methodological frameworks/content. Accountability in terms of budget spending turned out to be an important driving force behind evaluating. The key instrument of the central evaluation system is 'policy appraisals', syntheses of existing evaluation work covering specific parts ('articles') of the policy budget. Policy appraisals ought to take place every four to seven years at most and are to be sent to Parliament. The policy divisions within departments are responsible for setting up 'their' own policy appraisals. This may have an effect on day-to-day evaluation practice, for example on the programming of studies.

Evaluation structure
In the Netherlands, the vast majority of evaluation studies are commissioned to external parties like bureaus, research institutes or university departments, in view of, according to our respondents, internal capacity restrictions and objectivity. Within policy departments, a variety of units assist/are involved in managing evaluations. The initiative to conduct evaluations is most often taken by the research co-ordinator within a policy division or directorate. None of the policy departments we studied has a separate evaluation division or unit, but there are a number of semi-independent research institutes performing also evaluations, among

which the Research and documentation centre (WODC) at the Ministry of Security and Justice. All policy departments have a financial-economic division, responsible for accountability and control. To varying extents, these financial-economic divisions help stimulate the planning, quality and use of policy evaluation research. For example, at the Ministry of Social affairs, this role is quite pronounced as opposed to other departments. Within the Economic Affairs department, since 2012-2013 a network of divisions and units, including the financial-economic division, has addressed the quality of monitoring and evaluation (m&e) activities, mostly in relation to *effectiveness* evaluations.
As opposed to financial-economic divisions, 'knowledge divisions' are more concerned with strategic, future-oriented research than with ex post evaluations. The latter account for just a fraction of their research activities.
Several of the organizational structures and units are quite fresh and require time to take shape in everyday practice. An exception is the Commission for Regulatory Evaluation (CER) at ZonMw, which has been operating in the domain of evaluations of health law since almost two decades.

Evaluation programming and obligations
In all annual departmental budgets, an evaluation program is included and published. Besides this, almost no parallel programs were retrieved. These evaluation programs partly also result from obligations in about 10-20 per cent of concrete laws, to (periodically) evaluate the law in question, most often in terms of effectiveness and efficiency. As indicated, these evaluation studies form the base material for the periodic policy appraisals, in which findings are to be synthesized to account for budget spending. The process of evaluation programming seems to be dominated by accountability motives and political incentives. Most departments first determine an overall research budget, within which individual evaluations ought to fit. Research institutions like the WODC or the Netherlands Institute for Social research (SCP) administer their own budget.
As concerns non-financial means, respondents often reported that policy divisions spend a limited percentage of their time on research and evaluation, judging from (among other things) limited capacity for the co-ordination of research activities. This is not without consequence: it puts pressure on the programming of evaluations. What is more, time pressure makes policy officials to postpone the start of an evaluation, so that measurement opportunities are missed, like forming a control group, performing a baseline measurement and/or setting up a monitoring system.
Over the past few years, financial-economic divisions and sometimes other types of organizational divisions have tried to make evaluation programs more pro-active, particularly in view of the obligatory periodic policy appraisals and the strive for establishing policy effectiveness. It is too early to establish the results of these initiatives.

Evaluation frameworks & programs (process and content)
In chapter 3, we described a number of programs that serve to manage or enhance both the process and content of evaluation studies. Three of these programs particularly addressed legislation – in the domains of the environment, health and administrative law, respectively. The evaluation programs are examples of (temporary) evaluation capacity and offer lessons as regards the programming and demarcation of evaluation questions, design and evaluation criteria. One overarching lesson resulting from these experiences is the importance of a multidisciplinary, empirical-legal approach when evaluating legislation. Effectiveness in a legal sense does not imply that a law will achieve the intended societal effects in practice – and vice versa. In the Program for Regulatory Evaluation (PER,

managed by the aforementioned CER) in the health domain, evaluations of laws need to fulfil seven main requirements.
Besides these programs, we described a number of general programs applicable to policy making in general (potentially including laws). At the department of Education, the knowledge division stimulates combinations of black-box (experimental) and clear-box (explanatory) evaluation approaches. At the department of Economic affairs, the financial-economic division and the new evaluation network developed a guide for enhancing *policy* quality, of which an integrated part is how to perform evaluations. So here, evaluation is regarded as one of a set of indispensable means to achieve the overarching aim of (more) effective policy making. The guide encompasses all stages of the policy cycle, from problem analysis, choice of instruments and ex ante appraisal to political decision-making, implementation and finally, ex post evaluation. It also contains tips and tricks for doing 'early' (small scale) policy experiments and for setting up a monitoring system.
The dominant approach in evaluation research differs between the policy departments studied, depending in part on the dominant departmental divisions or units involved. For example, if the financial-economic division is very active, summative approaches often prevail over explanatory ones. For example, at the Economic Affairs department, commissioners can make use of guides and checklists for the commissioning and content of (experimental) *effectiveness* evaluations, including checklists for assessing tenders, TORs and final reports. If knowledge divisions or scientific research institutes are involved in evaluation, more often a mix of evaluation approaches is encountered (e.g. at the Education department).
In a recent ex post evaluation program at the Ministry of Infrastructure and the Environment, the ambition is to hark back to earlier findings from ex ante regulatory impact assessments of the policy or law.

Safeguarding evaluation knowledge and experience
Like internationally, the commissioning party is increasingly regarded as a key player in the process and content of evaluation research by the Dutch policy departments. To what extent and how do policy departments make sure that commissioners have the know-how and experience to adequately manage evaluations? We addressed: which parts of the departmental organization fulfill a role in this regard, the presence of evaluation training courses, how external evaluators are selected and then supervised, and how evaluation experience and knowledge is retained and shared (collective memory).
In most policy departments, research coordinators – often senior policy officials – fulfil a key function in this regard; they are relatively experienced and help forward less experienced colleagues. Besides this, support functions vary significantly between departments. For example, the Ministry of Social Affairs has a *Chief Science Officer* who supports the co-ordination and methodology of research including evaluations. At the Education Ministry, the Knowledge division fulfils a similar role as well as a Policy Analyst Team at the Ministry of Economic Affairs that forms part of the aforementioned evaluation *network*. At the Ministry of Security and Justice, project managers of (evaluation) research more or less 'take over' the supervision of evaluations after the TOR-stage. Nevertheless, the commissioner from the Ministry (most often a policy official) is member of a supervisory committee, which is always installed.
A limited number of departments and 'government academies' offer special courses for policy officials, fully or partly about evaluation.
The selection of evaluation research teams is most often determined in the competitive tender procedure, in which each ministry uses its own (more or less explicit) set of criteria. Some ministries have a more elaborate selection procedure than others. For example, the Education Ministry that uses long-term research

contracts. At the Ministry of Security and Justice, the WODC keeps a database containing characteristics and capacities of external research teams. Particularly in case of evaluations of laws, a multidisciplinary team is crucial.
Once the evaluation process has started, supervisory committees are regularly used to monitor progress, quality and content. In some policy departments, this is standard practice and there are procedures as regards the role and composition of these committees. For example, the WODC requires the chair of the supervisory committee to be a senior academic.
The exchange of departmental experiences with evaluation projects is often done by informal means, for example, in bilateral or division meetings. Respondents doubt whether this will safeguard 'collective evaluation memory'. Between departments, periodic exchange of evaluation experiences takes place between at least the financial-economic divisions and the Ministry of Finance (often in the light of policy appraisals). This was valued positively by the respondents involved.
The aforementioned 'fresh' evaluation structures and programs appear to be vulnerable when taking into account: departmental cultures as regards research and evaluation, reorganizations and large turnover of policy staff. This is in line with foreign experiences that enhancing and consolidating evaluation capacity is a long-winded task that requires perseverance.

Promoting evaluation utilization
Evaluation use is an indispensable aspect of evaluation capacity. Without processing the fruits of evaluation in the development of laws and policies, each of the capacity aspects described above will lose meaning. Contact and consultation between commissioner and evaluator enhances the probability of evaluation use according to the literature as well as this study. Timeliness and timing in the policy cycle are also important factors. In general, departments publish evaluation reports and often the reports are sent to Parliament. In many cases, an official cabinet response is attached. This response often includes statements as to how a Minister intends to deal with the findings and (if available) recommendations. Besides this, evaluators themselves often valorize evaluation findings by means of presentations for policy officials or academics and articles in professional or academic journals.
In the end, it is up to policy officials and the political context to what extent and how findings are used. There are no fixed or prescribed routes for this, apart from horizon clauses in financial regulations. These regulations expire after five years and can only be prolonged if sufficiently sustained by, among other things, ex post evaluation findings.
To increase their learning value, users argue that ex post evaluations should not only look back, but be oriented towards future action to a larger degree. Evaluations are also deemed instructive if policy assumptions (policy theory) are given a reality check. Finally, more knowledge repositories could be established in which findings from previous evaluation studies are systematized.

Conclusions and discussion (research question 4)

Many of the aforementioned evaluation structures and programs are still quite young, making it hard to determine their impact on learning and accountability and ultimately, on the effectiveness of laws an policies. Hence, it would be interesting to closely follow future developments in the context of evaluation capacity.
One key condition for evaluation capacity is that initiatives to enhance policy relevant knowledge and learning experiences are welcomed rather than punished, in the political context as well as in the central evaluation system. In relation to this, some respondents argued that to learn from evaluations requires formative (process and intermediate) evaluations next to summative work about end effects.

What do the above findings imply for law making divisions and the commissioner of this study, the Knowledge Centre for legislation and legal affairs (Legislative quality division)? Quite rarely are law making divisions involved in the process of ex post evaluations – as opposed to ex ante activity. Only in a limited number of cases, an ex post evaluation of legislation was coordinated by a legal division – for example, when several policy divisions were involved (e.g. regarding the railway system). Overall, policy divisions rather than jurists are responsible for ex post evaluations. Nevertheless, involving jurists could enhance the content and quality of evaluations of laws. After all, combined legal-empirical approaches proved to add value. What is more, in the very process of law-making, jurists could design evaluation clauses (obligations to evaluate within X years) in such a way that the resulting evaluations become more meaningful. For example, by choosing realistic evaluation periods to allow effects sufficient time to materialize. Finally, legal divisions could be made more aware of the relevance of the *explanatory memorandum* in the clarification of the main policy aims and assumptions as regards effectiveness. A clear policy theory will make it more easy to focus evaluations of legislation at a later stage. The Knowledge Centre for legislation and legal affairs has developed a guide for writing explanatory memoranda.

Literatuur

Algemene Rekenkamer. (2011). *Leren van subsidie-evaluaties*. Den Haag: Algemene Rekenkamer.

Algemene Rekenkamer. (2012). *Effectiviteitsonderzoek bij de rijksoverheid*. Den Haag: Algemene Rekenkamer.

Algemene Rekenkamer. (2013). *Effectiviteitsonderzoek bij de rijksoverheid: vervolgonderzoek*. Den Haag: Algemene Rekenkamer.

Algemene Rekenkamer, TK 1993-1994, 23710 nrs. 1-2. *Wetgeving: organisatie, proces en produkt*. Verkregen van http://resourcessgd.kb.nl/SGD/19931994/PDF/SGD_19931994_0006932.pdf .

Astbury. B., & Leeuw, F.L. (2010). Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation,* 31(3), 363-381.

American Evaluation Association (AEA) (2004). *Guiding principles for evaluators*. Washington DC: AEA, www.eval.org.

American Evaluation Association (AEA) (2013). *Evaluation Roadmap for a more effective government* (revised). Washington DC: AEA, www.eval.org.

Backes, C.P., Hardy, E.M.J., Jansen, A.M.L., Polleunis, S., & Timmers, R. (2014). *Evaluatie Bestuurlijke Lus Awb en internationale rechtsvergelijking*, Maastricht: Universiteit Maastricht (i.o.v. WODC).

Bakker, R. (2012). *Evaluatie als fundering voor beleid*. Den Haag: Ministerie van Binnenlandse Zaken. Verkregen van www.rijksoverheid.nl/documenten/rapporten/2012/06/01/evaluatie-als-fundering-van-beleid .

BEC-secretariaat (z.j.). *Handreikingenboekje beleidskwaliteit en beleidsevaluatie*. Intern document van het Ministerie van EZ.

Bex, P.M.H.H., Bovens, F., & Goo, F.S. (2010). *Evaluatie Wet extern klachtrecht: onderzoek naar de organisatie en effecten van het extern klachtrecht bij gemeenten*, Eindrapportage (versie 1.0). Nieuwegein: SIRA Consulting.

Bemelmans-Videc, M.L. (2002). Evaluation in the Netherlands 1990-2000: Consolidation and Expansion. In J.-E. Furubo, R.C. Rist & R. Sandahl (Red.). *International Atlas of Evaluation.* New Brunswick: Transaction Publishers.

Bourgeois, I., Hart, R.E., Townsend, S.H., & Gagné, M. (2011). Using hybrid models to support the development of organizational evaluation capacity: a case narrative. Evaluation and program planning, 34, 228-235. Doi: 10.1016/j.evalprogplan.2011.03.003.

Bourgeois, I. en Cousins, J.B. (2013). Understanding Dimensions of Organizational Evaluation Capacity. *Americal Journal of Evaluation, 34*(3), 299-319. Doi: 10.1177/1098214013477235.

Bovens, M., 't Hart, P., & Kuijpers, S. (2008). The politics of policy evaluation. In M. Moran, M. Rein & R.E. Goodin, (red.), *The Oxford Handbook of public policy* (pp. 319-335). Oxford: Oxford University Press

Bussmann, W. (2010). Evaluation of Legislation: Skating on Thin Ice. *Evaluation, 16*(3), 279-293. Doi: 10.1177/1356389010370252.

De Waard, B.W.N. (red.) (2011). *Ervaringen met bezwaar*. Tilburg/ Den Haag: Tilburg University/ WODC.

Centraal Planbureau (2011). *Zicht op Effectiviteit. Studie naar evaluatieontwerpen voor onderwijs- en wetenschapsmaatregelen.* Den Haag: CPB achtergronddocument. Verkregen van www.beleidsevaluatie.info/wp-content/uploads/2011/06/cpb-achtergronddoument-zicht-op-effectiviteit-van-beleid.pdf .

Clements, P., Chianca, T., & Sasaki, R. (2008). Reducing world poverty by improving evaluation of development aid. *American Journal of Evaluation* 29(2), 195-214. Doi: 10.1177/1098214008318657.

Coglianese, C. (2012). *Measuring Regulatory Performance. Evaluating the Impact of Regulation and Regulatory Policy*. Parijs: OESO.

Commissie Hoekstra, Visitatiecommissie Juridische functie en Wetgeving (2007). *Met recht verbonden. Verslag van de interdepartementale Visitatie Juridische functie en Wetgeving.* Den Haag.

Commissie Theeuwes. (2012). *Durf te meten. Eindrapport Expertwerkgroep Effectmeting*. www.rijksoverheid.nl/documenten/rapporten/2012/11/23/durf-te-meten-eindrapport-expertwerkgroep-effectmeting

Cooksy, L.J. (2012). Influences on evaluation quality. *American Journal of Evaluation, 33*(1), 79-87. Doi: 10.1177/1098214011426470.

Cooksy, L.J., & Caracelli, V.J. (2009). Metaevaluation in practice. Selection and application of criteria. *Journal of MultiDisciplinary Evaluation 6(11).* www.jmde.com.

De Boer, L., & Larsen, V. (2013). *Nieuwe werkpraktijken MKBA. Een inventarisatie van aansprekende procesvernieuwingen*. Den Haag: LPBL (i.o.v. Ministerie van Ienm)

Den Ouden, W., Kortmann, C.N.J., Jongbloed, A.W., Van den Brink, J.E., & Tjepkema, M.K.G. (2013). *De bestuursrechtelijke geldschuldenregeling. Titel 4:4 Awb geëvalueerd*. Universiteit Utrecht / Universiteit Leiden. Den Haag: WODC.

EIM (2012). *Evaluatie WBSO 2006-2010*. Zoetermeer: EIM.

European Commission (EC) (2013a). *Evalsed Guide, the resource for the evaluation of Socio-Economic Development.* Verkregen van http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide_evalsed.pdf.

EC (2013b). *Designing and commissioning of counterfactual impact evaluations (CIE). A practical guide for ESF managing authorities*. Luxemburg: Publication office of the EU.

EC (2015a). *Monitoring and evaluation of cohesion policy (2014-2020). Evaluation plan Guidance Document on Evaluation Plans. Terms of Reference for Impact Evaluations. Guidance on Quality Management of External Evaluations.* Verkregen van http://ec.europa.eu/regional_policy/sources/docoffic/2014/working/evaluation_plan_guidance_en.pdf

EC (2015b). Better Regulation Guidelines. Verkregen van http://ec.europa.eu/smart-regulation/guidelines/index_en.htm.

EC (2015c). *Better Regulation Toolbox. VI: Evaluations and fitness checks.* Verkregen van http://ec.europa.eu/smart-regulation/guidelines/toc_tool_en.htm

EC (2015d). *Evaluatieplan Europees Sociaal Fonds (ESF) 2014-2020*.

Eijlander, Ph. (2003). De Awb-evaluatie geëvalueerd. Hoe is de tweede evalautie van de Awb aangepakt en wat heeft die opgeleverd? *Nederlands Juristenblad*, 9(9), 265-269.

Farrington, D.P., & Welsh, B.C. (2005). Het belang van experimentele evaluaties in de criminologie. *Justitiële verkenningen,* 31(8), 11-41.

FEZ-EZ (z.j.) *Handreiking evaluatieonderzoek EZ*. Intern document van EZ.

Forss, K., & Carlsson, J. (1997). The quest for quality – or can evaluation findings be trusted? *Evaluation*, 3(4), 481-501.

Furubo J-E., Rist R.C., & Sandahl, R. (red.) (2002) *International Atlas of Evaluation*. New Brunswick, NJ and London: Transaction Publishers.

GAO (2012) *Designing evaluations, 2012 revision*. Washington: US Government, Government Accountability Office 12-208G.

GfK (2015). *Het rijksinkoopstelsel; gebruik, kennisniveau en informatiebehoeften. Kwantitatief en kwalitatief onderzoek*. Presentatie. Verkregen via www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2015/08/01/het-rijksinkoopstelsel-gebruik-kennisniveau-en-informatiebehoeften/20150902-rapportage-het-rijksinkoopstelsel-def.pdf.

Gough D., Oliver, S., & Thomas, J. (2012). *An introduction to systematic reviews*. London: Sage.

Van Gils, G., & Leeuw, F.L. (2010). *Leren van Evalueren: Onderzoek naar het gebruik van evaluatieonderzoek bij het Ministerie van Buitenlandse Zaken*. Den Haag: Ministerie van Buitenlandse Zaken, p. 47.

HM Treasury (2011a). *Green book. Appraisal and evaluation in central government*. Treasury Guidance, London: TSO. Verkregen van: www.gov.uk/government/uploads/system/uploads/attachment_data/file/220541/green_book_complete.pdf.

HM Treasury (2011b). *Magenta book. Guidance for evaluation*. Verkregen van www.gov.uk/government/publications/the-magenta-book

Heijnen, S., Van der Steeg, M., & Ooms, T. (2015). *Evaluatie van de lerarenbeurs: Aanvragers, Deelname en Afronding*. CPB Achtergronddocument (behoort bij CPB Discussion Paper 305). Den Haag: CPB. www.cpb.nl/publicatie/evaluatie-van-de-lerarenbeurs-aanvragers-deelname-en-afronding

Hertin, J., Turnpenny, J., Jordan, A., Nilsson, M., Russel, D., & Nykvist, B. (2007). *Rationalising the policy mess? Ex ante policy assessment and the utilisation of knowledge in the policy process*. Berlin: Freie Universität Berlin, FFU report 03-2007.

Jacob, S., Speer, S., & Furubo, J.E. (2015). The institutionalization of evaluation matters: updating the International Atlas of Evaluation 10 years later. *Evaluation* 21(1), 6-31.

Johnson, K., Greenseid, L.O., Toal, S., King, J.A., & Lawrenz, F. (2009). Research on evaluation use: a review of the empirical literature from 1986 to 2005, *American Journal of Evaluation* 30(3) **77-410. D**oi: 10.1177/1098214009341660

Kennisplatform Crow. (2014). *Leren en beleid bijstellen*. Geraadpleegd op 8 september 2015 op www.crow.nl/mobiliteit-en-gedrag/weblog/november-2014-%281%29/leren-en-beleid-bijstellen.

King, J. A., Stevahn, L., Ghere, G., & Minnema, J. (2001). Toward a taxonomy of essential evaluator competencies. *American Journal of Evaluation*, 22, 229-247.

Klein Haarhuis, C.M. en Niemeijer, E. (2008). *Wet en werkelijkheid. Bevindingen uit evaluaties van wetten*. Den Haag: Boom, WODC Onderzoek en Beleid.

Klein Haarhuis en Oosterwaal (2014). *Handvatten voor evaluaties van effectiviteit*. www.kcwj.nl/sites/default/files/attachments/handvatten-voor-evaluaties-van-effectiviteit.pdf

Kort, M., Van Klink, L., & Van Gils, N. (2014). *Tussentijdse evaluatie Taxiwet.* Rotterdam: Rebel Group.

Kushner, S. (2005). Qualitative Control. A Review of the Framework for Assessing Qualitative Evaluation. *Evaluation, 11*(1), 111-122. Doi: 10.1177/1356389005053194.

Lahey, R. (2010). *The Canadian M&E system: lessons learned from 30 years of development*. ECD Working paper series, 23. Washington DC: World Bank Independent Evaluation Group.

Legemaate, J. (1997). Evaluatie van wetgeving in de gezondheidszorg, *Sociaal Recht,* 251-256.

Leeuw, F.L., Rist, R.C., & Sonnichsen, R.C. (1994). *Can Governments Learn? Comparative Perspectives on Evaluation & Organizational Leaning*. New Brunswick: Transaction Publishers.

Leeuw, F.L. (2009). Evaluation Policy in the Netherlands. In W.M.K. Trochim, M.M. Mark en J. Cooksy (red.). *Evaluation policy and evaluation practice. New Directions for Evaluation, 123*, 87-102.

Lösel, F. (1995). The efficacy of correctional treatment: A review and synthesis of meta-evaluations. In J. McGuire (red.), *What works: Reducing reoffending* (pp. 79-114). Chichester, UK: Wiley.

Mackay, K. (2006). Evaluation Capacity development. Institutionalization of Monitoring and Evaluation Systems to Improve Public Sector Management. *ECD Working Paper Series, 15*, Washington D.C.: The World Bank.

Mark, M.M. (2012). Influences on evaluation quality: reflections and elaborations. *American Journal of Evaluation*, 33(1), 85-87. Doi:10.1177/1098214011426470.

Mastenbroek, E., Van Voorst, S., & Meuwese, A. (2015). Closing the regulatory cycle? A meta-evaluation of ex-post legislative evaluations by the European Commission. *Journal of European Public Policy 17(x).* Doi: 10.1080/13501763.2015.1076874.

Mayne, J.L., Bemelmans-Videc, M.L., Hudson, L., & Conner, R. (1992). *Advancing Public Policy Evaluation: Learning From International Experiences*. Amsterdam: Elsevier.

Mayne, J. (2008). Contribution analysis: an approach to exploring cause and effect. ILAC (Institutional learning and change) Brief. www.cgiar-ilac.org

Michiels, F.C.M.A. (2010) *Het evalueren van de Awb: een voortdurend proces*. .. p. 41-55. In

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

Ministerie van BZK (2010). Strategische kennisagenda: investeren in legitimiteit. Verkregen van www.rijksoverheid.nl/documenten/rapporten/2010/03/02/strategische-kennisagenda-investeren-in-legitimiteit

Ministerie van Financiën. (2003). *Handreiking Evaluatieonderzoek ex post*. Den Haag: Directie Begrotingszaken. Geraadpleegd op 28 augustus 2015 op www.minfin.nl/vbtb.

Ministerie van Financiën (2006). *Regeling Periodiek Evaluatieonderzoek*. Staatscourant Nr. 83.

Ministerie van Financiën. (2014). *Regeling Periodiek Evaluatieonderzoek*. Staatscourant Nr. 27142.

Ministerie van Financiën (2016). *Handreiking beleidsdoorlichtingen* http://www.rijksbegroting.nl/beleidsevaluaties/evaluaties-en-beleidsdoorlichtingen/handreiking

Ministerie van IenM (2013). Strategische kennis- en innovatieagenda (SKIA) 2012-2016. Verkregen van www.rijksoverheid.nl/documenten/brochures/2012/06/29/ienm-maakt-ruimte-strategische-kennis-en-innovatieagenda-skia-infrastructuur-en-milieu.

Ministerie van Justitie (2000). *Nota Wetgevingskwaliteitsbeleid en wetgevingsvisitatie.* Tweede Kamer, vergaderjaar 2000–2001, 27475, nr. 2. Den Haag: Sdu Uitgevers.

Ministerie van Justitie (1991). *Nota Zicht op wetgeving.* Tweede Kamer, vergaderjaar 1990-1991, 22008, nrs. 1-2. Den Haag: Sdu Uitgevers.
Ministerie van OCW. (2015a). *Notitie Zicht op Effectiviteit*. Intern document van Directie Kennis van het ministerie van OCW.

Ministerie van OCW. (2015b). *Kennisagenda OCW*. Verkregen van www.rijksoverheid.nl/documenten/publicaties/2015/03/13/kennisagenda-ocw-2015.

Ministerie van SZW (2009). *Organisatie-, mandaat- en volmachtbesluit FEZ (artikel 4).* Verkregen van http://wetten.overheid.nl/BWBR0024956/2016-04-05.

Ministerie van SZW (2013). Strategische kennisagenda SZW. Verkregen van www.rijksoverheid.nl/documenten/rapporten/2013/08/29/kennisagenda-szw-2013.

Ministerie van V&W (2008). *Spoor in beweging. Eindrapport evaluatie spoorwetgeving*. Den Haag: V&W, Projectorganisatie evaluatie spoorwetgeving.

NAO (2009). *Post Implementation Review of Statutory Instruments: analysis of the extent of review by government departments. Briefing for the house of lords merits of statutory instruments committee, December*. London: National Audit Office.

Nelen, J. (2000). *Gelet op de wet. Evaluatie van strafwetgeving onder de loep*. Amsterdam: VU (dissertatie).

Nielsen, S.B., Lemire, S., & Skov, M. (2011). Measuring Evaluation Capacity – Results and Implications of a Danish Study. *American Journal of Evaluation*, 1-21. Doi: 10.1177/1098214010396075.

OESO (2014). *Framework for regulatory policy evaluation*. Verkregen van www.oecd.org/regreform/framework-for-regulatory-policy-evaluation.htm

OESO (2015a). *Regulatory Policy Outlook.* Parijs: OECD.

OESO (2015b). *Governance at a glance 2015, Chapter 8. Ex post evaluation of regulation*. Paris: OECD Publishing. doi: http://dx.doi.org/10.1787/gov_glance-2015-40-en

Olsthoorn-Heim, E. (2003). *Vijf jaar evaluatie regelgeving via ZonMw. Verslag van een analyse van het programma Evaluatie Regelgeving in opdracht van het Bestuur van ZonMw*. Amsterdam: Met Recht onderzoek en advies in gezondheidsrecht.

Oosting, M. (1987). De veelzijdigheid van wetsevaluatie. In J.H.T.H. Andriessen et al. (red.), *Wetsevaluatie tussen wetenschap en beleid*. Zwolle: Tjeenk Willink.

Panel van Advies IOB (2011). *Eerste advies van het panel van advies,* Den Haag: Inspectie Ontwikkelingssamenwerking en beleidsevaluatie.

Panel van Advies IOB (2013), *Tweede advies van het panel van advies,* Den Haag: IOB.

Panel van Advies IOB (2014), *Derde advies van het panel van advies,* Den Haag: IOB.

Parapuf, A. (2016). *Zicht op evaluatiecapaciteit. Een onderzoek naar het vermogen van Nederlandse ministeries om (wets)evaluaties te verrichten.* Nijmegen: Radboud Universiteit (Masterscriptie Bestuurskunde, Organisatie & management).

Pater, C., Sligte, H., & Van Eck, E. (2012). *Verklarende evaluatie, een methodiek.* Amsterdam: Kohnstamm Instituut.

Patton, M.Q. (2008). *Utilization-Focused Evaluation*, 4th Ed. Thousand Oaks, CA: Sage.

Patton, M.Q. (2010). Utilization-focused evaluation. Real-time and prospective aspects. In Wereldbank (red.), *Exploring the potential of real-time and prospective evaluations, summary of a workshop* (pp. 6-11). Washington DC: World Bank Independent Evaluation Group.

Pattyn V., & Verweij S. (2014). Beleidsevaluaties tussen methode en praktijk: Naar een meer realistische evaluatiebenadering. *Burger, Bestuur en Beleid, 4*, 260-267.

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. London: Sage.

Power, M. (1997). *The audit society: rituals of verification*. Oxford: Oxford University Press.

Prins, A. (2012a), *Gebruik en gebruikers van WODC publicaties: Een bibliografische analyse van het WODC 2008-2010*, Groningen: Ad Prins

Prins, A. (2012b), *Citatieanalyse van overige WODC publicaties 2007-2011*, Groningen: Ad Prins.

Preskill, H. en Boyle, S. (2008). A Multidisciplinary Model of Evaluation Capacity Building. *American Journal of Evaluation, 29*(4), 443-459. DOI: 10.1177/1098214008324182.

Programma Versterking Juridische Functie Rijk (2010)

Regering van Australië (2011). Identifying and evaluating regulation reforms. Productivity commission research report. Canberra: Productivity Commission. www.pc.gov.au.

Regering van Zuid-Australië (2011). *Better Regulation Handbook. How to design and review regulation, and prepare a Regulatory Impact Statement.* Melbourne: Regering van Zuidaustralië.

Regering van Australië (2014). Post implementation reviews. Guidance note. Office of Best Practice Regulation. www.dpmc.gov.au/office-best-practice-regulation/publication/post-implementation-reviews-guidance-note.

Rijksacademie voor Financiën, economie en bedrijfsvoering (2015). *Beleid in perspectief* (cursus). Verkregen van www.rijksacademie.nl/opleiding/bip/.

Rijksoverheid. *Inkopen door het Rijk*. Verkregen van www.rijksoverheid.nl/onderwerpen/inkopen-door-het-rijk/inhoud/inkooporganisatie-van-het-rijk.

Ringeling, A. (1997) *Evaluatie: hoe en wat?, Regelmaat* p 143-150

Rossi, P., & Freeman, H. (1992). *Evaluation: A systematic approach.* Newbury Park, CA: Sage.

Schueler, B.J., Blekemolen, M., Dijkersloot, A.P.W., Modderman, C.B., Ortlep, R., & Scholtes, H.H.M. (2013). *Evaluatie van een drietal versnellingsinstrumenten uit de Awb,* Utrecht: Universiteit Utrecht, Twynstra Gudde Adviseurs en Managers.

Schwartz, R., & Mayne, J. (2005). Assuring the quality of evaluative information: theory and practice. *Evaluation and Program Planning* 28, 1–14.

Scriven, M. (1994). The fine line between evaluation and explanation. *Evaluation Practice,* 15(1), 75-77.

Smismans, S. (2015). Policy Evaluation in the EU: the challenges of linking ex ante and ex post appraisal, *EJRR*, 2015(1), p. 6-25.

Stern, E. (2009). Evaluation policy in the European Union and its institutions. In W.M.K. Trochim, M.M. Mark, & L.J. Cooksy (red.), *Evaluation policy and evaluation practice. New Directions for Evaluation, 123,* 67–85.

Stevahn, L., King, J.A., Ghere, G., & Minnema, J. (2005). Establishing Essential Competencies for Program Evaluators. *American Journal of Evaluation 26* (1), 43-59. Doi: 10.1177/1098214004273180.

Suarez-Balcazar, Y., & Taylor-Ritzler, T. (2013). Moving from science to practice in evaluation capacity building. *American Journal of Evaluation*, 35(1) 95-99. Doi: 10.1177/1098214013499440.

Taylor-Ritzler, T., Suarez-Balcazar, Y., Garcia-Iriarte, E., Henry, D.B. en Balcazar, F.E. (2013). Understanding and measuring evaluation capacity. A model and instrument validation study. *American Journal of Evaluation* 34(2), 190-206. doi: 10.1177/1098214012471421.

TIER (2011). *De evidence based beleidsmaker en toezichthouder. Een master voor professionals in Evidence Based Policy and Evaluation*. Verkregen van www.tierweb.nl/dtpa/opleidingsinformatie/master-evidence-based-policy-and-evaluation.html.

United Nations Evaluation Group (UNEG) (2005a). *Norms for evaluation in the UN system.* Verkregen van www.iom.int/jahia/webdav/site/myjahiasite/shared/shared/mainsite/about_iom/eva_techref/UNEG_Norms_for_Evaluation.pdf.

UNEG (2005b). *Standards for evaluation in the UN system*. Verkregen van www.iom.int/jahia/webdav/site/myjahiasite/shared/shared/mainsite/about_iom/eva_techref/UNEG_Standards_for_Evaluation.pdf.

UNEG (2010a). *Quality Checklist for Evaluation Terms of Reference and Inception Reports*. UNEG Guidance document 2010/1. Verkregen van www.uneval.org/document/library .

UNEG (2010b). *Quality Checklist for Evaluation Reports*. UNEG Guidance document 2010/1. Verkregen van www.uneval.org/document/library .

Uylenburg, R., de Boer, J., Peeters, M.G.W.M., & Oosterhuis, F.H. (2011). *Het STEM-geluid bij de evaluatie van milieuwetgeving*. Arnhem: STEM Publicatie 2011/1.

Van Aeken, K. (2007). From vision to reality: ex post evaluation of legislation. *Legisprudence: international journal for the study of legislation* 5(1), 41-68.

Van Gils, G., & Leeuw, F.L. (2010). *Leren van Evalueren: Onderzoek naar het gebruik van evaluatieonderzoek bij het Ministerie van Buitenlandse Zaken*. Den Haag: Ministerie van Buitenlandse Zaken.

Van Hoesel, P.H.M., Mevissen, J.W.M., & Dekker, L.H. (2015). *Kennis voor Beleid - Beleidsonderzoek in Nederland.* Assen: Van Gorcum.

Van der Laan, A., Beijaard, F., Geurts, T., Kalidien, S., & Straalen, E.K. van (2013). *Zelfevaluatie WODC, periode 2006-2012.* Den Haag: WODC.

Van Thiel, S., & Leeuw, F.L. (2002). The performance paradox in the public sector. *Public Performance & Management Review*, 25(3) 267-281.

Vedung, E. (1997). *Public Policy and Program Evalution*. New Brunswick: Transaction Publishers.

Veerman, G.J. m.m.v. Hendriks-de Lange, S.R. (2007). *Over wetgeving. Principes, paradoxen en praktische beschouwingen*. Den Haag: Sdu Uitgevers.

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

Veerman, G.J. (2014). Klaarheid over het Clearing House. *RegelMaat,* 29 (4), 199-211.

Vereniging voor Beleidsonderzoek (VBO) (2004). *Twaalf tips over het offertetraject voor opdrachtgevers van beleidsonderzoek.* Nijmegen: VBO. Verkregen van www.beleidsonderzoek.nl

Von Meyenfeldt, L., Schrijvershof, C., & Wilms, P. (2008) *Tussenevaluatie beleiddoorlichting*. Den Haag: APE, rapport nr. 570.

Wereldbank (1992). *Governance and development*. Washington, DC : The World Bank. Verkregen van http://documents.worldbank.org/curated/en/1992/04/440582/governance-development

Wereldbank (2002). *Evaluation capacity development: a growing priority*. Washington DC: World Bank Operations Evaluation Department. Précis, nr. 229.

Wereldbank (2004a). *Monitoring and evaluation: Some tools, methods & approaches.* Washington DC: World Bank Operations Evaluation Department.

Wereldbank (2004b). *Ten steps to monitoring & evaluation systems*. Washington DC: World Bank.

Wereldbank (2006). *Impact evaluation – the experience of the World Bank Independent Evaluation Group.* Washington DC: World Bank, IEG.

Wereldbank (2010). *Exploring the Potential of Real-Time and Prospective Evaluations, summary of a workshop*. Washington DC: World Bank Independent Evaluation Group.

Winter, H.B. (1996). *Evaluatie in het wetgevingsforum*. Deventer: Kluwer (dissertatie).

Winter, H.B., & Klein Haarhuis, C.M. (2007). Wetsoverstijgende evaluaties. *Tijdschrift voor Gezondheidsrecht* 31(8), p. 620-626

Witte Huis (2014). *Economic Report of the President (Chapter 7: Evaluation as a tool).* Washington: US Government printing office. Verkregen van www.whitehouse.gov/sites/default/files/docs/full_2014_economic_report_of_the_president.pdf

WODC (2012). *Wegwijzer bij contractonderzoek in opdracht van het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) van het ministerie van Veiligheid en Justitie*. Den Haag: WODC.

WODC (2013). *Evaluatie van WODC-onderzoek door aanvragers 2006-2012*, Den Haag: WODC.

WODC (2014). Rapport van de Visitatiecommissie, Commissie Prins. Tilburg/Den Haag.

WODC (2015). *Onderzoeksprogramma WODC*. Geraadpleegd 8 september 2015 op www.wodc.nl/onderzoek/onderzoeksprogramma/index.aspx.

Wollman, H. (red.). (2003). *Evaluation in public sector reform: Concepts and practice in international perspective.* Cheltenham/Northampton: Elgar.

Yarbrough, D.B., Shulha, L.M., Hopson, R.K., & Caruthers, F.A. (2011). *The Program Evaluation Standards. A guide for evaluators and evaluation users.* USA: Sage (3rd ed.).

ZonMw, Commissie Evaluatie Regelgeving (2013), *Brochure Programma Evaluatie Regelgeving*, Den Haag: ZonMw.

Bijlage 1 Begeleidingscommissie

Prof.mr.dr.	E. Niemeijer	Academie voor Wetgeving & Academie voor Overheidsjuristen
Mr.dr.	J. L.H.M. Stoop	Ministerie van Infrastructuur en Milieu / Beleids- en Bestuursondersteuning / Hoofddir. Bestuurlijke & Juridische Zaken / Directie Milieu, Ruimte en Water / Afdeling Milieu
Drs.	D. de Groot	Ministerie van Veiligheid en Justitie, Directie Wetgeving en Juridische Zaken, Wetgevingskwaliteitsbeleid
Dr.	E Mastenbroek	Radboud Universiteit Nijmegen, Faculteit der Managementwetenschappen

Bijlage 2 Respondenten en interviews*

Departement	Onderdeel	Datum (2015)
BZ (geen wetsevaluaties)	IOB	2 september
BZK	Financieel-economische Zaken (FEZ) samen met directie Kennis & Verkenningen DG Wonen en Bouwen	13 oktober
EZ	FEZ / Beleidskwaliteit- en evaluatiecommissie (BEC) DG Bedrijfsleven & Innovatie - Regiegroep M&E (later BAT) Directie Algemene Economische Politiek (AEP) *Behavioural Insights Team* (BIT) Wetgeving & Juridische Zaken (5 pers.)	15 juli
Financiën (alleen over centrale evaluatiestelsel)	Inspectie der Rijksfinanciën (IRF), Bureau Strategische Analyse (2 pers.)	26 augustus
IenM	Hoofddirectie Bestuurlijke en Juridische Zaken (HBJZ) (2 pers.) & FEZ (Financien, Management en Control- FMC) Wetgevingsjurist HBJZ	16 juli 24 augustus
OCenW	Directie Kennis (3 pers) Onderzoekscoördinator DG Hoger Onderwijs Directie Wetgeving & Juridische Zaken (DWJZ)	1 juli 26 augustus 14 juli (e-mail en telefonisch)
SZW	FEZ & Kennisunit DG SZI	15 juli
VenJ	WODC (EWB, 4 pers.) en leiding Onderzoekscoördinator Ambtenaar betrokken bij Awb – evaluatie Beleidsmedewerker	8 juni, 20 oktober 15 juni (telefonisch) 11 augustus 25 augustus
VWS	FEZ Commissie Evaluatie Regelgeving (CER) bij ZonMw	15 september 8 april
Overige instituties		
Algemene Rekenkamer	2 pers.	1 september
Raad van State	Geen interview, korte telefonische communicatie	24 november
WRR	Top	16 november

*Omdat we de namen van de respondenten niet prijsgeven, is alleen hun organisatieonderdeel weergegeven.

Bijlage 3 Uitwerkingen van het evaluatiekader (internationaal)

Europese Commissie: Richtlijnen voor *Terms of Reference* en andere richtlijnen voor aanvragers (EC, 2013a)
&
UNEG: Criteria voor de beoordeling van ToR's (UNEG, 2010)

- Wettelijk/ regulerend, institutioneel kader
- Scope van de evaluatie: evaluatieobject, onderzoeksperiode, locatie in de beleidscyclus en (geografische) gebiedsafbakening
- Belangrijkste stakeholders en gebruikers van de evaluatie
- Evaluatieve en onderzoeksvragen
- Beschikbare kennis over het programma en eventuele effecten (mede van invloed op volgende stap)
- Belangrijkste te gebruiken methoden en technieken
- Planning
- Indicatief budget
- Benodigde kwalificaties van het team dat de evaluatie uitvoert
- Structuur van de offerte
- Regels m.b.t. offreren en beoordelingscriteria

European Evaluation Society (EES): Competenties van evaluatieonderzoekers (verkort)

1. Evaluation knowledge

1.1 Appreciates the distinctive role played by evaluation in society
1.11 Exhibits familiarity with evaluation theories, schools and approaches
1.12 Shows awareness of evaluation history and trends
1.13 Appreciates the linkages between evaluation and the social sciences
1.14 Understands program theory and its implications for evaluation
1.15 Aims at evaluation independence and excellence in all contexts

1.2 Masters the antecedents of evaluation quality
1.21 Uses appropriate evaluation concepts and correct evaluation terms
1.22 Displays a capacity to identify relevant evaluation questions
1.23 Knows how to engage constructively with all stakeholders
1.24 Comprehends the value of diverse evaluation approaches
1.25 Adapts evaluation designs and methods to specifi c contexts

1.3 Understands the potential and limits of evaluation instruments and tools
1.31 Data collection and analysis
1.32 Experimental and quasi experimental methods
1.33 Qualitative, participatory and mixed methods
1.34 Case studies, surveys, interviews, expert panels
1.35 Indicators, rating and monitoring systems

2. Professional practice

2.1 Demonstrates capacity to manage and deliver evaluations
2.11 Responds to legitimate stakeholders' needs and concerns
2.12 Assesses the evaluation context and identifi es the program logic
2.13 Manages resources and skills prudently so as to achieve results
2.14 Gathers, uses and interprets evidence with care and judgment
2.15 Reports fairly and encourages effective use of evaluation results

2.2 Displays interpersonal skills
2.21 Writes fluently and communicates clearly
2.22 Values team work and leads by example
2.23 Uses sound negotiating and confl ict resolution skills
2.24 Demonstrates gender awareness and cultural sensitivity
2.25 Nurtures professional relationships

3. Dispositions and attitudes

3.1 Upholds ethical standards and democratic values in the conduct of evaluations
3.2 Reaches out to clients and stakeholders
3.3 Evinces independence of mind and appearance
3.4 Displays self-awareness and pursues continuous professional development
3.5 Contributes to the evaluation community

Europese Commissie (EC): Criteria voor de beoordeling van evaluatierapporten (EC, 2013a)

Please assess the evaluation report in terms of your judgements as to how positively or negatively it met each criterion specified below:	**Very positive**				**Very negative**
1. Meeting needs: The evaluation report adequately addresses the requests for information formulated by the commissioners and corresponds to the terms of reference	☐	☐	☐	☐	☐
2. Relevant scope: The rationale of the programme, its outputs, results, impacts, interactions with other policies and unexpected effects have been carefully studied	☐	☐	☐	☐	☐
3. Open process: The interested parties – both the partners of the programme and the other stakeholders – have been involved in the design of the evaluation and in the discussion of the results in order to take into account their different points of view	☐	☐	☐	☐	☐
4. Defensible design: The design of the evaluation was appropriate and adequate for obtaining the results (within their limits of validity) needed to answer the main evaluative questions	☐	☐	☐	☐	☐
5. Reliable data: The primary and secondary data collected or selected are suitable and reliable in terms of the expected use	☐	☐	☐	☐	☐
6. Sound analysis: Quantitative and qualitative data were analysed in accordance with established conventions, and in ways appropriate to answer the evaluation questions correctly	☐	☐	☐	☐	☐
7. Credible results: The results are logical and justified by the analysis of data and by suitable interpretations and hypotheses	☐	☐	☐	☐	☐
8. Impartial conclusions: The conclusions are justified and unbiased	☐	☐	☐	☐	☐
9. Clear report: The report describes the context and goal, as well as the organisation and results of the programme in such a way that the information provided is easily understood	☐	☐	☐	☐	☐
10. Useful recommendations: The report provides recommendations that are detailed enough to be implemented	☐	☐	☐	☐	☐
In view of the contextual constraints bearing on the evaluation, the evaluation report is considered to be:	☐	☐	☐	☐	☐

UKES Guidelines for Commissioners

To ensure good practice in evaluation, it would be helpful if Commissioners:

- acknowledge the benefits of external, independent evaluation.
- operate fair tendering situations in which competitors ideas are not exploited or intellectual property misused as a result of commissioning.
- hold preliminary consultations with all parties to the evaluation to support a relevant, realistic and viable specification.
- specify the purpose and audience(s) for the evaluation with appropriate background material to encourage relevant tenders.
- operate a tendering procedure that is open and fair ensuring that appropriately qualified assessors are involved and making explicit criteria upon which a tender decision will be made.
- clarify the constraints that commissioners operate under, e.g. timescales, budgets, deadlines, and accountability.
- adhere to the terms agreed in the contract and consult with evaluators and other interest groups if significant changes are required to the design or delivery of the evaluation.
- specify the legal terms and responsibilities of the evaluation in the contract.
- match the aims and potential outcome of the evaluation to the knowledge and expertise of the potential evaluator(s).
- provide access to documentation and data required for evaluation purposes.
- establish clear principles for the reporting and dissemination of evaluation reports funded by public monies, consistent with acknowledged procedures which ensure quality evaluation and reporting.
- have realistic expectations on what an evaluation might provide including sufficient time for evaluators to respond to an initial invitation to tender and produce a proposal.
- include experienced evaluators (who are not potential applicants for funding) in initial drafts of evaluation specifications, including feasible budget and realistic timescales.
- have trust in and mutual respect for all stakeholders, participants, commissioners and evaluator(s).
- take advice of evaluators on research methodologies for collecting and analysing data.
- communicate openly and keep the evaluation team informed of changes in circumstances affecting the evaluation.
- recognise where evaluators need to keep their sources of information anonymous.
- preserve the integrity of the findings, e.g. by not quoting or publicising findings out of context.

Bijlage 4 Enkele uitwerkingen van het evaluatiekader bij beleidsdepartementen

Box B4.1 Tijdlijn van ontwikkelingen in het centrale evaluatiestelsel

1971-2 Nederland kende begin jaren zeventig de COBA - Commissie beleidsanalyse die als doel had: bevorderen van het gebruik van beleidsanalyse en evaluatie, binnen en buiten de rijksoverheid, door middel van opleidingen, publicaties en advisering, in samenwerking met wetenschappers en andere professionals. Deze commissie ontwikkelde voor de departementen producten als 'departementale doelstellingsanalyse'. Er was naar analogie met het Amerikaanse Planning *Programming and Budgeting System* (PPBS) een nauwe relatie tussen planning en de financiële invalshoek. COBA werd geen succes omdat er aan de top weinig draagvlak voor was.[161]
In 1981 is gestart met een reeks Heroverwegingsonderzoeken in het kader van de beheersing van de overheidsfinanciën, met als onderdeel van elk onderzoek een analyse van de gevolgen voor doelbereiking en doelmatigheid bij 20 procent budgetkorting (in 2014 her-ingevoerd als onderdeel van de beleidsdoorlichtingen).
In de jaren tachtig verscheen een aantal kritische rekenkamerrapporten waar uit bleek dat het inzicht in doeltreffendheid en doelmatigheid van overheidsbeleid (zeer) beperkt is.[162] In 1990 verscheen het rapport van de Commissie Programma evaluatie met een pleidooi voor institutionalisering van de evaluatiefunctie binnen de overheidsorganisatie.
Na een kabinetsstandpunt uit 1991 (TK 1991–1992, 22 032, nr. 27) ontwikkelde zich in de jaren negentig het officiële evaluatiebeleid van de regering (Leeuw, 2009). In het regeringsdocument wordt beleidsevaluatie aangeduid als vast onderdeel in beleidsprocessen (binnen de kaders van de al bestaande Comptabiliteitswet en andere regelingen). Ministers en topambtenaren zijn verantwoordelijk voor de uitvoering van evaluaties. Binnen de departementen dient voorzien te worden in een min of meer onafhankelijke evaluatiefunctie. Directies FEZ worden verantwoordelijk voor het stimuleren en coördineren van de evaluatiefunctie en voor het nagaan van kosten en opbrengsten van evaluaties. Uitkomsten van evaluaties dienen een rol te spelen in begrotingsbeslissingen; wat blijkt uit jaarverslagen waarin departementen duidelijk maken welke voorgenomen en afgeronde evaluaties er zijn, wat de resultaten zijn en hoe deze zijn gebruikt in de budgettaire besluitvorming.
Het document bevatte tevens een pleidooi voor het vaker uitvoeren van evaluaties van *wetgeving en juridische procedures*.
In 1999 start de VBTB-operatie (Van Beleidsbegroting tot Beleidsverantwoording) waarin beleidsinformatie en financiële informatie gekoppeld worden en waarin prestatie-indicatoren worden ingevoerd, gekoppeld aan begrotingsposten.
In de jaren nul neemt de betekenis van beleidsevaluatie verder toe doordat het een steeds duidelijker functie krijgt in de structuren voor beleidsvoering (VBTB, wetsevaluatie; Leeuw 2000). Het Regeerakkoord 1998–2002 bevat het voornemen tot bevordering van de brede implementatie van het resultaatgerichte sturingsmodel (outputsturing), door systematisch interdepartementale beleidsonderzoeken (IBO's) te verrichten naar aansturing en bedrijfsvoering. Het aantal organisaties (o.a. Algemene Rekenkamer en departementale FEZ-directies) dat evaluaties en controles uitvoert, groeit en tevens het aantal evaluaties en meta-evaluaties. Op het vlak van bijvoorbeeld milieu, emancipatie en veiligheid neemt ook het aantal *ex ante* effectevaluaties toe (Leeuw 2000).
In 2001 wordt de VIDE, ofwel de Dutch Evaluation Society (www.videnet.nl) opgericht.
In hetzelfde jaar treedt een nieuwe Comptabiliteitswet in werking, waarin evaluatie officieel een plek krijgt. De Regeling prestatiegegevens en evaluatieonderzoek rijksoverheid (RPE)

[161] Zie ook: http://www.beleidsevaluatie.info/home/een-korte-tijdlijn-van-beleidsevaluatie-in-nederland/.
[162] Wel vonden sinds 1981 Heroverwegingsonderzoeken plaats, die departementsoverstijgend waren en een directe relatie hadden met de begrotingscyclus.

wordt van kracht, met evaluatieverplichtingen voor de departementen, ex post en ex ante. De departementen dienen een duidelijk evaluatiebeleid te ontwikkelen en te voeren. De kwaliteit van evaluaties dient op peil te zijn, de organisatie van het evaluatieproces dient duidelijk belegd te zijn, en er dienen voorzieningen te zijn om het gebruik van de resultaten te bevorderen. In de aangepaste RPE van 2006 wordt als nieuw doelmatigheidsinstrument de *beleidsdoorlichting* geïntroduceerd teneinde beleid 'vooraf beter te onderbouwen en achteraf beter te evalueren' (IOFEZ, 2004). Elk departement wordt geacht om bevindingen uit verschillende doeltreffendheids- en andere evaluaties hierin samen te brengen (synthetiseren) op het niveau van de beleidsdoelen (artikelen) in de begroting.[163]
Sindsdien kent het centrale verantwoordingsstelsel twee hoofdinstrumenten: (1) de IBO's, in aard en aanpak vooral ex ante en gericht op alternatieven van beleid cq. heroverweging[164] en (2) de beleidsdoorlichtingen.
2008. In de Regeringsnota Vertrouwen in wetgeving waarin de nadruk ligt op terughoudendheid met wetgeving en ruimte wordt door de minister van Justitie de oprichting aangekondigd van een 'Clearing House voor systematische wetsevaluatie', met het oog op 'ervaringskennis op het terrein van voorbereiding van wetgeving, effecten van wetgeving en de uitvoering daarvan'.
2010 Het kabinet biedt de rapporten van 20 werkgroepen Brede Heroverwegingen aan de Tweede Kamer aan. De Brede Heroverwegingen, ter voorbereiding van politieke keuzes over toekomstige bezuinigingen, maken deel uit van het kabinetsbeleid om de financiële en economische crisis het hoofd te bieden (www.rijksoverheid.nl/documenten/kamerstukken/2010/04/01/aanbieding-heroverwegingsrapporten).
2012-2013 De Rekenkamer publiceert twee opeenvolgende overzichten van evaluatiestudies op rijksniveau die gaan over doeltreffendheid en doelmatigheid (OER). Iets minder dan de helft van het geld dat ministers jaarlijks uitgeven aan beleid (€94 miljard in 2011) is sinds 2006 42% (€ 36 miljard) daadwerkelijk op doeltreffendheid onderzocht – zoals verplicht op grond van artikel 20 Cw. Deels heeft dit te maken met gebrekkige aansluiting tussen evaluatieonderzoek en de formulering van begrotingsdoelstellingen. Voor een deel bestaan hier redenen voor bij beleidsmakers die de AR vervolgens heeft nagetrokken en in haar rapport van een algemene reactie voorziet. (1) Het is niet duidelijk hoe het beleid op doeltreffendheid kan worden geëvalueerd (2) Het is nog te vroeg om de doeltreffendheid te kunnen vaststellen (3) Het beleid wordt al door anderen op doeltreffendheid geëvalueerd (4) Doeltreffendheidsonderzoek voegt weinig toe (5) Het beleid is inmiddels stopgezet (6) De kosten van doeltreffendheid zijn te hoog (Algemene Rekenkamer, 2013:4).
Op 1 januari 2013 treedt een herziene 'Regeling Periodiek Evaluatieonderzoek' (RPE) in werking die verbetering van overzichten van evaluatie- en overig onderzoek vereist en beter inzicht in de programmering van beleidsdoorlichtingen in begrotingen en jaarverslagen. Daarnaast moet met Verantwoord Begroten (VB) met ingang van de begroting 2013 de informatievoorziening over evaluaties worden verbeterd.[165] In 2015 is de RPE opnieuw herzien. Met VB is het belang van beleidsdoorlichtingen verder toegenomen doordat de begrotingen en

[163] De rapporten zijn openbaar op de website van de Rijksbegroting: www.rijksbegroting.nl/beleidsevaluaties/evaluaties-en-beleidsdoorlichtingen/archief.
[164] Bijv. de IBO wetenschappelijk onderzoek (2014) had als centrale vraag: "Is de huidige inzet van middelen voor wetenschappelijk onderzoek optimaal voor het bereiken van een maximale maatschappelijke output?" voor de beantwoording waarvan internationale studies en ander materiaal is verzameld en voorts vragenlijsten zijn uitgezet en interviews zijn gehouden. Daarna zijn 5 beleidsvarianten beschreven met het oog op de toekomst – waaronder handhaving van de status quo.
[165] Voor elke evaluatie worden de titel, het betreffende artikel en start- en einddatum vermeld. Van afgeronde evaluaties wordt de vindplaats aangegeven, waar mogelijk met een hyperlink naar de evaluatie zelf. Ook wordt gestimuleerd om evaluatieonderzoek te gebruiken bij de motivering van beleidswijzigingen (http://www.rijksbegroting.nl/algemeen/overzichten,Verantwoord-begroten.html link naar flyer).

jaarverslagen minder informatie bevatten over de beleidsresultaten (website verantwoordingsonderzoek Rekenkamer).

Bijlage 4 *(vervolg)*

Voorbeeld experimentele en verklarende evaluatie (OCW)

Eén van de bestudeerde clusters van beleidsmaatregelen betreft 'Centrale toetsing en zeer zwakke scholen' – voor een deel gerelateerd aan het toenmalige *conceptwetsvoorstel Toetsing* (2011). In bijlage 4 geven we de door CPB geschetste designs weer voor wat betreft het deel centrale toetsing. Het beleid gericht op centrale toetsing bestaat grofweg uit vier componenten:
- Verplichte eindtoets
- Verplicht leerlingvolgsysteem
(beide verplichtingen zijn nieuw, maar in de praktijk gebruikt 85% van de basisscholen de Cito-eindtoets; 90-95% gebruikt een of meer onderdelen van het Cito-leerlingvolgsysteem, p. 21)
- Absolute kwaliteitsnorm voor scholen
- Invoering begintoets

Voor voortijdig experimenteren met de eerste twee componenten was geen ruimte, omdat het voornemen was deze componenten in te voeren met ingang van schooljaar 2012-2013; voor de laatste twee componenten is die ruimte er wel, omdat die voor niet eerder dan 2014-2015 zijn voorzien.

Effectenevaluatie
Voor de evaluatie van de effecten van het beleid zijn er verschillende mogelijkheden. Deze mogelijkheden hangen af van de ruimte die beschikbaar is voor het uittesten van het nieuwe beleid.
Een eerste mogelijkheid voor evaluatie is het analyseren van bestaande gegevens. Op dit moment gebruikt ongeveer 85% van de scholen de Cito-eindtoets en 90 tot 95% van de scholen gebruikt één of meer onderdelen van het Cito-leerlingvolgsysteem. Een mogelijkheid is om scholen te zoeken die de afgelopen jaren zijn gestart met één van beide of beide componenten van centrale toetsing en de resultaten van deze scholen te analyseren voor en na de start van het werken met deze componenten. Aangezien er nog veel meer factoren in de tijd kunnen zijn veranderd is het goed om bij deze analyse ook een controlegroep te gebruiken. Dat zou de groep scholen kunnen zijn die al werkte met de genoemde componenten van centrale toetsing. Een mogelijkheid is om hiervoor gebruik te maken van de gegevens uit het PRIMA-onderzoek. In het PRIMA-onderzoek wordt gemeten of scholen een leerlingvolgsysteem gebruiken en of ze deelnemen aan de CITO-toets. Onderzocht kan worden of er scholen zijn die eerst niet deelnamen aan de CITO-eindtoets en later wel, of scholen die eerst geen leerlingvolgsysteem gebruikten en later wel. Basisscholen moesten al een leerlingvolgsysteem hebben voor hun zorgleerlingen. Het PRIMA-onderzoek bevat echter ook gegevens over het gebruik van het leerlingvolgsysteem. Voor de scholen die één van beide componenten (anders) zijn gaan gebruiken kunnen de resultaten op de PRIMA-toetsen voor en na het gebruik van deze componenten worden vergeleken en kunnen *difference-in-differences*-modellen worden geschat waarbij scholen die altijd al deelnamen aan de Cito-eindtoets als controlegroep worden gebruikt. Ook kan, analoog aan het onderzoek naar het Amsterdamse beleid (Chorny et al. 2010) gekeken worden of er in meer steden of regio's al beleid is ingevoerd dat leidde tot verplichte deelname aan de CITO-eindtoets.
(...)
Een tweede mogelijkheid betreft een vergelijking van een voor- en nameting. Voor de invoering van de verplichte eindtoets is dit echter niet goed mogelijk, omdat er geen vergelijkbare toetsgegevens zijn bij de voormeting (scholen doen immers nog niet mee aan de eindtoets). Deze aanpak kan echter wel worden gebruikt voor het vaststellen van het effect

van de invoering van een leerlingvolgsysteem. De voor- en nameting zouden dan nog kunnen worden uitgebreid met een controlegroep van scholen die al wel een leerlingvolgsysteem gebruiken, waardoor een difference-in-differencesopzet ontstaat.'

Verklarende evaluatie
Voor een verklarende evaluatie van het onderdeel **centrale toetsing** is het van belang na te gaan:
a. welke evidentie er is op basis van onderzoek en ervaring van leerkrachten, schoolleiders, besturen, inspectie en deskundigen voor de afzonderlijke veronderstellingen in deze sets;
b. hoe scholen en leerkrachten op centrale toetsing verwachten te reageren (en waarom) in de wijze van onderwijs geven, de accenten in het lesprogramma, de mate van differentiëren tussen leerlingen, de wijze van toetsen, enz. (ex ante);
c. hoe zij feitelijk reageren en waarom (pilots, ex post);
d. of en hoe leerkrachten, schoolleiders, besturen, inspectie en deskundigen na enkele jaren toe- of afgenomen leerwinst verklaren (in hoeverre is de centrale toetsing en in hoeverre zijn andere veranderingen daar debet aan). Visies van leerkrachten en deskundigen op deze punten kunnen deels op basis van beschikbare data worden getoetst. (...)

Stappenplan verklarende evaluatie (OCW)

Pater et al. (2012) dragen onderzoekers een aantal stappen aan om tot verklarende evaluatie te komen, en passen deze vervolgens toe op een tweetal cases rond voortijdig schoolverlaten (VM2 en De Wijkschool). Deze laten zich als volgt samenvatten:

1. **Vooraf: Desk research**
- Verzamel/ activeer kennis over het betreffende beleidsthema (wetenschappelijke / evaluaties; inzichten van experts)
- Reconstrueer globaal de beleidscontext waar de beleidsinterventie uit voortkomt (documenten, wetsvoorstellen, nota's, adviezen etc.)
- Ga na wat het veronderstelde probleem is en of het plausibel is dat het probleem overwonnen wordt door de veronderstelde veranderingsmechanismen
- Indien beschikbaar, analyseer de al bestaande effectevaluaties met het doel te bepalen of de effecten gerelateerd zijn of kunnen worden aan verschillen in de context waarin interventies zijn gepleegd.

2. **Reconstructie van de beleidstheorie: hoe wordt dit beleid geacht te werken?**
- Reconstrueer de beleidsinterventie en de achterliggende beleidstheorie aan de hand van beleidsdocumenten. Construeer schema's van causale ketens aan de hand van veronderstellingen in de beleidstheorie en evt. empirisch onderzoek.
- Expliciteer de beleidstheorie in dialoog met beleidsmakers
- Analyseer de verkregen informatie uit interviews op impliciete en expliciete causaliteiten, mechanismen, differentiatie in contexten, doelgroepen, belanghebbenden, bedoelde en onbedoelde effecten. Pas de reconstructie van de beleidstheorie en de schema's van causale ketens aan.

3. **Confrontatie van de beleidstheorie met de literatuur** (in ieder geval bij ex ante, optioneel bij proces- en ex post evaluatie)
4. **Veldonderzoek** (alleen bij proces- en ex post evaluatie). In het veldonderzoek wordt de gereconstrueerde beleidstheorie getoetst aan de doelgroep en andere betrokkenen. Dit kan uiteraard alleen als de interventie al in gang gezet is. Formuleer voor elke informant relevante en te beantwoorden vragen en denk na over de meest geschikte onderzoeksmethode voor elk type betrokkenen.

De auteurs betogen dat het een iteratief proces betreft: 'De stopregel wordt gevormd door het niet meer verkrijgen van volledig nieuwe informatie.' (2012:90).

Tot slot hoort de **kritische analyse en evaluatie** tot de kern van de verklarende methodiek. Dit is de analyse waarin de confrontatie tussen beleidstheorie en praktijk wordt aangegaan, waarin in de diepte ontrafeld wordt welke nieuwe of aangepaste mechanismen de interventie heeft veroorzaakt met gewenste gedragsveranderingen als uitkomst.
Dan zou ook duidelijk moeten worden of een interventie (deels) wel en/of (deels) niet heeft gewerkt. Dat kan in essentie aan het volgende liggen: 1. Het beleidsontwerp was goed, maar het is niet goed geïmplementeerd; het is gebleken dat uitvoerders van het beleid niet doen wat de beleidsbedoeling was. Als de implementatie wel conform de bedoeling is gedaan, kan het zijn dat: 2. het onderliggende mechanisme niet op gang is gebracht, bijvoorbeeld omdat de juiste context ontbrak, of 3. er een breuk is in de causale keten, één schakel heeft niet gewerkt.
De auteurs bieden op basis van de literatuur voorts richtpunten voor wat bij kwalitatief onderzoek als dit, bruikbare methodologische kwaliteitscriteria (kunnen) zijn.[166]

Kwaliteitscriteria voor verklarende evaluatie volgens de evaluatoren van Kohnstamm (2012)

- Validiteit – via de reconstructie van het proces : de gebeurtenissen in een case worden geordend en er wordt nagegaan in hoeverre die als causaal proces met elkaar in verband kunnen worden gebracht.
- Betrouwbaarheid of repliceerbaarheid – via aanpak-transparantie en triangulatie van bronnen
- Generaliseerbaarheid – via o.a. *multiple case study* designs[167]
- Gevoeligheid – onder welke condities gaat het beter/ minder?
- Causale validiteit – hiertoe dient een evaluatie zo ingericht te worden, dat een effect van andere variabelen vooraf en/of achteraf zoveel mogelijk uitgesloten wordt.
- Bruikbaarheid – de evaluatie bevat aanknopingspunten voor beleid
- Fijnmazigheid – welk (bijv. psychologisch) mechanisme ligt precies aan gevonden fenomenen ten grondslag

Checklist evaluatie<u>opzet</u> FEZ (EZ)

1. Aanleiding

- De aanleiding voor de evaluatie is duidelijk omschreven

2. Beleidsonderbouwing

- De probleemanalyse is beschreven (wat is het probleem en waarom interventie)
- Het verband tussen de overheidsinterventie (bijvoorbeeld subsidie), activiteiten en *outcome* is beschreven
- Er zijn heldere meetbare doelen en indicatoren benoemd

3. Probleemstelling en onderzoeksvragen

- De probleemstelling of centrale onderzoeksvraag is duidelijk

[166] Bij black-box of experimenteel onderzoek is interne validiteit het belangrijkste kwaliteitscriterium: hoe hard is een relatie te leggen tussen interventie(s) en uitkomsten? (o.a. Farrington c.s. 2004) – hoewel de genoemde criteria ook hier zeker van toepassing zijn.
[167] Waarbij bevindingen uit een eerste case als veronderstelling worden ingebracht in de tweede etc., of waarbij de bevindingen van verschillende, onafhankelijk uitgevoerde cases naast elkaar worden gelegd en vergeleken (zie bijvoorbeeld Miles en Huberman, 1994).

- De probleemstelling of centrale onderzoeksvraag sluit aan bij de aanleiding van de evaluatie en de probleemanalyse
- Bij een beleidsdoorlichting: alle elementen/vragen uit de RPE zijn benoemd
- Bij een individuele evaluatie: de benoemde onderzoeksvragen geven antwoord op alle relevante elementen van een beleidsdoorlichting om zodoende als basis gebruikt te worden voor de beleidsdoorlichting

4. Onderzoeksaanpak

- De causale relatie tussen beleidsinstrument(en), de activiteiten en het vermeende effect worden onderzocht
- Er is aangegeven of een nulmeting is gedaan
- Er is aangegeven of er mogelijk een controlegroep bestaat
- De beperkingen en onmogelijkheden ten aanzien van het meten van effecten worden benoemd
- Duidelijk is op welke alternatieve wijze de doeltreffendheid wordt onderzocht, indien er geen nulmeting en controlegroep is (bepalen plausibiliteit)..
- Verkennen van en advies over verbeteringen in de mogelijkheden tot effectmeting.
- De wijze waarop data worden verzameld staat beschreven (databases, interviews, enquête)
- De beperkingen ten aanzien van uitspraken over de doeltreffendheid bij een alternatieve onderzoeksaanpak (geen nulmeting/controlegroep) zijn benoemd

5. Uitbesteden of zelf doen

- Er is aangegeven of en waarom het onderzoek wordt uitbesteed of zelf wordt gedaan

6. Tijdpad

- Het tijdpad van het onderzoek (start- en einddatum en tussenproducten) is helder

7. Organisatie en kosten

- Aanspreekpunt voor het onderzoek (projectleider) is benoemd
- Er is aangegeven hoe het onderzoek wordt begeleid
- Er is aangegeven welk budget beschikbaar is voor het onderzoek

8. Communicatie

- Er is aangegeven of de evaluatie wel/niet wordt aangeboden aan de Tweede Kamer

Checklist evaluatie<u>rapport</u> FEZ (EZ)

1. Volledigheid

- Het rapport geeft antwoord op de centrale onderzoeksvraag
- Het rapport geeft antwoord op alle (deel)vragen uit de evaluatieopzet/offerte

2. Duidelijkheid

- De conclusie over de doeltreffendheid van het beleid is helder (het beleid is wel/niet effectief of kan niet worden aangetoond, maar wel indicaties dat ……)
- Het rapport maakt duidelijk onderscheid tussen de bevindingen, de analyse en de conclusies die daaruit getrokken worden

3. Methodologische verantwoording

- Het rapport geeft aan in welke mate de bevindingen valide en betrouwbaar zijn
- De strekking ('hardheid') van de conclusies sluit aan op de validiteit en betrouwbaarheid van de bevindingen
- Er staat voldoende informatie in het rapport om te kunnen beoordelen hoe het onderzoek is uitgevoerd, hoe de variabelen zijn gemeten

4. Aanbevelingen

- De aanbevelingen zijn een logisch gevolg van de conclusies
- De aanbevelingen zijn bruikbaar (toepasbaar en zinvol)
- Het relatieve belang van de aanbevelingen komt duidelijk naar voren

De evaluatie van de WBSO (EZ)

De Wet Bevordering Speur- en Ontwikkelingswerk (WBSO, 1994) heeft als doelstelling om onderzoek en ontwikkeling door Nederlandse bedrijven en ondernemers te stimuleren door de loonkosten van onderzoekend personeel te verlagen door middel van een fiscale tegemoetkoming. Het budget voor de WBSO bedraagt in 2012 872 miljoen euro. Sinds 1994 wordt het budget van de WBSO jaarlijks vastgesteld in het Belastingplan, de realisatie kan afwijken van het budget. Het fiscale instrumentarium binnen de wet is voor de periode 2006-2010 door EIM kwantitatief geëvalueerd; overigens niet voor het eerst. In de loop der jaren zijn door technologische vooruitgang steeds geavanceerdere evaluatietechnieken toepasbaar geworden. In de huidige evaluatie – gebaseerd op een dynamisch paneldatamodel dat geschat is met behulp van non-lineaire regressie – zijn een aantal vragen (hypothesen) getoetst (EIM, 2012).

Kernhypothese
1) De WBSO heeft een positief effect op de totale R&D-loonuitgaven van een onderneming.
Hypothese ter ondersteuning kernhypothese
2) Bedrijven die relatief veel subsidie krijgen geven relatief meer uit aan R&D-loon dan bedrijven die een lager subsidiepercentage hebben.
Hypotheses voor toetsing uitstralingseffecten
3) De WBSO zorgt voor meer innovatie op bedrijfsniveau.
4) De WBSO leidt tot *spill-overs* in de eigen sector.
5) De WBSO leidt tot *spill-overs* buiten de eigen sector.
6) De WBSO zorgt voor een hogere toegevoegde waarde per werkende.

Het betreft een technisch-kwantitatieve evaluatie op basis van een geabstraheerd model, zoals de onderstaande formule illustreert[168]:

$$\ln Y_{it} = \alpha_i + \gamma \ln Y_{it-1} + \delta \frac{WBSO_{it-1}}{Y_{it-1}} + \beta X_i + \mu_t + U_{it}$$

We gaan hier verder niet in op de bevindingen, maar het voorbeeld laat zien dat wetsevaluaties behalve kwalitatief-juridisch ook kwantitatief-economisch kunnen zijn.

[168] Hierin is *Y* de totale R&D-loonsom in een bedrijf in een jaar, *ai* een fixed effect per bedrijf, *WBSO* het bedrag dat vanwege de WBSO minder aan belasting hoeft te worden betaald, *X* een reeks aan controlevariabelen en *μt* een fixed effect voor de jaren. *U* is de storingsterm.