

Hoofdstuk 7

Beleidsvaluatie

Preprint versie van volgende publicatie—all rights reserved.

*Pattyn Valérie (2020). 'Hoofdstuk 7 Beleidsvaluatie'. In: Wayenberg Ellen. **Bouwen aan beleid. Het proces van de overheid** Gent: Owl press (pp.182-204).*

Werkt het beleid? De relevantie van beleidsvaluatie

‘De plenaire vergadering van 27 mei 2020 keurde de oprichting goed van een Commissie ad hoc ter evaluatie van het Vlaamse COVID-19-beleid en tot opzetten van een post-COVID-19-traject. Deze tijdelijke commissie heeft als opdracht een evaluatie op te maken van de voorbije COVID-19-crisisperiode, hieruit de nodige lessen te trekken voor de toekomst, en eerste aanzetten te geven voor een post-COVID-19-traject.’ (Vlaams Parlement 2020)

Met bovenstaande aankondiging werd de oprichting van een Parlementaire Commissie ad hoc voor de evaluatie van het Vlaamse COVID-19-beleid een feit. De COVID-19-crisis noodzaakte overheden tot een sterk ingrijpen in ieders vrijheden, en heeft geleid tot heel wat nieuwe maatregelen, met als doelstelling bijvoorbeeld het (zoveel als mogelijk) beperken van de negatieve economische effecten van de crisis. De vraag stelt zich nu of deze COVID-19-maatregelen ook zinvol zijn. Werden de gewenste effecten behaald? Hoe beoordelen belanghebbenden de maatregelen? En hoe kan het in de toekomst eventueel beter? Dergelijke vragen komen idealiter aan de orde in de Commissie ad hoc evaluatie van het Vlaamse COVID-19-beleid, en schreeuwen om een systematisch antwoord.

Uiteraard is het beantwoorden van de vragen over de effecten van beleid niet enkel in crisistijd cruciaal. Het is een kernelement van goed bestuur dat altijd aan de orde is. De overheid maakt bepaalde keuzes over welke beleidsinstrumenten worden ingezet om bepaalde democratisch afgesproken doelstellingen te bereiken, en keuzes over de publieke middelen die hiertoe worden aangewend. Het is dan ook belangrijk te weten of ‘het goede werd gedaan’ en ‘of dit goed werd gedaan’. Met andere woorden: van beleid mogen we verwachten dat het de beoogde resultaten oplevert, en dat deze resultaten ook op een efficiënte en doelmatige manier werden bereikt (van der Knaap et al. 2020). Goede kennis is dan ook cruciaal. Deze kennis wordt geleverd door beleidsvaluatie.

Hieronder bespreken we wat precies moet worden verstaan onder beleidsvaluatie, en wat hierbij de kernelementen zijn. Zoals zal blijken, is beleidsvaluatie een veelomvattend concept dat meerdere functies, types en actoren betreft. Daarna verkennen we de theorievorming in beleidsvaluatieonderzoek. We bespreken achtereenvolgens meer ‘prescriptieve theorievorming’ voor evaluatie, en theorievorming over evaluatie.

Beleidsvaluatie: een draak met meerdere koppen

Het bepalen van de meerwaarde van beleid

Wat is beleidsevaluatie? Zoals het woord zelf suggereert, omvat het twee sleutelaspecten: ‘beleid’ en ‘e-value-ren’. De laatste term verradt de essentie van beleidsevaluatie: Beleidsevaluatie is van nature **normatief**, en gaat over het vellen van een oordeel over de vraag of het overheidsingrijpen ‘goed’ is geweest. Een van de meest geciteerde kernachtige definities van evaluatie maakt dit normatieve element eveneens duidelijk: ‘Evaluatie betreft het bepalen van de meerwaarde van datgene wat geëvalueerd wordt’ (Scriven 1991).

Zulke waardeoordelen worden uiteraard nogal frequent in de mond genomen, en vele burgers hebben bijvoorbeeld over de COVID-19-maatregelen een eigen **opinie** klaar, of het beleid volgens hen al dan niet succesvol is of ‘werkt’. Beleidsevaluatie onderscheidt zich echter van deze meer informele opinies door haar **systematische karakter**. Beleidsevaluatie gebeurt volgens een bepaalde methodiek, die kan variëren afhankelijk van de evaluatiebehoefte, en gebeurt op een specifiek moment, dus niet zomaar ad hoc.

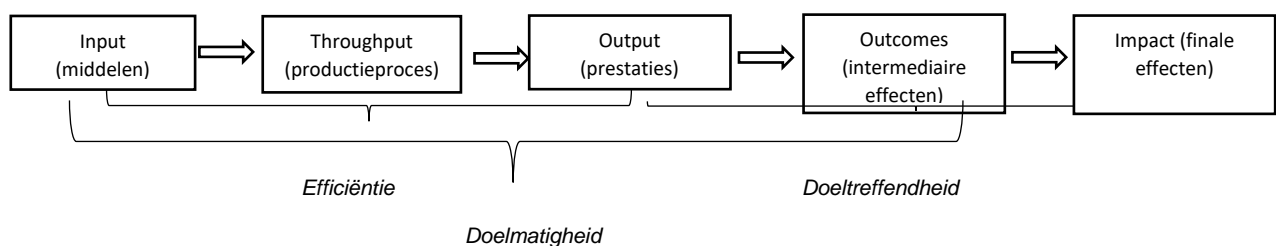
Beleidsevaluatie laat zich ook het beste interpreteren als een vorm van **toegepast onderzoek**, waarbij bepaalde gegevens, informatie, en kennis worden vergaard. Vaak kan men zich hiervoor beroepen op bestaande administratieve gegevens van bijvoorbeeld Statistiek Vlaanderen, maar in de meeste gevallen dient men ook nieuwe, primaire data te vergaren specifiek voor het evaluatieonderzoek, via bijvoorbeeld interviews, focusgroepen of een enquête.

Verder zal beleidsevaluatie zich richten op bepaalde **criteria** of **maatstaven** waaraan het beleid zal getoetst worden. De doeltreffendheid (een synoniem voor effectiviteit), doelmatigheid en efficiëntie van beleid zijn de meest courante evaluatiecriteria waaraan beleidsmaatregelen zullen worden getoetst. Zoals bijvoorbeeld de Vlaamse beleidsnota Economie, Wetenschapsbeleid en Innovatie 2019-2024 vermeldt: ‘We evalueren de effectiviteit en efficiëntie van strategische transformatiesteun en ecologiesteun en maken gebruik van de resultaten van deze evaluatie om deze maatregelen te optimaliseren zodat deze maximaal bijdragen aan de internationalisering van Vlaamse innovatie en het aantrekken van kennisintensieve buitenlandse investeringen.’ (Vlaamse Regering 2019a)

In bovenstaand voorbeeld vormt de ‘strategische transformatie- en ecologiesteun’ het zogenaamde **evaluandum** van de beleidsevaluatie. Dit evaluandum verwijst eenvoudigweg naar ‘wat er geëvalueerd wordt’. Beleidsevaluatie richt zich op ‘beleid’. In de praktijk kan dit evaluandum een individuele beleidsmaatregel zijn, een pakket van maatregelen, of een programma.

De vraag stelt zich vervolgens wat we precies bedoelen met een **effect** en met een doeltreffend, doelmatig of efficiënt beleid. Onderstaand input-impactmodel (figuur 7.1) maakt een en ander duidelijk.

Figuur 7.1: Input-Impactmodel



Dit model bestaat uit een aantal bouwstenen (zie ook van der Knaap et al. 2020):

- de in te zetten middelen, mensen en tijd (in evaluatiejargon: *input*); te meten in euro's, maanden, jaren...;
- de instrumenten en de wijze waarop deze worden ingezet in de uitvoering van beleidsmaatregelen (*throughput*); te meten in processen die worden ontplooid;
- de te behalen prestaties of operationele doelen (*output*). Deze prestaties verwijzen naar de producten of diensten die door de overheid of andere actoren in de uitvoering worden geleverd. De output kan worden gemeten in aantallen en kwaliteit van prestaties;
- de intermediaire effecten (*outcomes*) die nodig zijn om de finale maatschappelijke effecten te bereiken;
- de uiteindelijk te bereiken finale doelen of maatschappelijke effecten die met de inzet van een beleidsinstrument worden beoogd. Naar deze finale effecten wordt in het jargon ook vaak verwezen met *impact*.

De pijlen in de figuur zijn niet onbelangrijk, en staan voor veronderstellingen die samen de **interventiologica** of 'beleidstheorie' achter het beleid weergeven. Met de inzet van bepaalde beleidsinstrumenten hebben beleidsmakers in principe als bedoeling bepaalde maatschappelijke effecten te realiseren. Vaak zijn deze veronderstellingen echter enkel impliciet aanwezig, maar kunnen ze via documentenanalyse of een interview met beleidsmakers worden gereconstrueerd.

Neem bijvoorbeeld de subsidie van zonnepanelen door de Vlaamse overheid. De veronderstelling hierbij is dat via de subsidie van zonnepanelen burgers sneller geneigd zullen zijn dergelijke panelen te installeren, en op die manier kan bijgevolg het aandeel duurzaam opgewekte energie toenemen. Het uitvoeren van deze subsidieregeling kost geld: voor de subsidies zelf en voor het uitvoeringsproces van de regeling (de inzet van mensen, middelen en tijd; input). Het aantal verstrekte subsidies betreft in dit geval de output van het beleid (of de beleidsprestaties). De intermediaire effecten verwijzen naar het beoogde effect van de verstrekte subsidies, dus de plaatsing van een bepaald aantal zonnepanelen. Een toename in aantal zonnepanelen wordt verder verondersteld bij te dragen aan een groter aandeel duurzaam opgewekte energie (impact).

De accolades geven weer waar bovengenoemde evaluatiecriteria efficiëntie, doelmatigheid, doeltreffendheid/**effectiviteit** voor staan:

- 'efficiëntie' betreft de verhouding tussen de aantallen en de kwaliteit van de geleverde prestaties enerzijds en de kosten anderzijds;
- 'doelmatigheid' wordt hier vaak mee verward, maar wijst op de verhouding tussen de ingezette middelen en de effecten. Doelmatig beleid is gericht op een goede balans tussen kosten en effecten, en zal vaak relatief worden bepaald, bijvoorbeeld in vergelijking in de tijd of met andere landen;
- het sleutelcriterium 'effectiviteit' betreft de mate waarin het bereiken van de doelstellingen toe te schrijven is aan het gevoerde beleid.

Belangrijk is dat de effectiviteit dus een stap verder gaat dan louter **doelbereik**. Zelfs als men via monitoringdata zou vaststellen dat het aandeel duurzame hernieuwbare energie is toegenomen, en de doelstellingen van de Vlaamse overheid zijn gerealiseerd, dient een

evaluator te onderzoeken of er geen andere factoren een positieve of negatieve invloed hebben gehad op het bereiken van de doelstellingen. Mogelijk hadden veel burgers ook zónder subsidies immers de stap gezet naar de aankoop van zonnepanelen, wat de **meerwaarde** van het beleid in een ander perspectief zet. Met andere woorden: een effectevaluatie richt zich op het openen van deze **causale black box** en zal onderzoeken ‘in welke mate’, of ‘hoe en waarom’ subsidies een effect hebben gehad op het bereiken van de doelstellingen.

Bovenstaande evaluatiecriteria bieden geen exhaustief plaatje. Een evaluator kan zich eveneens richten op een veelheid aan andere criteria, met onder meer de relevantie van beleid (komt het beleid tegemoet aan de noden van de doelgroep?), de duurzaamheid van beleid (de mate waarin de positieve effecten na het beëindigen van het beleid aanwezig blijven), de coherentie van beleid (de mate waarin het beleid contradictorisch werkt ten aanzien van gelijkaardige initiatieven) als mogelijke alternatieven. Aan het begin van elke evaluatiestudie zal moeten worden bepaald welke evaluatiecriteria meer specifiek moeten worden onderzocht. Hoe je evalueert, zal immers afhangen van wat en waarom je evalueert.

Een veelheid aan evaluatiefuncties

Dit brengt ons bij de hamvraag waarom evaluatiestudies gebeuren: wat kan ermee worden bereikt? We maakten al eerder gewag van het belang van beleidsevaluatie in het kader van **goed bestuur**. In de evaluatieliteratuur worden traditioneel minimaal een viertal functies omschreven (Vedung 1997; De Peuter et al. 2007; van der Knaap et al. 2020): beleidsleren, verantwoorden, het ondersteunen van beleidsplanning en de strategisch-politieke functie van evaluatie.

De twee spilfuncties van beleidsevaluatie zijn, vanuit een rationeel perspectief, ‘leren’ en ‘verantwoorden’, met als algemeen doel het verbeteren van beleid dat werkt, of het stoppen van beleid dat niet werkt (en dus het stoppen van uitgeven van publieke middelen aan onvoldoende effectief beleid). Ook de voormalige Amerikaanse president Barack Obama verwees ernaar in zijn inaugurale rede (20 januari 2009): *‘The question we ask today is not whether our government is too big or too small, but whether it works – whether it helps families find jobs at a decent wage, care they can afford, a retirement that is dignified. Where the answer is yes, we intend to move forward. Where the answer is no, programs will end.’*

Beleidsleren kan op verschillende manieren worden bekeken, waarbij het onderscheid tussen *single loop learning* en *double loop learning*, of **bijsturend of vernieuwend leren**, het meest gangbaar is. Bijsturend leren betreft het aanpassen of verbeteren van beleid binnen de grenzen van de bestaande doelstellingen en aanpak. Vernieuwend leren is fundamenteeler van aard, en heeft betrekking op de onderliggende doelstellingen en veronderstellingen van beleid. Beleidsmaatregelen kunnen dan volledig worden vervangen ten gevolge van een evaluatie. Wanneer een evaluatie van het subsidiebeleid van zonnepanelen bijvoorbeeld leidt tot een aanpassing aan de manier waarop subsidies verdeeld worden, of aan het subsidiebedrag, betreft dit *single loop learning*. Wanneer daarentegen in plaats van subsidies gekozen wordt voor een heel ander beleidsinitiatief (bijvoorbeeld eisen bij nieuwbouw instellen) om het aandeel duurzame hernieuwbare energie te doen toenemen, is *double loop learning* aan de orde (van der Knaap et al. 2020). Hoewel leren heel vaak wordt gepredikt in discours, blijft dit in de feiten vaak beperkt tot *single loop* leren. Het fundamenteel herzien van doelstellingen of beleidsmaatregelen ligt politiek dikwijls gevoeliger.

Het afleggen van **verantwoording** is een volgende belangrijke evaluatiefunctie, en kan zowel vanuit democratisch als vanuit managementoogpunt worden benaderd. Overheidsorganisaties hebben de plicht om verantwoording af te leggen over de werking van de diensten enerzijds en de resultaten van het beleid anderzijds. Sturings- en verantwoordingsrelaties bestaan in verschillende vormen en soorten: onder meer tussen parlement en regering, centrale administraties en decentrale of verzelfstandigde agentschappen, tussen overheid en burger. Evaluatierapporten kunnen een rol spelen bij het legitimeren van maatregelen en uitgaven, en kunnen letterlijk de vraag beantwoorden of uitgaven ‘verantwoord’ zijn, in het licht van bijvoorbeeld het behaalde effect van de maatregelen (van der Knaap et al. 2020).

Volgend voorbeeld uit de Vlaamse Beleidsnota Energie 2009-2014 (Vlaamse Regering 2009a) verraadt eveneens een verantwoordingsmotief: ‘Op voorwaarde dat ze een positieve evaluatie kregen, verlengen we de energieconsulenten.’ Ook in de COVID-19-crisis werd er vanuit diverse hoeken expliciet opgeroepen tot een sterkere evaluatie- en verantwoordingscultuur: ‘Is dat wel goed besteed, zetten we in op de juiste prioriteiten? ... Elke aanbeveling in dit land heeft maar zin als ze het belang van een betere bestuurs- en evaluatiecultuur daarbij onderlijnt.’ (Van de Cloot 2020)

Naast beleidsleren en verantwoorden, wordt **beleidsplanning** steevast als een andere mogelijke functie van evaluatie gezien. Een brede waaier van evaluatiecriteria kan hierbij ingezet worden. Zo kan bijvoorbeeld de draagwijdte, dringendheid of haalbaarheid van een beleidsvoornemen getoetst worden. Ook beleidsrelevantie kan het voorwerp zijn van een evaluatie vooraleer men beslist tot de uitvoering ervan. Verder kan men alternatieve beleidsinitiatieven of varianten van eenzelfde beleidsinitiatief onderling afwegen en vergelijken om een keuze te onderbouwen. Evenzeer kan de interne coherentie of samenhang tussen maatregelen en doelstellingen beoordeeld worden, net als de vraag of het voorgestelde beleid consistent is met initiatieven binnen andere beleidsdomeinen. Een voorbeeld uit de Vlaamse Beleidsnota Mobiliteit en Openbare Werken 2009-2014 (Vlaamse Regering 2009b) richt zich expliciet op deze planningsfunctie: ‘De studies voor de realisatie van een opengetijdedok worden voortgezet. Er zal een maatschappelijke kosten-batenanalyse worden uitgevoerd van de verschillende planopties.’

Tot slot dienen we ook expliciet de **strategisch-politieke functie** van evaluatie te erkennen. Een evaluatie kan bijvoorbeeld worden gehanteerd om bepaalde thema’s op de politieke agenda te zetten, of om bepaalde keuzes te legitimeren. In een consensusdemocratie als België worden evaluaties ook regelmatig als pasmunt ingezet om wetten of decreten door te voeren (Sociaal-Economische Raad van Vlaanderen 2015). Men stemt dan in met het voeren van een evaluatie na afzienbare tijd, in ruil voor steun aan bepaalde regelgeving. Als voorbeeld wordt in dit verband wel eens naar de abortuswetgeving verwezen (Varone et al. 2005; Rekenhof 2018).

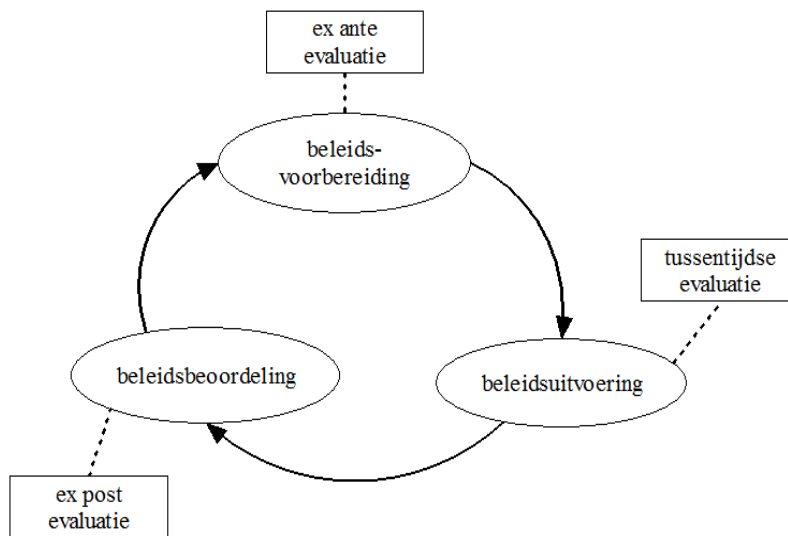
Evaluatieonderzoek kan ook voor meer **duistere** functies worden gehanteerd, waarbij evaluatieonderzoek kan worden gebruikt om besluitvorming uit te stellen, politieke tegenstanders in een kwaad daglicht te stellen en verantwoordelijkheden te ontlopen. Beleidsevaluatie blijft bovenal intrinsiek een politiek gebeuren, ook al worden studies op een zo objectief mogelijke en gedegen wetenschappelijke manier uitgevoerd.

Types beleidsevaluatie: terugblikken én vooruitblikken

De expliciete vermelding van beleidsplanning als een van de motieven om te evalueren mag enigszins bevreemdend overkomen vanuit een **traditionele visie** op beleidsevaluatie, en de

traditionele visie op de beleidscyclus in het bijzonder. In de standaardindeling van het beleidsproces wordt beleidsevaluatie inderdaad enkel beschouwd als de fase ná de implementatie van beleid.¹ De overheid maakt een keuze voor bepaalde beleidsmaatregelen om een gepercipieerd probleem op te lossen, implementeert deze, en vervolgens worden de resultaten beoordeeld (beleidsevaluatie). Gieten we dit proces in een typisch cyclusmodel, dan wordt nog geen recht gedaan aan de meer **comprehensieve visie** op beleidsevaluatie, waarbij beleidsevaluatie zowel vooraf, tijdens als na de uitvoering van het beleid een volwaardige plaats heeft. Er wordt in dit geval gesproken van respectievelijk *ex-ante*-evaluatie, *ex-durante*-evaluatie (tussentijdse) of *ex-post*-evaluatie, zoals afgebeeld in figuur 7.2.

Figuur 7.2: Beleidsevaluatie doorheen de beleidscyclus



Bron: De Peuter et al. (2007)

Het doel van de beleidsevaluatie *ex ante* is precies om voorafgaand aan de besluitvorming van nieuw beleid meer inzicht te verkrijgen in het te verwachten doelbereik, de doeltreffendheid, doelmatigheid en uitvoerbaarheid van beleidsmaatregelen, of de beleidslogica te toetsen in een pilootsetting. Een gedegen *ex-ante*-evaluatie kan op die manier de kansen op succesvol beleid doen toenemen, en kan helpen de logica achter beleid op punt te stellen of te toetsen. De uitkomsten van een *ex-ante*-evaluatie worden bij voorkeur voorgelegd aan de organen in het openbaar bestuur waar beleidsbeslissingen genomen worden en budgetten worden toegekend. Op die manier worden ze in de gelegenheid gesteld om debat te voeren, wat tot beter geïnformeerde besluitvorming over beleid kan leiden. Merk op dat dus ook de verantwoordingsfunctie en de leerfunctie van evaluatie een plaats hebben in *ex-ante*-evaluatie.

Ex-durante-evaluaties worden uitgevoerd om zicht te krijgen op de implementatie van beleid of op mogelijke verbeteringen, wanneer er bijvoorbeeld aanwijzingen zijn dat de uitvoering hapert en/of de doelstellingen niet bereikt worden. Ook kunnen *ex-durante*-evaluaties nuttig zijn om een vroeg inzicht te krijgen in de effecten van beleid, inclusief eventuele positieve of negatieve neveneffecten. Vanuit dit perspectief zijn *ex-durante*-evaluaties eigenlijk een naar

¹ Zie hoofdstuk 1, Overheidsbeleid: wablief?

voren gehaalde *ex-post*-evaluatie (van der Knaap et al. 2020). Naar analogie met *ex-ante*-evaluatie, is het duidelijk dat *ex-durante*-evaluaties zowel van nut zijn in functie van verantwoorden als van leren. Vanzelfsprekend is de functie ‘beleidsplanning’ hier minder van toepassing.

Ex-post-evaluaties reflecteren, zoals gesteld, de traditionele visie op beleidsevaluatie en zijn waar beleidsevaluatie voor de meeste mensen voor staat. Strikt genomen vindt *ex-post*-evaluatie plaats na de uitvoering, al is het in de praktijk niet altijd duidelijk wanneer de implementatie precies is voltooid. *Ex-post*-evaluatie blikt in essentie terug in de tijd, maar levert ook lessen op voor toekomstig beleid: beslissen we om beleid te beëindigen, te veranderen, of in verbeterde vorm voort te zetten? Ook een *ex-post*-evaluatie kan dus zowel nuttig zijn in functie van beleidsleren als van verantwoorden. Onderstaande tabel zet een en ander op een rijtje.

Tabel 7.1: Types beleidsevaluaties gelinkt aan evaluatiemotieven

Types beleidsevaluatie, afhankelijk van fase in beleidsproces	Evaluatiemotieven		
	Beleidsplanning	Beleidsleren	Verantwoording
<i>Ex-ante</i> -evaluatie	X	X	X
<i>Ex-durante</i> - of tussentijdse evaluatie		X	X
<i>Ex-post</i> -evaluatie		X	X

Regelmatig wordt in wetten of decreten bepaald wanneer een evaluatie precies dient te worden uitgevoerd. Zo bepaalt Artikel 51 van het integratie- en inburgeringsdecreet (Vlaams Parlement 2013) bijvoorbeeld het volgende: ‘Uiterlijk drie jaar na de inwerkingtreding van dit decreet, en vervolgens om de drie jaar, evalueert de Vlaamse Regering de uitvoering van het Vlaamse integratiebeleid en legt ze de evaluatie voor aan het Vlaams Parlement.’ Dit voorbeeld maakt ook het moeilijke onderscheid tussen *ex-durante*- en *ex-post*-evaluaties duidelijk.

Een rijk landschap van (onafhankelijke?) actoren

Dat beleidsevaluaties in essentie onderzoeksstudies zijn, betekent niet dat enkel universitaire onderzoekers beleidsevaluaties uitvoeren, wel integendeel. Het landschap aan **evaluatoren** is zeer breed, en betreft eveneens consultants, beleidsmedewerkers binnen de administratie, auditeurs aan het Rekenhof², onderzoekers aan planbureaus, adviesraden etc.

Vanuit het perspectief van het evaluandum beschouwd, kunnen we ruwweg twee types evaluaties onderscheiden: **interne of externe evaluaties**. Interne evaluaties betreffen evaluaties waarvan de evaluator behoort tot de organisatie waarvan het beleidsinitiatief wordt geëvalueerd. Men spreekt van externe evaluaties als de evaluator niet behoort tot de organisatie waarvan het beleidsinitiatief wordt geëvalueerd (De Peuter et al. 2007). In de praktijk ziet men meer en meer combinaties van beide vormen, waarbij bijvoorbeeld bepaalde delen van de evaluatie worden uitbesteed (een externe evaluator wordt bijvoorbeeld gerekruteerd om een survey uit te sturen en te analyseren), en waarbij de rest van het proces intern wordt gehouden. Elke vorm heeft bepaalde voor- en nadelen, waarbij de voordelen van

² Het Rekenhof is een belangrijke actor in het Belgische evaluatielandschap. Sinds 1999 voert het Rekenhof ‘thematische audits’ uit. Terwijl deze audits in het verleden eerder focusten op de voorwaarden voor effectief en efficiënt beleid en op doelbereik, kunnen meer en meer van deze performantie-audits als beleidsevaluaties worden beschouwd, waarbij ook een substantief oordeel wordt geveld over de effecten en impact van beleidsinitiatieven (van der Knaap et al. 2020).

een interne evaluatie de nadelen zijn van externe evaluatie, en vice versa. Onderstaande tabel somt de belangrijkste voor- en nadelen op.

Tabel 7.2: Voor- en nadelen van interne en externe evaluatie

Voordelen interne evaluatie	Nadelen interne evaluatie
Gedetailleerde kennis over het evaluandum Toegang tot informatie Contextgevoeligheid Minder gepercipieerd als bedreigend	Rolcomplexiteit Mogelijke belangenvermenging
Voordelen externe evaluatie	Nadelen externe evaluatie
Sterke (gepercipieerde) onafhankelijkheid Specialisatiekennis op methodisch vlak Aanbreng van vernieuwende perspectieven Financiële transparantie	Moeilijkere toegang tot informatie Minder mogelijkheden tot interactie Disfuncties concurrerende markt

Bron: De Peuter et al. (2007)

Een interne evaluator heeft doorgaans als voordeel dat deze bij voorbaat meer detailkennis heeft van de logica achter een beleidsmaatregel, makkelijker toegang heeft tot informatie, en meer *inside* informatie heeft over de eventuele gevoeligheden die gepaard zijn gegaan met de keuze voor een bepaalde beleidsmaatregel (of de uitvoering ervan). Ook wordt een interne evaluator vaak als minder bedreigend gepercipieerd. Zeker in het geval van een evaluatie die wordt uitgevoerd vanuit een verantwoordingsobjectief, wordt een externe evaluator nogal eens als ‘controleur’ beschouwd. Daartegenover staat dat een interne evaluator kan worstelen met rolcomplexiteit. Men moet immers de nodige afstand kunnen nemen van de eigen organisatie of het beleid waarin men een rol heeft gespeeld. Dit kan ook wel eens leiden tot een perceptie van belangenvermenging (‘slager keurt eigen vlees’).

Het inschakelen van een externe evaluator heeft als troef dat deze bepaalde geavanceerde methodologische kennis of vernieuwende perspectieven kan inbrengen, waarover men binnen de organisatie niet zelf beschikt, of waar men niet aan denkt. Vanuit budgettair perspectief is het laten uitvoeren van een evaluatiestudie door een externe evaluator ook vaak transparanter, men weet immers vooraf precies wat de evaluatie zal kosten. De (personeels)kost is in geval van een interne evaluatie soms moeilijker in te schatten. Daarnaast heeft een externe evaluator het voordeel dat deze als relatief onafhankelijker wordt beschouwd, althans in de perceptie. **Onafhankelijkheid** is, zeker gezien de politieke context waarin evaluaties plaatsvinden, essentieel voor de geloofwaardigheid van een evaluatie, en het ontbreken hiervan kan de legitimiteit van de bevindingen en van de opdrachtgever zelf in het gedrang brengen.

In een evaluatiecontext wordt onafhankelijkheid typisch geassocieerd met onpartijdigheid en objectiviteit. **Onpartijdigheid** refereert aan een gebrek aan vooringenomenheid. Objectiviteit verwijst meestal naar een weergave van de werkelijkheid die niet wordt beïnvloed door persoonlijke voorkeuren of overtuigingen. Kortom, een evaluator moet een onbevangen kritisch perspectief kunnen brengen op het geëvalueerde beleid. Hoewel een externe evaluator het voordeel heeft van de perceptie van een grotere onafhankelijkheid, is er uiteraard geen garantie dat dit effectief het geval is. Externe evaluatoren begeven zich immers in een markt, en zijn ook afhankelijk van hun broodheren/opdrachtgevers voor eventuele toekomstige opdrachten. Dit maakt hen ook kwetsbaar voor eventuele beïnvloeding.

Volledig onafhankelijk onderzoek, zonder invloed van welke voorkeur dan ook, is uiteraard in zekere zin een ideaal en heel moeilijk haalbaar (Onderzoeksraad voor Veiligheid 2018).

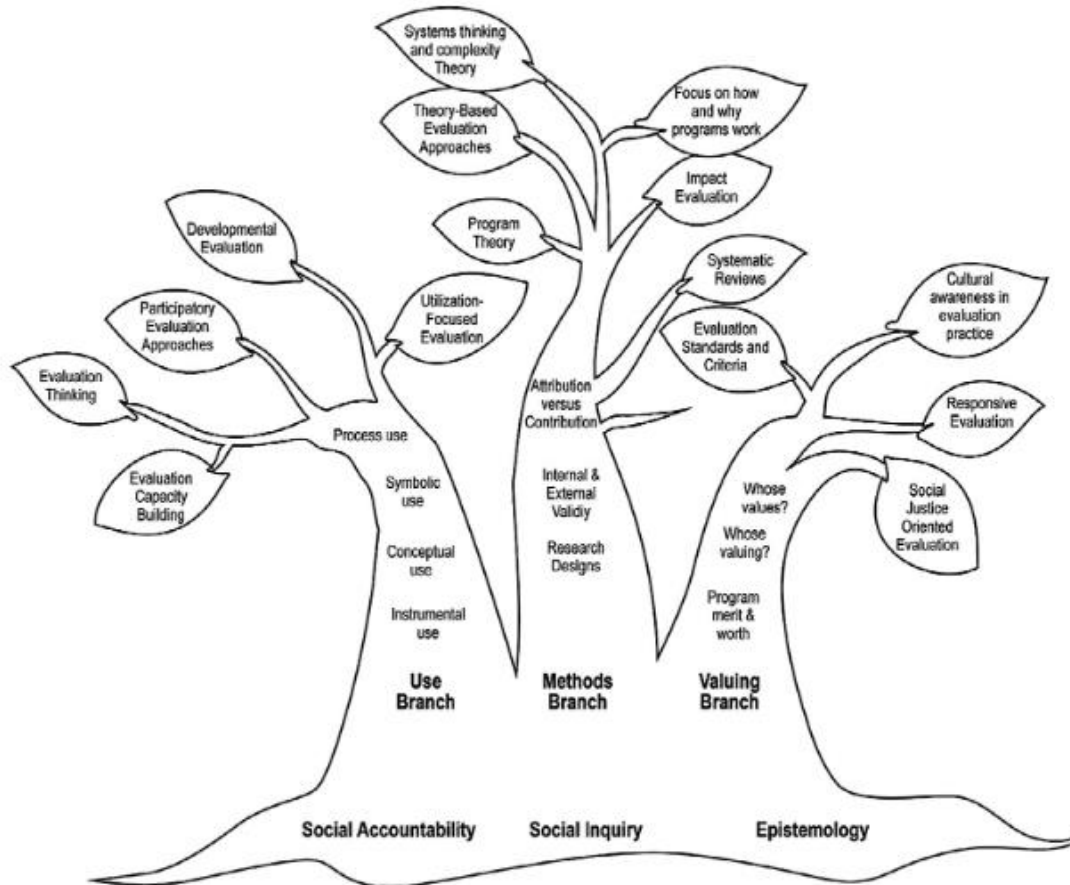
Cruciaal is echter dat evaluatoren zich bewust zijn van en expliciet zijn over eventuele afhankelijkheden, en deze in de mate van het mogelijke trachten te voorkomen (van der Knaap et al. 2020).

Onafhankelijkheid is dus een cruciaal **kwaliteitscriterium voor evaluatiestudies**. Daarnaast gelden voor evaluaties dezelfde kwaliteitscriteria als voor traditioneel wetenschappelijk onderzoek, zoals validiteit en betrouwbaarheid van de gebruikte methoden. Verder is ook de **bruikbaarheid** van het onderzoek voor beleidsmakers een essentieel kwaliteitscriterium. Zoals in de inleiding gesteld, is evaluatie een bron van kennis om te weten of het ‘goede’ werd gedaan in de aanpak van maatschappelijke problemen, en of dit goed werd gedaan. Het is dan ook belangrijk dat een studie relevant is in functie van het beantwoorden van deze sleutelvragen, zodat de evaluatie zelf ook een meerwaarde vormt voor beleidsmakers.

De evaluatietheorieboom: onderzoek *voor* evaluatie

De stand van zaken in theorievorming over beleidsevaluatie beschrijven is geen evidente oefening, aangezien het praktijkveld van beleidsevaluatie zeer heterogeen is. Beleidsevaluatie wordt uitgevoerd in een veelheid van beleidsdomeinen, door een veelheid van actoren, op een diversiteit van types beleid. Om enigszins vat te krijgen op het onderzoek over evaluatie, kunnen we ons beroepen op het welbekende classificatiesysteem van Christie en Alkin (2013) dat in evaluatiecontext gekend staat als *the evaluation theory tree*. De boom is meer recent door Lemire et al. (2020) aangepast in het licht van recente ontwikkelingen in beleidsevaluatie. Onderstaande figuur geeft deze meer actuele versie weer, en biedt ons een houvast om een aantal belangrijke stromingen in het evaluatieonderzoek te beschrijven.

Figuur 7.3: Evaluatietheorieboom



Bron: Lemire et al. (2020)

Vooraleer we de boom zelf bespreken, is het belangrijk even stil te staan bij de aard van de vermelde theorieën. Hoewel wordt gesproken van de *evaluation theory tree*, zijn de onderdelen eerder te beschouwen als frameworks of denk- en analysekaders, zoals we deze behandelden in hoofdstuk 2 (tabel 2.8). Ze bieden immers geen empirisch getoetste verklaringen over de evaluatiepraktijk, maar eerder prescriptieve richtlijnen die aangeven wat een goede of effectieve evaluatiestudie is. Ze geven met andere woorden (vaak breed gedeelde) opvattingen en overtuigingen weer zoals deze in het evaluatieveld circuleren (Fitzpatrick et al. 2009), en betreffen dus vooral onderzoek ‘voor’ evaluatie en evaluatoren.

Zoals de figuur weergeeft, is theorievorming in beleidsevaluatie ‘geworteld’ in een streven naar publieke verantwoording (aantonen dat maatschappelijke programma’s een positief effect of impact hebben), het voeren van **systematisch** sociaalwetenschappelijk onderzoek, en **epistemologie** (reflecties over de aard en validiteit van de geleverde kennis). Deze wortels hebben geleid tot specifieke ‘takken’ die tot op de dag van vandaag de theorie en praktijk van evaluatie richting geven. De takken geven belangrijke karakteristieken van beleidsevaluatie weer, die ook al eerder in dit hoofdstuk aan bod kwamen: beleidsevaluatie gaat over het geven van een waardeoordeel (*valuing*); gebeurt via een systematische methodiek (*methods*) en is gericht op het gebruik door beleidsmakers (*use*). Volgens Christie en Alkin (2013) dient elke evaluatietheorie zich over elk van deze aspecten uit te spreken, al legt elke theorie in de praktijk wel bepaalde klemtonen. We staan hieronder kort stil bij elke **evaluatietak**, en bespreken per tak een aantal belangrijke kenmerken van het gevoerde evaluatieonderzoek.

Methodentak

De centrale tak in de boom is de **methodentak**. Deze tak betreft theorievorming over onderzoeksmethodologie, waaronder onderzoeksdesigns en technieken voor het uitvoeren van evaluaties. Heel concreet bevat de methodentak vooral reflecties en benaderingen over hoe op de meest betrouwbare en valide manier kennis kan worden vergaard, gezien de vaak complexe beleidscontext waarin evaluaties plaatsvinden. Belangrijk om aan te stippen is dat beleidsevaluatieonderzoekers traditioneel sterk putten uit sociaalwetenschappelijke methoden. Ook economische methoden worden in een evaluatiecontext klassiek onder deze noemer geplaatst. Het werk van Donald Campbell (1957), en later van Campbell en Stanley (1966), is voor deze methodentak bijzonder invloedrijk geweest (Carden & Alkin 2012). Zij ontwikkelden de basis voor de **experimentele** en **quasi-experimentele designs** die vandaag de dag nog steeds door velen als ‘gouden standaard’ worden gezien om effectiviteit te meten, en die ook in de Belgische beleidscontext af en toe worden toegepast (zie voorbeeld in onderstaand kader).

Er zijn diverse soorten experimenten. Eenvoudig gesteld worden bij een ‘zuiver’ experiment (*randomized controlled trial* genoemd) twee willekeurig samengestelde groepen vergeleken, waarbij de ene onderhevig is aan een beleidsmaatregel (de beleidsgroep), en de andere niet (controlegroep), om het effect te bepalen. Naast deze zuivere experimentele vorm, kennen we nog een aantal afgeleiden, aangezien het niet altijd mogelijk is om random groepen samen te stellen, of om een controlegroep te vormen. In dergelijk quasi-experimenteel onderzoek worden dan bijvoorbeeld enkel metingen gedaan bij de beleidsgroep voor én na de invoering van een beleidsmaatregel.

In 2018 voerde het Directoraat-Generaal Personen met een Handicap in 2018 een experiment uit, met als doelstelling de tot dan hoge absenties bij geplande controles bij een evaluerend arts aan te pakken. De evaluerend arts bepaalt of iemand recht heeft op erkenning voor zijn handicap. Het experiment, gebaseerd op gedragsonderzoek, hield in dat het effect van verschillende communicatievarianten werd getest in Gent en Hasselt. Personen werden op een random manier verdeeld over verschillende groepen: men ontving ofwel de oproepingsbrief die tot dan toe werd gehanteerd, een nieuwe oproepingsbrief, een nieuwe oproepingsbrief en een brochure, of een nieuwe oproepingsbrief en twee herinnerings-sms'en. Elke keer iemand aanwezig was op de doktersafspraak of tijdig annuleerde, werd dit genoteerd als een positieve uitkomst. Onderstaande tabel (resultaat voor Gent) toont overduidelijk dat het sturen van de nieuwe oproepingsbrief in combinatie met sms de meeste vruchten afwerpt.

<i>Oost-Vlaanderen (centrum medische expertise Gent)</i>	<i>Controlegroep (oude brief)</i>	<i>Nieuwe brief</i>	<i>Nieuwe brief + brochure</i>	<i>Nieuwe brief + sms</i>
<i>Totale steekproefgrootte</i>	305	292	272	186
<i>Aandeel positieve uitkomsten</i>	86%	95%	96%	99%

Bron: Federale Overheidsdienst Sociale Zekerheid (2020)

Recenter wordt echter ook de meerwaarde van andere evaluatiemethoden onderkend, zeker wanneer het gaat over het onderzoeken van de omstandigheden waaronder beleid werkt, of de manier waarop belanghebbenden de meerwaarde van beleid ervaren. Denk hierbij aan kwalitatieve methoden als *Process Tracing* of *Qualitative Comparative Analysis* (QCA) (Pattyn et al. 2019). Dergelijke methoden focussen niet zozeer op de attributie of de netto-toegevoegde waarde van een beleidsinterventie op een bepaald effect (in welke mate heeft een informatiecampagne mensen gestimuleerd om inderdaad op te dagen tijdens een medische controle?), maar eerder op de contributie (of bijdrage) van beleid aan dit effect, naast andere

factoren die ook een rol kunnen spelen. Zeker in geval van complex (of *wicked*) beleid is het immers niet evident te bepalen in welke mate precies een beleidsinitiatief een verschil heeft teweeggebracht. Het bepalen van de bijdrage van beleid in een welbepaalde context is dan veel realistischer.³

Ook de beschikbaarheid van nieuwe massale gegevens (*big data*) door nieuwe databronnen, betere ontsluiting van bestaande ICT-toepassingen en de opkomst van sociale media, heeft de afgelopen jaren geleid tot veel meer nieuwe mogelijkheden voor evaluatieonderzoek (Lemire et al. 2020). Een goed voorbeeld hiervan is de ontsluiting van het Datawarehouse Arbeidsmarkt en Sociale Bescherming dat door de Kruispuntbank voor Sociale Zekerheid (KSZ) wordt beheerd. Dit Datawarehouse voorziet in een koppeling van sociaaleconomische gegevens afkomstig van verschillende openbare instellingen van sociale zekerheid, en biedt zo een massa aan data die als input kan worden gebruikt voor evaluatieonderzoek. Een onderzoek naar de kinderbijslag voor kinderen met een aandoening maakte hiervan bijvoorbeeld handig gebruik (Rekenhof 2018). Ook kunnen sociale media nu handig worden benut om op grote schaal vragenlijsten te verspreiden, denk bijvoorbeeld aan de veelheid aan bevestigingen over COVID-19 die onderzoeksinstellingen via sociale media lanceerden.

Waarderingstak

Aan de rechterkant van de methodentak bevindt zich de **waarderingstak**. Deze tak is zeer sterk geïnspireerd door het werk van Michael John Scriven, wiens definitie over evaluatie we eerder aanhaalden. Het was Scriven (1967) die sterk betoogde dat evaluatie geen evaluatie kan zijn zonder een normatief waardeoordeel. Wat betreft dit waardeoordeel, kunnen binnen het evaluatieveld twee strekkingen worden geïdentificeerd: een objectieve en een subjectieve strekking.

Vanuit het perspectief van de **objectieve strekking**, en ook deze van Scriven, is het de evaluator zelf die met een waardeoordeel dient te komen. Deze rol is niet voor andere actoren weggelegd. Vanuit een meer **subjectief perspectief** daarentegen wordt de realiteit gezien als een dynamisch proces, waarbij verschillende personen een verschillende opvatting kunnen hebben over de werking van een bepaalde beleidsinterventie. In deze visie is het niet opportuun dat de evaluator zelf een waardeoordeel velt, maar dient de subjectieve interpretatie van de evaluatie-informatie ook in rekening te worden gebracht (Carden & Alkin 2012). Deze laatste opvatting leunt ook aan bij wat we in hoofdstuk 2 beschreven: volgens sommigen kan het beleid niet objectief bestudeerd worden, omdat het altijd om waarden draait.

Dit wordt ook gereflecteerd in bepaalde zogenaamde **constructivistische stromingen** in het evaluatieonderzoek. Vanuit het oogpunt van constructivisten krijgen doelen, effecten en beleid pas betekenis in het perspectief van belanghebbenden. Elke persoon ervaart een eigen werkelijkheid, en het is aan evaluatoren om deze verschillende opvattingen expliciet te achterhalen. Van groot belang in dit verband is het (extreem) constructivistische werk van Guba en Lincoln (1989), die zich met hun werk onder de titel *Fourth Generation Evaluation* expliciet hebben afgezet tegen onder andere het werk van Scriven. Volgens hen schuilt in de meer objectieve benaderingen het gevaar dat alleen wat meetbaar is belangrijk wordt geacht, en doen objectieve stromingen onvoldoende recht aan de inzichten van belanghebbenden uit de praktijk. In een constructivistische visie dient evalueren dan ook in de eerste plaats een

³ Zie bijvoorbeeld van der Knaap et al. (2020) voor een gedetailleerd overzicht van evaluatiemethoden voor *ex-ante*-, *ex-durante*- en *ex-post*-evaluatieonderzoek, en de evolutie in trends inzake benaderingen op evaluatieonderzoek.

interactief proces te zijn, waarbij de evaluator actief werkt rond de informatiebehoefte en verwachtingen van de belanghebbenden. De bedoeling hierbij is dat belanghebbenden inzicht krijgen in elkaars percepties, opvattingen over de beleidsopties of beleidseffecten (zie ook van der Knaap et al. 2020 voor een meer gedetailleerde bespreking). We dienen wel te benadrukken dat de objectieve of meer rationeel-analytische visie nog steeds de meest courante stroming is in de evaluatiepraktijk, eerder dan de constructivistische benaderingen.

Gebruikstak

De linkerkant van de boom ten slotte wordt gevormd door de **gebruikstak**. Deze tak betreft verschillende benaderingen om na te denken over het gebruik van evaluatieonderzoek, en de factoren die hierin een belangrijke rol spelen. Het werk van Carol Weiss (1979) is hierin sterk bepalend geweest. Zij benadrukte de meervoudigheid van het concept 'gebruik', waarbij ze een onderscheid maakte tussen instrumenteel gebruik, conceptueel gebruik en symbolisch gebruik.

Instrumenteel gebruik is dat gebruik waarbij een evaluatie direct beslissingen over het beleid beïnvloedt. Op basis van de aanbevelingen van een evaluatie wordt dan bijvoorbeeld besloten om een beleidsprogramma stop te zetten of aan te passen. Zo leidde een evaluatie naar de communicatie van het vrijwillig terugkeerbeleid van mensen zonder papieren, op basis van de onderzoeks-aanbevelingen, tot aanpassing van de affiches en brochures waarmee asielzoekers worden ingelicht over de mogelijkheden van vrijwillige terugkeer. Instrumenteel gebruik kan betekenen dat er geleerd is van de evaluatie, maar dit is niet per se het geval. In sommige gevallen brengt een evaluatie nieuwe kennis, en vindt er daardoor gebruik plaats. In andere gevallen kan de kennis al bekend zijn, maar biedt de evaluatie wel mogelijkheden om een bepaalde keuze te maken of uit te voeren. Instrumenteel gebruik vindt meestal plaats direct volgend op de evaluatie, maar kan ook pas veel later gebeuren.

Carol Weiss (1979) benadrukte verder het belang van **conceptueel** gebruik of zogenaamd *enlightenment use*. Conceptueel gebruik vindt plaats als de evaluatie zorgt voor een verandering in denken, of voor nieuwe ideeën over bijvoorbeeld bepaalde concepten, de onderliggende aannames onder beleid, en de prioriteiten en doelen van het beleid. Denk bijvoorbeeld aan een grotere bekommernis over het milieu, als gevolg van een toenemend aantal milieu-impacttoetsen (*environmental impact assessments*). Conceptueel gebruik is minder makkelijk te identificeren dan instrumenteel gebruik. Het speelt zich immers af in de hoofden van beleidsmakers, en komt misschien pas na langere tijd expliciet tot uiting via bijvoorbeeld de referentie naar evaluatiestudies tijdens een parlementair debat.

Symbolisch gebruik raakt aan het strategisch-politiek gebruik dat we al eerder schetsten, en vindt plaats als de gebruiker de evaluatie aanwendt om een al bestaand standpunt kracht bij te zetten of te legitimeren. De gebruiker is hierbij niet geïnteresseerd in de evaluatie, behalve als legitimering. Zo kan men bijvoorbeeld verwijzen naar een evaluatiestudie om een bepaalde beleidsbeslissing te rechtvaardigen, maar zou men zonder studie ook dezelfde beslissing nemen.

Het verschil tussen instrumenteel en symbolisch gebruik is niet altijd makkelijk te bepalen. Bij instrumenteel gebruik is het ook mogelijk dat de beïnvloeding bestaat uit het geven van meer zekerheid over een beslissing of de argumentatie om een gewenste verandering door te voeren. Het verschil tussen symbolisch en instrumenteel gebruik ligt dus in de intentie van de beleidsmaker: bij symbolisch gebruik is er geen oprechte interesse in de bevindingen, alleen

in zoverre ze kunnen bijdragen aan legitimering, terwijl er bij instrumenteel gebruik wel een oprechte interesse is (Widmer & Neuenschwander 2004).

Het gebruik, in het bijzonder instrumenteel en conceptueel gebruik, kan voortkomen uit twee verschillende bronnen: het evaluatieproces en de evaluatiebevindingen (Kirkhart 2000; Taut 2007). Dit onderscheid is van belang omdat het erop wijst dat niet enkel het eindrapport van belang is bij gebruik, maar dat ook het proces van de evaluatie zelf van enorme meerwaarde is. Vooral de klemtoon op **procesgerelateerd gebruik** heeft recent veel meer aandacht gekregen in de evaluatieliteratuur. In het licht daarvan werden diverse **participatieve benaderingen** ontwikkeld, die gericht zijn op de betrokkenheid van stakeholders tijdens het evaluatieproces. In deze participatieve benaderingen worden de doelgroepen van beleid bij de evaluatie zelf betrokken, waarbij prioriteit wordt gegeven aan zwakkere en kwetsbare groeperingen in de samenleving (mensen in armoede, jeugddelinquenten, langdurig werklozen, mensen zonder papieren...). Dit participatieve element kan een of meerdere fasen van het evaluatieproces betreffen, zoals bijvoorbeeld de selectie van relevante evaluatievragen, de keuze van dataverzamelingmethoden, de dataverzameling zelf en de interpretatie van de bevindingen. Het idee is dat de participatie zelf kan bijdragen aan een positiever zelfbeeld van de deelnemers, wat een mooie bijvangst is van de evaluatie. Ook kunnen participatieve evaluaties helpen om bepaalde aspecten van de probleemdefinitie aan het licht te brengen die anders soms over het hoofd worden gezien. Participatieve evaluaties met personen in kansarmoede hielpen bijvoorbeeld de aandacht van beleidsmakers te vestigen op niet-monetaire aspecten van kansarmoede (Mestrum 2000).

Theorievorming over evaluatie: geen boom, maar zaadjes

Bovenstaande evaluatietheorieboom betreft, zoals gesteld, vooral prescriptieve evaluatietheorieën die voorschrijven hoe een evaluatie moet gebeuren, en die in eerste instantie ten dienste staan van evaluatoren zelf en beleidsmakers die evaluaties uitbesteden of uitvoeren (theorieën *voor* evaluatie). De vele takken en blaadjes aan de prescriptieve evaluatietheorieboom tonen een sterk ontwikkeld praktijklandschap. Hetzelfde kan evenwel niet worden gezegd over het onderzoek *over* beleidsevaluatie. Het praktijkgerichte karakter van beleidsevaluatie ging enigszins koste van systematische empirische theorievorming over deze beleidsondersteunende functie. Zoals Carden en Alkin (2012) stelden: *‘There is a paucity of research about evaluation. We simply do not know (in a scientific way) the results of particular evaluation actions taking place within particular contextual settings. Research does not provide those answers. Perhaps the complexity of program contexts and evaluator actions are so vast that we will never be able to create a descriptive evaluation theory.’*

Dat wil natuurlijk niet zeggen dat er geen theorievorming *over* evaluatie beschikbaar is. De afwezigheid van een empirische evaluatietheorieboom maakt het evenwel moeilijker hiervan een exhaustief beeld te schetsen. Als we de drie takken van de evaluatieboom bekijken, is vooral de ‘gebruikstak’ sterk voorwerp geweest van studies (Christie 2003). Men stelde immers vast dat vele evaluatiestudies niet leidden tot evaluatiegebruik. Carol Weiss zei hierover: *‘We are often disappointed. After all the " Sturm und Drang " of running an evaluation, and analyzing and reporting its results, we do not see much notice taken of it. Things usually seem to go along much as they would have gone if the evaluation had never been done. What is going wrong here? And what can we do about it?’* (Weiss 1987; zie ook van der Knaap et al. 2020)

De meer gedetailleerde ontrafeling van het concept ‘evaluatiegebruik’ heeft inmiddels een en ander in perspectief gezet, door aan te tonen dat de invloed van evaluatie ruimer is op te vatten dan louter instrumenteel gebruik. Anderzijds zijn in de evaluatieliteratuur inmiddels heel wat studies uitgevoerd over de factoren die het gebruik van evaluaties kunnen bevorderen of verhinderen (voorbeelden: Leviton & Hughes 1981; Shulha & Cousins 1997; Johnson et al. 2009). Het is hierbij duidelijk dat het al dan niet gebruiken van evaluaties in besluitvorming meestal het resultaat is van een complex samenspel van factoren, waarbij is gebleken dat vooral volgende elementen een rol spelen in het verklaren van gebruik:

1. de kenmerken van de evaluatie (de kwaliteit van het onderzoek, de haalbaarheid van de aanbevelingen, de timing van de studie);
2. de kenmerken van de evaluator (wordt deze als geloofwaardig gezien?);
3. de kenmerken van de potentiële gebruiker (heeft deze interesse in de evaluatie?);
4. de kenmerken van de organisatie die de evaluatie uitvoert of laat uitbesteden (wat is de attitude van het management ten aanzien van de evaluatie?);
5. de politieke context van een evaluatie (bijvoorbeeld de politieke gevoeligheid van het evaluandum).

Gelinkt aan het onderzoek naar het gebruik van evaluatie, worden meer en meer studies ondernomen die beleidsevaluatie beschouwen vanuit een bestuurskundige invalshoek. Dergelijke studies focussen zich onder meer op verschillen over de **institutionalisering** van evaluatie in organisaties (bijvoorbeeld Jacob et al. 2015; Stockmann et al. 2020), de evaluatiecapaciteit van organisaties, de evaluatiematuriteit van verschillende sectoren en landen (bijvoorbeeld Furubo et al. 2002), en verschillende evaluatiestijlen (bijvoorbeeld De Francesco & Pattyn 2020). In vergelijking met het extensieve onderzoek dat is gebeurd *voor* evaluatie en beleid, is er dus nog wel een weg af te leggen wat het onderzoek *over* evaluatie betreft. We spreken in dit geval dan ook beter van zaadjes voor een verdere onderzoeksagenda, eerder dan van een robuust ontwikkelde theorieboom.

Tot slot

Beleidsevaluatie is erop gericht de meerwaarde van het evaluandum te beoordelen, wat essentieel is voor goed bestuur. Afhankelijk van het type beleidsevaluatie (*ex ante*, *ex durante*, *ex post*) staan specifieke functies meer of minder centraal: beleidsplanning, verantwoording of beleidsleren (waarom?). Beleidsevaluatie is in essentie een vorm van toegepast onderzoek (wat?) waarbij de evaluator al dan niet kan behoren tot de organisatie waarvan het beleidsinitiatief wordt geëvalueerd (wie?). Ongeacht de manier van evalueren, is onafhankelijkheid van onderzoek een essentieel kwaliteitscriterium. Over theorievorming zijn vooral prescriptieve kaders sterk ontwikkeld voor de evaluatiepraktijk, eerder dan empirisch getoetste theorieën over beleidsevaluatie (hoe?). We onderscheiden hierin drie hoofdlijnen van onderzoek: methoden van evaluatie, het waarderen van het evaluandum en het gebruik van evaluatie.

Test jezelf

1. In de Vlaamse Beleidsnota Wonen 2019-2024 staat volgende passage over de evaluatie van de verhuur van leegstaande sociale huurwoningen in afwachting van renovatie:

‘Bij grote complexen duurt het vaak een tijd vooraleer al deze verhuisbewegingen zijn afgerond. In tussentijd staan de woningen waarvan de bewoners wel al zijn verhuisd leeg. Om deze leegstaande sociale huurwoningen in tussentijd nuttig te gebruiken, bevat het sociaal huurstelsel de mogelijkheid om deze woningen in afwachting van de renovatie tijdelijk te verhuren buiten het sociaal huurstelsel. Ik evalueer dit regime en ga na of het moet aangepast worden opdat het beter kan ingezet worden om de leegstand in afwachting van renovatie te verminderen.’ (Vlaamse Regering 2019b)

Wat zijn de functies waarvoor beleidsevaluatie kunnen worden gebruikt, en welke van deze functies herken je in deze passage?

2. De Vlaamse Beleidsnota 2019-2024 Welzijn, Volksgezondheid, Gezin en Armoedebestrijding bevat volgende passage:

‘We evalueren de lopende projecten rond vernieuwende aanbodsvormen in Initiatieven voor Beschut Wonen en Psychiatrische Verzorgingstehuizen, met name specifieke woonvormen voor jongvolwassenen met psychische problemen en aanklappende zorg voor zorgmijdende mensen met psychische problemen die sociaal huren. Bij een positieve evaluatie gaan we na of we extra ruimte kunnen creëren voor deze vernieuwende zorg. We onderzoeken in welke mate we de erkenningsnormen moeten aanpassen om tussenvormen mogelijk te maken van begeleiding, afgestemd op de zelfstandigheid van de zorggebruikers.’ (Vlaamse Regering 2019c)

Welke vorm van leren tracht men met deze evaluatie te realiseren: bijsturend of vernieuwend leren? Leg beide concepten eerst uit, om er daarna je antwoord op te baseren.

3. De Vlaamse ministers van Binnenlands Bestuur en van Onderwijs lanceerden samen een oproep tot organisatie van zomerscholen in het lager en secundair onderwijs in juli en augustus 2020 (Departement Onderwijs en Vorming 2020). Sommige leerlingen hebben door het opschorten van de lessen op school als gevolg van de COVID-19-pandemie een leerachterstand opgelopen. Bij anderen is de bestaande kloof vergroot, en nog andere leerlingen werden in deze periode onvoldoende bereikt. Met deze zomerscholen wil de Vlaamse Regering maatwerk op individueel niveau aanbieden aan kleine groepen kinderen en jongeren, zodat zij (en hun leerkrachten en ouders) sterker kunnen starten in het volgende schooljaar.

Vermits dit de eerste keer is dat de Vlaamse overheid dergelijke zomerscholen organiseert op grote schaal, laat ze deze evalueren. De projectoproep vermeldt onder meer volgende ambitie: ‘zicht krijgen op de mate dat de geplande zomerscholen effectief gerealiseerd werden, wat wel of niet is gerealiseerd, wat de oorzaken van niet-realiseren waren en welke drempels of succesfactoren hier een rol in speelden’ (Departement Onderwijs en Vorming, 2020).

Welk evaluatiecriterium staat centraal in deze evaluatie? En kan je zelf nog een evaluatievraag formuleren, gelinkt aan een ander evaluatiecriterium dat relevant zou zijn om te onderzoeken in het kader van de zomerscholen?

4. Begin 2019 kwam in Nederland het laatste van drie onderzoeksrapporten uit naar aanleiding van de zogeheten WODC-affaire. Deze ontstond na een uitzending van het televisieprogramma Nieuwsuur in december 2017, waarin de onafhankelijkheid van het WODC in twijfel werd getrokken. Het WODC verricht wetenschappelijk onderzoek naar het beleid van het Nederlands ministerie van Justitie en Veiligheid (J&V). Daarnaast adviseert het over dat beleid en de beleidsprogramma's. Minister Grapperhaus van Justitie en Veiligheid verbond aan de rapporten de conclusie dat de balans tussen nabijheid en onafhankelijkheid verstoord was. Het WODC zou op grotere afstand van het beleidsdepartement worden geplaatst, onder meer door een fysieke verhuizing uit het gebouw van het ministerie, en door aanscherping van regels en protocollen (Scheepmaker 2019).

Wat is onafhankelijkheid precies in een evaluatiecontext, en waarom is dit een belangrijk kwaliteitscriterium voor beleidsevaluaties?

5. 'Om het doel te bereiken dat er een maatschappelijk geaccepteerde norm zou ontstaan dat huiselijk geweld onacceptabel is, meer dan op dat moment het geval was, zijn in Nederland herhaaldelijk publiekscampagnes ingezet. Vanaf 2007 is jaarlijks een landelijke publiekscampagne gepresenteerd in verschillende landelijke media. Die campagne is erop gericht om bij het brede publiek (en in één geval ook specifiek voor daders) de norm aan te geven ten aanzien van huiselijk geweld en de ernst van huiselijk geweld te benadrukken. Daarnaast wordt aangegeven waar men voor hulp terecht kan. In 2005 en 2006 was al aandacht gegeven aan ouderenmishandeling in de vorm van informatiecampagnes van de landelijke overheid aan gemeenten. Sinds 2008 bestaat het Platform Ouderenmishandeling.' (Jongebreur et al. 2011)

Bovenstaande passage geeft een beleidstheorie weer. Wat wordt bedoeld met een beleidstheorie? Kan je op basis hiervan de output en beoogde effecten van het betreffende beleid identificeren?

Referentielijst

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54, 297-312.

Campbell, D. T., & Stanley J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago. IL: Rand McNally.

Carden, F., Alkin, M.C. (2012). Evaluation roots. An international perspective. *Journal of Multidisciplinary Evaluation* 8(17), 102-118.

Christie, C.A. (2003). What Guides Evaluation? A Study of How Evaluation Practice Maps onto Evaluation Theory. *New Directions for Evaluation*, 97, Spring 2003

Christie, C. A., & M.C. Alkin (2013). "An Evaluation Theory Tree." In M.C. Alkin (Ed.). *Evaluation Roots: Tracing Theorists' Views and Influences*, 2nd ed. Thousand Oaks, California: Sage.

- De Francesco, F. en Pattyn, V. (2020- te verschijnen). Evaluation styles. In: Howlett, M. & Tosun, J. (Eds.) Handbook of policy styles. Routledge.
- De Peuter, B., Desmedt, J., Bouckaert, G. (2007). Handleiding Beleidsevaluatie. Deel 1: Evaluatiedesign en -Management. Steunpunt Bestuurlijke Organisatie Vlaanderen.
- Departement Onderwijs en Vorming (2020). *Oproepen voor onderzoeksvorstellen: Zomerscholen*. Geraadpleegd 16 augustus 2020.
<https://onderwijs.vlaanderen.be/nl/oproepen-voor-onderzoeksvorstellen>.
- Federale Overheidsdienst Sociale Zekerheid 2020. Gemiste doktersafspraken niet bestraffen maar voorkomen - nudgingexperiment bij de DG Personen met een handicap. <https://socialsecurity.belgium.be/nl/nieuws/gemiste-doktersafspraken-niet-bestrafen-maar-voorkomen-nudgingexperiment-bij-de-dg-personen> (geraadpleegd 17 augustus 2020).
- Fitzpatrick, J., Christie, C., & Mark, M. M. (2009). Evaluation in action: Interviews with expert evaluators. California: Sage.
- Furubo, J. E., Rist, R. C., & Sandahl, R. (2002). *International atlas of evaluation*. Transaction Publishers.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Sage.
- 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., L. O. Greenesid, S. A. Toal, J. A. King, F. Lawrenz, en B. Volkov (2009) Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation* 30 (3) 377-410.
- Kirkhart, K.E. (2000) Reconceptualization evaluation use: An integrated theory of influence. *New Directions for Evaluation*. 88, 5-22.
- Knaap van der P., Pattyn V., Hanemaayer D. (2020). *Beleidsevaluatie in theorie en praktijk*. Den Haag: Boom.
- Lemire, S., Peck, L. R., & Porowski, A. (2020). The Growth of the Evaluation Tree in the Policy Analysis Forest: Recent Developments in Evaluation. *Policy Studies Journal*, 48, S47-S70.
- Leviton, L.C. en E.F.X. Hughes (1981) Research on the utilization of evaluations: A review and synthesis. *Evaluation Review* 5 (4) 525-548.
- Menstrum, F. (2000). Vrouwen en armoede. Wat zeggen de Verenigde Naties? *Samenleving en Politiek* 7(2), 21-29.
- Nederlands Ministerie van Financiën (2003). Handreiking evaluatieonderzoek ex post—Een praktisch handvat voor de opzet en uitvoering van evaluatieonderzoek ex post. Den Haag: Ministerie van Financiën.
- Onderzoeksraad voor Veiligheid (2018). Onafhankelijk onderzoek in het publiek belang. Den Haag: Onderzoeksraad voor Veiligheid.

- Pattyn, V., Molenveld, A., en Befani, B. (2019). Qualitative comparative analysis as an evaluation tool: Lessons from an application in development cooperation. *American Journal of Evaluation*, 40(1), 55-74.
- Rekenhof (2018). *Beleidsevaluatiecapaciteit van de federale overheidsdiensten. Verslag van het Rekenhof aan de Kamer van Volksvertegenwoordigers*. Brussel: Rekenhof.
- Scheepmaker, M. (2019). Inleiding. *Justitiële Verkenningen*. 45(2), 508.
- Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum evaluation (American Educational Research Association Monograph Series on Evaluation No. 1)*. Chicago, IL: Rand McNally.
- Scriven, M. (1991). *Evaluation Thesaurus*. Newbury Parck: Sage Publications.
- Shulha, L.M. and J. B. Cousins (1997) Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice* 18 (3) 195-208.
- Sociaal-Economische Raad van Vlaanderen (2015). *Tien denksporen voor ex post decreetsevaluatie in en door het Vlaams Parlement. Advies SERV*. Brussel: SERV.
- Stockmann, R., W. Meyer, en L. Taube. *The Institutionalisation of Evaluation in Europe*. London: Palgrave Macmillan.
- Taut, S. (2007) Methodological and conceptual challenges in studying evaluation process use. *Canadian Journal of Program Evaluation* 22 (2) 1-19.
- Varone, F., Jacob, S., & De Winter, L. (2005). Polity, politics and policy evaluation in Belgium. *Evaluation*, 11(3), 253-273.
- Vedung, E. (1997). *Public policy and program evaluation*. Abingdon: Routledge.
- Vlaams Parlement (2013). *Decreet van 7 juni 2013 betreffende het Vlaamse integratie- en inburgeringsbeleid*. *Belgisch Staatsblad* 26/7/2013. Brussel: Vlaams Parlement.
- Vlaams Parlement (2020). *Dossier corona*. <https://www.vlaamsparlement.eu/dossiers/corona> (geraadpleegd 12 augustus 2020).
- Vlaamse Regering (2009a). *Beleidsnota 2009-2014 Energie*. Brussel: Vlaamse Regering.
- Vlaamse Regering (2009b). *Beleidsnota 2009-2014 Mobiliteit en Openbare Werken*. Brussel: Vlaamse Regering.
- Vlaamse Regering (2019a). *Beleidsnota 2019-2024 Economie, Wetenschapsbeleid en Innovatie*. Brussel: Vlaamse Regering.
- Vlaamse Regering (2019b). *Beleidsnota 2019-2024 Wonen*. Brussel: Vlaamse Regering.
- Vlaamse Regering (2019c). *Beleidsnota 2019-2024 Welzijn, Volksgezondheid, Gezin en Armoedebestrijding*. Brussel: Vlaamse Regering.

Weiss, C. H. (1979). The many meanings of research utilization. *Public administration review*, 39(5), 426-431.

Weiss, C.H. (1987). Keynote speech. Uitgesproken tijdens de conferentie van de American Evaluation Association, 16 oktober 1987.

Widmer, T. en Neuenschwander P. (2004) *Embedding evaluation in the Swiss Federal Administration; Purpose, institutional design and utilization*. *Evaluation* 10 (4) 388-409.