



Spoor Beleid en Monitoring

HANDLEIDING BELEIDSEVALUATIE

DEEL 1: EVALUATIEDESIGN EN -MANAGEMENT

Bart DE PEUTER

Joris DE SMEDT

Prof. Dr. Geert BOUCKAERT

**Rapport
D/2006/10106/013**

2007

INHOUDSOPGAVE

1. Inleiding.....	11
1.1 Aanleiding en opzet	11
1.2 Doelpubliek	12
1.3 Opbouw van de handleiding	12
2. Beleidsevaluatie en monitoring in een overheidscontext	15
2.1 Evaluatie bij de overheid: ontstaan, ontwikkeling en trends.....	15
2.1.1 <i>Ontstaan</i>	15
2.1.2 <i>Beleidsevaluatie als praktijkveld: de mijlpalen</i>	16
2.1.3 <i>Beleidsevaluatie als theorieveld: stromingen, kenmerken en tendensen</i>	23
2.1.4 <i>Recente trends in evaluatie</i>	27
2.2 Definities van beleidsevaluatie.....	29
2.3 Motieven om beleid te evalueren	31
2.4 Positionering van evaluatie in de beleidscyclus en typologieën van beleidsevaluatie	35
2.4.1 <i>Evaluatietypes vanuit de beleidscyclus</i>	36
2.4.2 <i>Evaluatietypes op basis van de inhoudelijke focus</i>	37
2.4.3 <i>Evaluatietypes op basis van het evaluatiecriterium</i>	38
2.4.4 <i>Interne en externe evaluatie</i>	39
2.4.5 <i>Besluit in verband met evaluatietypologieën</i>	43
2.5 Het evaluatieproces	43
2.6 Bijzondere vraagstukken.....	45
2.6.1 <i>Capaciteitsopbouw</i>	45

2.6.2 Evaluatie van domeinoverschrijdend beleid	49
2.6.3 Evaluatie van interbestuurlijk beleid	51
Literatuur.....	53
3. Evaluatie in België en Vlaanderen.....	55
3.1 Impulsen	55
3.2 Enkele trends en uitdagingen	56
Literatuur.....	59
4. Ontwerpen en plannen van een evaluatie	61
4.0 Leeswijzer.....	61
4.1 Het bepalen van de scope van een evaluatie.....	61
4.1.1 Definiëren van het evaluandum.....	62
4.1.2 Het antwoord op de vraag: “Wat willen we weten over het evaluandum?”	64
4.1.3 identificeren van stakeholders	64
4.1.4 Opstellen en selecteren van de evaluatievragen en –criteria.....	66
4.2 Het plannen van de uitvoering van de evaluatie.....	73
4.2.1 Inleiding: enkele begrippen en kernprincipes	74
4.2.2 Planning van de gegevensverzameling.....	76
4.2.3 Het instrumentarium voor analyse en beoordeling.....	80
4.3 Uitbesteding van een evaluatie.....	80
4.3.1 Opmaak van een bestek voor een evaluatieopdracht.....	81
4.3.2 Selectie van de uitvoerder van de evaluatieopdracht.....	85
Literatuur.....	87

5. Evaluatietypes	89
5.1 Typologieën en dwarsverbanden	89
5.2 De rol van beleidstheorie voor evaluatie	93
5.3 Naar een conceptueel raamwerk van beleidstheorie	95
5.3.1 <i>Het veranderingsmodel</i>	95
5.3.2 <i>Het implementatiemodel</i>	99
5.3.3 <i>Het conceptueel raamwerk</i>	101
5.4 Ex ante evaluatie	102
5.4.1 <i>Inleiding</i>	102
5.4.2 <i>Relevantie van het beleid</i>	103
5.4.2.1 Opzet en vraagstelling	103
5.4.2.2 Meerwaarde.....	104
5.4.2.3 Aanpak en technieken	106
5.4.2.4 Voorbeelden	107
5.4.2.5 Kritieke slaagfactoren	108
5.4.3 <i>Interne logica van het beleid of de beleidstheorie</i>	108
5.4.3.1 Opzet en vraagstelling	108
5.4.3.2 Meerwaarde.....	109
5.4.3.3 Aanpak en technieken	109
5.4.3.4 Kritieke slaagfactoren	112
5.4.4 <i>Coherentie en consistentie van beleid</i>	114
5.4.4.1 Opzet en vraagstelling	114
5.4.4.2 Meerwaarde.....	116

5.4.4.3 Aanpak en technieken	116
5.4.4.4 Voorbeeld	119
5.4.4.5 Kritieke slaagfactoren	120
<i>5.4.5 Implementatie van het beleid.....</i>	<i>121</i>
5.4.5.1 Opzet en vraagstelling.....	121
5.4.5.2 Meerwaarde	121
5.4.5.3 Aanpak en technieken.....	122
5.4.5.4 Voorbeeld	123
5.4.5.5 Kritieke slaagfactoren	124
<i>5.4.6 Verwachte effecten en impact</i>	<i>124</i>
5.4.6.1 Opzet en vraagstelling.....	124
5.4.6.2 Meerwaarde	125
5.4.6.3 Aanpak en technieken.....	125
5.4.6.4 Voorbeeld: multimodaal verkeersmodel	126
5.4.6.5 Kritieke slaagfactoren	127
<i>5.4.7 Kosten en baten en kosteneffectiviteit.....</i>	<i>128</i>
5.4.7.1 Opzet en vraagstelling.....	128
5.4.7.2 Meerwaarde	129
5.4.7.3 Aanpak en technieken.....	129
5.4.7.4 Kritieke slaagfactoren	130
<i>5.4.8 Afwegen van beleidsalternatieven.....</i>	<i>131</i>
5.4.8.1 Opzet en vraagstelling.....	131
5.4.8.2 Meerwaarde	133
5.4.8.3 Aanpak en technieken.....	133

5.4.8.4 Kritieke slaagfactoren	135
5.5 Ex post evaluatie.....	137
5.5.1 <i>Inleiding</i>	137
5.5.2 <i>Doelbereiking</i>	138
5.5.2.1 Opzet en vraagstellingen	138
5.5.2.2 Meerwaarde.....	139
5.5.2.3 Strategieën en technieken	142
5.5.3 <i>Beleidseffectiviteit: inleiding</i>	145
5.5.3.1 Opzet en vraagstellingen	145
5.5.3.2 Meerwaarde.....	149
5.5.3.3 Strategieën en technieken	149
5.5.4 <i>Experimenteel design</i>	150
5.5.4.1 Evaluatiekader: “policy on” vs. “policy off”	150
5.5.4.2 Vergelijkingsbasis	151
5.5.4.3 Meten en vergelijken.....	152
5.5.4.4 Randvoorwaarden	155
5.5.4.5 Beperkingen op het hanteren van het experimenteel design.....	155
5.5.5 <i>Quasi-experimenteel design</i>	158
5.5.5.1 Evaluatiekader: “policy on” vs. “policy off”	158
5.5.5.2 Het bias probleem.....	158
5.5.5.4 Randvoorwaarden en beperkingen.....	168
5.5.6 <i>Alternatief of kwalitatief design</i>	169
Literatuur.....	169
6. Evaluatiemanagement en –capaciteit.....	171

6.1 Verantwoordelijkheden en budget	171
6.2 De status van de evaluator	172
6.3 Opvolging van de evaluatie	175
6.3.1 <i>Communicatie tussen de evaluator en de stakeholders</i>	175
6.3.2 <i>Stuurgroepformule</i>	176
6.4 Rapportering en feedback	178
6.4.1 <i>Algemene aandachtspunten</i>	179
6.4.2 <i>Resultaten, conclusies en aanbevelingen</i>	180
6.4.2.1 Resultaten	181
6.4.2.2 Conclusies	181
6.4.2.3 Aanbevelingen.....	182
6.4.3 <i>Het evaluatierapport</i>	185
6.5 Kwaliteitsbewaking en meta-evaluatie	188
Literatuur.....	197
7 Ontwikkeling en professionalisering van evaluatiecapaciteit	199
7.1 Ontwikkeling van institutionele capaciteit	199
7.2 Groeimodel voor de inbedding van een evaluatiecapaciteit op maat	200
Bronnen	209

Lijst van figuren en tabellen

Figuur 1: De 'evaluation theory tree'	24
Figuur 2: Het traditionele schema van de beleidscyclus	35
Figuur 3: Evaluatietypes gesitueerd in de beleidscyclus	36
Figuur 4: Evaluatiecriteria afgeleid uit de beleids- en beheerscyclus als open systeemmodel	38
Figuur 5: Het evaluatiemotief en de keuze tussen interne of externe evaluatie	40
Figuur 6: Sterktes en zwaktes van interne evaluatie	42
Figuur 7: Sterktes en zwaktes van externe evaluatie	42
Figuur 8: Logische sequens in een evaluatieproces	62
Figuur 9: Evaluatiecriteria voor ex ante, tussentijdse en ex post evaluatie	72
Figuur 10: Bronnen voor gegevensverzameling	77
Figuur 11: Kennisgraad en objectiviteit bij interne en externe evaluatie	81
Figuur 12: Evaluatie gepositioneerd in de beleidscyclus	90
Figuur 13: Evaluatiecriteria doorheen de beleids- en beheerscyclus	91
Figuur 14: Sequens voor de selectie van een evaluatietype	91
Figuur 15: Stroomdiagram voor het afleiden van een evaluatietype	92
Figuur 16: Het veranderingsmodel	96
Figuur 17: Eenvoudig veranderingsmodel achter antidrugsbeleid	97
Figuur 18: Uitgebreid veranderingsmodel achter antidrugsbeleid	98
Figuur 19: Veranderingsmodel achter beleid in de sector werkgelegenheid	98
Figuur 20: Het implementatiemodel	99
Figuur 21: Conceptueel raamwerk van beleidstheorie	101
Figuur 22: Evaluatie van de beleidsrelevantie	104

Figuur 23: Constructie van het veranderingsmodel via terugredeneren	110
Figuur 24: Constructie van het veranderingsmodel via voorwaarts redeneren.....	111
Figuur 25: Complexiteit van beleid	114
Figuur 26: Coherentie en consistentie van beleid als evaluatiecriteria	116
Figuur 27: Generiek veranderingsmodel bij complex beleid	117
Figuur 28: Coherentie bij complex beleid	118
Figuur 29: Boomstructuur van doelstellingen uit het verkeersveiligheidsbeleid.....	120
Figuur 30: Logisch diagram van verwachte impact: het regionaal programma voor Hengouwen voor Europese steun (1994-1999).....	120
Figuur 31: Kosteneffectiviteit als evaluatiecriterium	129
Figuur 32: Strategische keuze: op welke determinanten richt het beleid zich?	132
Figuur 33: Operationele keuze: welke variant van de beleidsoutput?.....	132
Figuur 34: Doelbereiking als evaluatiecriterium	138
Figuur 35: Beleidseffectiviteit als evaluatiecriterium.....	146
Figuur 36: Het veranderingsmodel	146
Figuur 37: Effectniveaus, effectverandering en beleidseffect.....	147
Figuur 38: Effecttypes op basis van tijdstip, duur en intensiteit.....	154
Figuur 39: Types van validiteit doorheen het evaluatieproces	159
Figuur 40: Individuele of geaggregeerde matching	162
Figuur 41: Factoren die de kwaliteit van een evaluatie beïnvloeden	189
Tabel 1: Overzicht van evaluatienetwerken in de VS, Canada en Europa	22
Tabel 2: Types van evaluatie gebaseerd op fase in de beleidscyclus	36

Tabel 3: Het verband tussen beleidsfase, evaluatietype en evaluatiemotief	37
Tabel 4: Clustering van evaluatietypes en motieven.....	37
Tabel 5: Checklist voor kwaliteitsvolle evaluatievragen	69
Tabel 6: Bestanddelen van een evaluatiebestek	82
Tabel 7: Checklist voor de beoordeling van evaluatieoffertes.....	86
Tabel 8: Detailniveaus van het multimodaal verkeersmodel.....	127
Tabel 9: Soorten effecten	130
Tabel 10: Koppeling van evaluatiemotieven en -types.....	137
Tabel 11: Evaluatiemotieven en evaluatie op doelbereiking.....	140
Tabel 12: Effectmetingen in een experimenteel design	153
Tabel 13: Comparatieve tijdreeksen voor een beleids- en vergelijkingsgroep.....	168
Tabel 14: Tijdreeksen opbouwen tijdens de uitvoering van het beleid	168
Tabel 15: Comparatieve tijdreeksen bij complex beleid.....	168
Tabel 16: Bestanddelen van een evaluatierapport.....	186
Tabel 17: Thema's doorheen de evaluatiestandaarden van de Zwitserse evaluatieassociatie.....	190
Tabel 18: Voorbeeld van checklist voor een meta-evaluatie.....	196
Tabel 19: Koppeling van evaluatiemotieven en evaluatietypologieën	202

1. Inleiding

1.1 Aanleiding en opzet

Wereldwijd groeit de aandacht voor beleidsevaluatie in de publieke sector. De vraag of en in welke mate de overheid een verschil maakt ten behoeve van de samenleving, is dan ook een zeer pertinente vraagstelling. Beleidsevaluatie levert een unieke bijdrage tot het antwoord.

Ook bij de overheid in België en Vlaanderen stellen we een stijgende interesse voor en vraag naar beleidsevaluatie vast. Slechts een aantal organisaties en sectoren heeft al een uitgebreide ervaring opgedaan met het evalueren van beleid. Voor een groot aantal actoren is beleidsevaluatie een tamelijk nieuw gegeven dat meer en meer aandacht krijgt. De toenemende vraag naar evaluatie ontstaat uit verschillende motieven: het duidelijk kunnen aantonen van de meerwaarde van beleidskeuzes en beslissingen, het verbeteren van de beleidsvoering, het kunnen aantonen en verklaren van bepaalde beleidseffecten, en leren uit bestaand beleid. Daarbij spelen zowel mechanismen van interne als externe druk op de overheid, als tendensen van organisatorische uitwaaiing en specialisatie en beleidsmatige coördinatie een belangrijke rol als katalysator. De versterkte aandacht voor beleidsevaluatie kadert in een algemeen streven naar een meer 'evidence-based policy', ofwel een geïnformeerd en onderbouwd beleid. De overheid en beleidsvoering zijn meer en meer kennisintensief geworden.

Tegenover de toenemende vraag naar beleidsevaluatie dient een adequaat aanbod te staan. Een belangrijk aspect daarvan is de opbouw van een slagvaardige evaluatiecapaciteit binnen de overheid. Een degelijke interne evaluatiecapaciteit is noodzakelijk om beleidsevaluaties voor te bereiden, om ze hetzij zelf uit te voeren hetzij om ze op te volgen indien ze worden uitbesteed, en om ze te gebruiken ter ondersteuning en onderbouwing van beleidsbeslissingen. Wanneer we spreken over evaluatiecapaciteit, kunnen we verschillende dimensies onderscheiden waaronder: financiële middelen, personeel, beleidskennis, en expertise over het evaluatieproces en evaluatietechnieken.

De toenemende aandacht voor beleidsevaluatie in combinatie met de vaststelling dat het aanbod nog te beperkt is, vormde de aanleiding voor het schrijven van deze handleiding. De handleiding vormt de output van een tweejarig onderzoeksproject binnen het Steunpunt Bestuurlijke Organisatie Vlaanderen, uitgevoerd door het Instituut voor de Overheid (K.U.Leuven).

De opzet van deze handleiding is een bijdrage te leveren tot de versterking van de evaluatiecapaciteit in de overheid en meer bepaald de laatstgenoemde dimensie ervan, de expertise inzake het evaluatieproces en -instrumentarium.

1.2 Doelpubliek

Deze handleiding richt tot iedereen die betrokken is bij het evalueren van beleid, hetzij aan de vraagzijde, hetzij aan de aanbodzijde. In de eerste plaats tot ambtenaren en hun overheidsadministraties op het lokale, provinciale, regionale of federale bestuursniveau, maar ook maar ook politici, academische onderzoeksgroepen, middenveldorganisaties, denktanks, en advies- en onderzoeksbureaus uit de private sector. De handleiding is generiek opgevat en niet beperkt tot een specifieke beleidssector.

1.3 Opbouw van de handleiding

De complete handleidingenset bestaat uit vijf complementaire volumes die de lezer een inzicht verschaffen in het evaluatieproces en –instrumentarium. Elk deel is zo opgevat dat het zelfstandig leesbaar is.

Deel 1 gaat in hoofdstuk 2 eerst in op de kenmerken van beleidsevaluatie in een overheidscontext. Hoe is beleidsevaluatie geëvolueerd als praktijk- en theorieveld? Welke recente trends manifesteren zich? We definiëren beleidsevaluatie en gaan in op de motieven om beleid te evalueren. Op basis daarvan plaatsen we evaluatie in de bredere beleidscyclus en gaan we in op verschillende types van evaluatie. Hoofdstuk 3 zoomt in op de Belgische en Vlaamse context voor beleidsevaluatie.

Vervolgens gaat deel 1 van de handleiding dieper in op de diverse stappen in het evaluatieproces. In hoofdstuk 4 staat het design van een evaluatie centraal. In hoofdstuk 5 gaan we in op de rol van beleidstheorie in evaluatie en vervolgens op de vraag: Welk evaluatietype is het best geschikt naargelang de informatiebehoefte? Voor een hele reeks subtypes van ex ante en ex post evaluatie bespreken we op de opzet en vraagstellingen, de meerwaarde en de aanpak en bijhorende technieken. Hoofdstuk 6 dekt een aantal aspecten van het management van evaluaties af en gaat in op de kwaliteitsdimensie. In hoofdstuk 7 ten slotte presenteren we een groeimodel voor de verdere ontwikkeling en professionalisering van evaluatiecapaciteit. Voorbeelden en checklists stofferen de verschillende hoofdstukken in dit eerste boekdeel.

Deel 2 van de handleiding gaat in op het monitoren van beleid. Monitoring vormt een potentieel belangrijke informatiebron voor beleidsevaluatie. We verkennen de onderlinge relatie tussen monitoring en evaluatie, hun gelijkenissen en verschilpunten. Verder presenteren we een algemeen stappenplan om een monitoringsysteem te ontwikkelen en te introduceren. We gaan uitgebreid in op het meten van beleid en beleidsresultaten, het werken met indicatoren en op de belangrijkste aandachtspunten daarbij. Daarnaast geven we ook een scenario aan voor de verdere uitbouw van bestaande prestatie-meetsystemen ter ondersteuning van het overheidsmanagement naar een breder beleidsperspectief.

Deel 3 van de handleiding is gewijd aan een brede waaier van technieken die men kan gebruiken om beleid te evalueren. Daarbij werd geen volledigheid nagestreefd: enkel de meest courante evaluatietechnieken werden geselecteerd. Dit is bedoeld als kompas voor de lezer om binnen een omvangrijk evaluatie-instrumentarium de meest adequate techniek te vinden op basis van de informatiebehoefte en de evaluatiecontext. Voor elke techniek geven we een omschrijving en telkens gaan we in op de functie en de context voor gebruik. Verder bespreken we in hoofdlijnen de concrete toepassing van de techniek en de sterktes en zwaktes. Ten slotte geven we nog een aantal referenties voor verdere lectuur.

Deel 4 van de handleiding gaat dieper in op beleidsevaluatie als praktijkveld. Aan de hand van een reeks praktijkvoorbeelden worden diverse elementen en aspecten van het evaluatieproces, die in de voorgaande delen van de handleiding aan bod kwamen, geïllustreerd en geduid. Doorheen de voorbeelden komt een mix aan bod van beleidssectoren, evaluatietypes (ex ante, tussentijds, ex post), evaluatievragen en -criteria, evaluatietechnieken, bestuursniveaus en landen. Enerzijds komen evaluaties binnen een beleidsveld of beleidsdomein aan bod, anderzijds twee bijzondere vraagstukken: evaluatie van domeinoverschrijdend beleid en van interbestuurlijk beleid.

Deel 5 van de handleiding is opgevat als een ondersteunend deel. Naast een glossarium met een verklaring van de voornaamste termen die in de voorgaande delen aan bod komen bevat het ook een reeks referenties naar aanvullende internet- en literatuurbronnen met betrekking tot evaluatie.

2. Beleidsevaluatie en monitoring in een overheidscontext

2.1 Evaluatie bij de overheid: ontstaan, ontwikkeling en trends

2.1.1 Ontstaan

Evaluatie is van alle tijden. Daarenboven kan men evalueren zien als een basisvorm van menselijk gedrag. In het alledaagse leven evalueert ieder van ons bij het nemen van beslissingen en meerbepaald bij het maken van keuzes, bijvoorbeeld bij de aankoop van een huishoudelijk apparaat, de keuze van een woonplaats of in onze inter-persoonlijke relaties. Dergelijke evaluaties zijn niet noodzakelijk een individueel proces. Ze kunnen ook in een groepsdynamiek tot stand komen.

Wanneer we evalueren in eerste instantie omschrijven als 'het onderzoeken en beoordelen, het bepalen van meerwaarde', dan ging de praktijk van evaluatie ongetwijfeld zijn definitie lang vooraf in de tijd met een oorsprong bij het begin van het menselijke bestaan. De Neanderthalers deden aan evaluatie wanneer ze vaststelden met welke types jonge bomen ze de beste speren konden vervaardigen. Hetzelfde geldt voor de strijdende Engelse kleine landeigenaren die hun eigen kruisbogen inruilden voor de handboog uit Wales. Ze stelden immers vast dat de handboog het meest solide pantser kon doorboren en bovendien niet één maar drie pijlen tegelijk kon lanceren. De Engelsen namen de handboog over, perfectioneerden hem en bleken onoverwinnelijk tijdens de 100-jarige oorlog. Ten nadele van de Fransen. Nochtans experimenteerden ook zij kort met de handboog maar zij grepen toch terug naar de kruisboog. Waarop hadden de Fransen zich dan wel gebaseerd? In elk geval bleven de Fransen verder strijden met een inferieur wapen, met strategische verliezen tot gevolg.

Bij alledaagse evaluaties en keuzes spelen vaak impressies en subjectieve percepties de voornaamste rol. Ze vormen de biotoop van de reclame en marketing. Eén van de alternatieven wordt daarbij op een informele wijze gekozen. Dat betekent: zonder het systematisch verzamelen van formele bewijselementen op een reeks van uitgesproken criteria waarmee de relatieve meerwaarde van elk alternatief wordt bepaald. Bij dergelijke *informele* evaluaties kunnen ervaring, instinct, veralgemening en redering elk de uitkomst van de evaluatie beïnvloeden. Ze kunnen de basis vormen voor een goed oordeel of 'wijze beslissing', maar evenzeer voor een foute inschatting. In het dagelijkse leven is informele evaluatie overigens vaak het enig praktische alternatief. Wie zal immers tijdens een culinaire uitstap, als voorbereiding van de eigen menukeuze, gegevens verzamelen via een exit poll bij andere restaurantbezoekers?

In deze handleiding staan echter de meer *formele* vormen van evaluatie centraal, en meerbepaald (formele) beleidsevaluatie in de overheidscontext. Formele beleidsevaluaties ondersteunen

beleidsbeslissingen door deze laatste te baseren op systematische inspanningen om expliciete criteria te bepalen en accurate informatie te verzamelen, waarmee de (meer)waarde van een beleidsoptie of gekozen en uitgevoerd alternatief kan worden beoordeeld.

Verderop staan we uitgebreid stil bij de motieven voor beleidsevaluatie. In wat nu volgt schetsen we twee evoluties: het chronologische verhaal van de algemene opgang van (formele) beleidsevaluatie bij de overheid als praktijkveld en de ontwikkeling van de inhoudelijke verhaallijnen binnen beleidsevaluatie als theorieveld.

2.1.2 Beleidsevaluatie als praktijkveld: de mijlpalen

Systematische evaluatie van overheidsbeleid was niet onbekend voor 1930 maar tot dan was het ook geen herkenbare trend. Socrates (5^{de} eeuw v. C.) hanteerde bijvoorbeeld verbale evaluaties als onderdeel van het onderwijsproces. Maar het duurde nog eeuwen vooraleer formele evaluaties wedijverden met religieuze en politieke overtuigingen als de ratio achter beslissingen met een sociale impact.

De 19^{de} eeuw werd zowel in Europa als de Verenigde Staten gekenmerkt door inspanningen tot hervormingen in een reeks van onderwijs- en sociale programma's, tegen de achtergrond van de industriële revolutie met haar grote maatschappelijke impact. Evaluaties van deze programma's en de gerelateerde instellingen waren meestal nog informeel en impressionistisch van aard. Onderwijs werd de eerste beleidssector waarin men de eerste systematische evaluaties van enige schaal opzette. In Groot-Brittannië werden koninklijke controlecommissies door de regering ingesteld om onderwijsinstuties op te volgen; de voorbode van de moderne externe inspecties. In de VS bestudeerde J. Rice de spellingscapaciteiten van 33.000 leerlingen binnen het schoolsysteem van een grootstad. Uit een vergelijking van scholen concludeerde hij dat er geen significante leereffecten toe te wijzen waren aan de relatief sterke aandacht die spellingsonderricht op dat moment kreeg in bepaalde scholen. Een andere belangrijke benadering tot evaluatie die werd ontwikkeld was deze van de gestandaardiseerde testmethode. Hiermee beoogde men een grotere efficiëntie en daardoor een groter bereik van leerlingen met onderwijsmiddelen.

Beleidsevaluatie mag pas recenter onderkend zijn als een apart praktijkveld, de moderne evaluatiepraktijk borduurt duidelijk verder op ideeën en technieken die in een ver verleden al werden toegepast. De eerste mijlpalen in de ontwikkeling van formele beleidsevaluatie vinden we terug in de Verenigde Staten. In een later stadium volgde de ontplooiing en verspreiding binnen Europa. Echte stuwkracht werd gegeven met het werk van R. Tyler in de VS, vanaf de jaren '30. Kenmerkend voor zijn benadering was de concentratie op duidelijk bepaalde doelstellingen. De facto definieerde hij daarmee evaluatie destijds als het onderzoek naar doelbereiking en verlegde hij de aandacht van middelen naar

resultaten. Zijn visie had een belangrijke invloed op de ontwikkeling en evaluatie van curricula en onderwijsmethoden in het Amerikaanse onderwijs.

De jaren '40 en '50 waren in de VS een herstelperiode na de Grote Depressie en WO II, met sterke sociaaleconomische verschillen in de maatschappij en tegelijk expansie, overconsumptie en sociale armoede. Dit klimaat straalde af op het onderwijs waar er weinig belangstelling voor verantwoording en evaluatie was. Weliswaar werden heel wat data verzameld en meet- en analysemethodes technisch verfijnd, dit alles gebeurde echter zonder een klare kijk te hebben op de informatiebronnen met het oog op de beoordeling en verbetering van het onderwijs. Evaluatie was toen vooral een zaak van de lokale afdelingen.

In de jaren '60 komt hierin verandering met de ontwikkeling van nationale onderwijsprogramma's en federale fondsen om deze te evalueren. Echter, de bestaande benaderingen van evaluatie en de federaal voorgeschreven methodiek brachten geen antwoord op de evaluatiebehoeften in het lokale werkveld. De benadering ontwikkeld door Tyler kon dan weer niet zomaar getransponeerd worden naar evaluaties op federale schaal. Meer algemeen kenmerkend voor dit tijdssegment was een zekere identiteitscrisis bij evaluatoren. Er bestond onzekerheid over hun rol: waren ze onderzoeker, ambtenaar, onderwijzer of filosoof? Over de vereiste kwalificaties en vaardigheden waarover een evaluator toen diende te beschikken was er evenmin duidelijkheid. Daarenboven was er een gebrek aan ondersteuning via professionele organisaties, vormingsmogelijkheden en gespecialiseerde literatuur waarin ideeën konden worden gedeeld en uitgewisseld. Het evaluatielandschap bleek vooralsnog vormloos en gefragmenteerd.

Vanaf de late jaren '60 en in de jaren '70 veranderde dit beeld opmerkelijk, terug met het onderwijs als voornaamste werkveld. Onder impuls van diverse auteurs kwamen nieuwe alternatieve opvattingen en bijhorende technieken op het vlak van evaluatie tot ontwikkeling. Het tegensprekelijke debat rond alternatieve benaderingen van evaluatie beleefde vruchtbare tijden. Dit kristalliseerde zich in tijdschriften en opleidingspakketten aan universiteiten. Deze dynamiek versterkte ook de aandacht voor meta-evaluatie of de zorg voor de kwaliteitsbewaking van evaluaties aan de hand van standaarden. De aandacht voor evaluatie steeg in een steeds breder wordende waaier van beleidsvelden.

Het is in dezelfde tijdsperiode dat de aandacht voor beleidsevaluatie naar Europa overgewaaid is en evaluatie als praktijkveld impulsen kreeg. De Europese landen vertonen echter een 'géometrie variable' waarmee ze een evaluatiecultuur hebben ontwikkeld.

Een belangrijke vraag hierbij is: Wanneer kan men spreken van een evaluatiecultuur? Furubo (2002) hanteert volgende criteria:

- Beleidsevaluatie gebeurt in veel beleidsdomeinen
- Evaluatoren komen uit verschillende disciplines

- Er is een nationaal discours met betrekking tot beleidsevaluatie
- Er zijn professionele organisaties
- De graad van institutionalisering in de uitvoerende macht (regering)
- De graad van institutionalisering in de wetgevende macht (parlement)
- Pluralisme van instanties of evaluatoren die evaluaties uitvoeren binnen elk beleidsdomein
- Evaluatie door het hoogste audit orgaan
- Proportie van evaluaties gefocust op effecten in relatie tot evaluaties gefocust op processen en output

Binnen Europa en tevens op internationaal vlak is de ontwikkeling en verspreiding van een evaluatiecultuur in een tweetal golven ontstaan. De eerste golf werd gevormd door een selecte kopgroep van landen (Zweden, Duitsland, het Verenigd Koninkrijk en ook Canada als niet-Europees land) waarbinnen men eind jaren '60, begin jaren '70 van een groeiende evaluatiecultuur kan spreken. Gedurende de jaren '80 kwamen slechts enkele landen aansluiten. De tweede golf komt er in de jaren '90 met een peloton van landen waaronder ook heel wat (West-)Europese staten.

Hoe is deze relatief trage en ongelijke verspreiding binnen Europa te verklaren? Het is aannemelijk dat een mix van kenmerken in de pionierslanden (1^{ste} golf) een *interne druk* heeft gecreëerd waaruit een evaluatiecultuur is ontstaan. Er wordt daarbij verondersteld dat deze kenmerken en omstandigheden zodanig specifiek voor deze landen waren, waardoor het niet vanzelfsprekend was dat dezelfde ontwikkeling zich herhaalde in andere landen.

Een drietal categorieën van factoren wordt in dit verband aangehaald. Een eerste betreft een *politië-culturele vatbaarheid* voor evaluatie; een sterke politieke traditie van een geloof in de maakbaarheid van de samenleving, instrumentalisme en een voorkeur voor rationalistische attitudes. Ten tweede, kenmerken die de verspreiding van ideeën over evaluatie binnen het politiekambtelijke systeem bevorderen, bijvoorbeeld de vertrouwdsheid van ambtenaren met sociaalwetenschappelijk denken en de mate waarin door intensere contacten en uitwisseling met de VS wat leidde tot een relatief sterkere invloed van de Amerikaanse evaluatietraditie. Ten derde was er in de pionierslanden *een grotere (gepercipieerde) nood aan evaluatie*, die mogelijk kan correleren met de graad van overheidsinterventie. De redenering achter deze laatste hypothese is enerzijds dat een sterk geloof in de maakbaarheid van de samenleving de aandacht richt op technische oplossingen en kennisopbouw via wetenschappelijke methoden en analyse. Anderzijds wordt verondersteld dat een grotere rol van de overheid (af te lezen aan het uitgavenniveau) leidt tot informatiebehoefte over de resultaten en effecten en de kostprijs van beleid. Deze laatste assumptie is empirisch te staven: de pionierslanden hadden ten tijde van hun

groeïende evaluatiecultuur (jaren '60-'70) inderdaad een hoger uitgavenpeil dan de landen die pas in de 2^{de} golf meer aandacht besteedden aan evaluatie.

Van belang is ook dat de versterkte aandacht voor evaluatie in de pionierslanden niet beperkt bleef tot de beleidsdomeinen waarbinnen de evaluatiepraktijk zich in de Verenigde Staten het sterkst had ontwikkeld: onderwijs en gezondheidszorg. Hoewel deze sectoren ook in deze landen bovenaan op de hervormingsagenda stonden, waren er weldegelijk oversijpelingseffecten van de evaluatiefilosofie naar tal van andere beleidsdomeinen.

Met de introductie van evaluatie in de eerste Europese landen werden ook ondersteunende structuren ontplooid (netwerken, vorming, expertisecentra, literatuur, enz.) die op hun beurt het pad effenden voor de verspreiding van evaluatie-ideeën naar andere Europese staten. Nochtans duurde het zoals eerder vermeld tot de jaren '90 vooraleer beleidsevaluatie ingang kreeg in een grote groep landen. De specificiteit van de aangehaalde factoren in de pionierslanden maakten dat de verspreiding destijds halt hield aan hun grenzen. Anderzijds speelde voor de landen in de 2^{de} golf niet zozeer een interne druk maar vooral een *externe druk* mee bij het introduceren van evaluatie in de overheid.

In het bijzonder de algemene trends van hervormingen en modernisering met betrekking tot het overheidsmanagement vormden een adequaat medium voor evaluatie-ideeën om verder verspreid te worden naar andere landen. Daarmee is wel meteen een verschilpunt aangehaald met de pionierslanden: terwijl daar de evaluatie-ideeën een zelfstandig geheel vormden en een strategische keuze vormden, kwamen deze ideeën bij de landen van de 2^{de} golf als onderdeel van een breder pakket van moderniseringsideeën (bv. het 'New Public Management' (NPM) van begin jaren '90) die de hervormingsagenda's mee bepaalden. Ondertussen is gebleken dat het brede NPM gedachtegoed in uiteenlopende mate en in veel gevallen ook 'à la carte' werd geïmplementeerd in de verschillende Europese landen. Dat in de jaren '90 plots een heel peloton van landen eveneens een evaluatiecultuur begon te ontwikkelen, heeft dan ook veel te maken met een gestegen externe druk vanwege internationale organisaties, voornamelijk de Europese Unie, de Wereldbank en de OESO.

Terwijl op internationaal vlak de Wereldbank een grote rol speelde doordat evaluatie een integraal onderdeel werd van de werkingsprocedures in het kader van ontwikkelingssteun, was ook de OESO een belangrijk forum voor de verspreiding van een evaluatiecultuur. Maar bovenal vormde (en vormt) het Europese integratieproces een katalysator voor de verspreiding van de ideeën en praktijk van beleidsevaluatie in de Europese landen. Evaluatie werd een essentiële voorwaarde in de toetredingsprocedure en een belangrijk aspect in de interbestuurlijke relaties (in een steeds verder openklappende waaier van beleidsdomeinen) van de Europese instellingen met de lidstaten en regio's. Ook niet-lidstaten als Noorwegen en Zwitserland, de facto sterk verweven met de EU, ontwikkelden

gelijktijdig een evaluatiecultuur en een bijhorend institutioneel raamwerk. Men mag veronderstellen dat ook voor de recentste lidstaten uit Centraal- en Oost-Europa evaluatie een belangrijker thema binnen de overheid zal worden.

Daarmee zien we binnen Europa een heel palet van trajecten met betrekking tot de mate waarin de ideeën en praktijk van beleidsevaluatie werden ontwikkeld binnen de overheid. Welk traject een bepaald land aflegt, hangt af van de omvang van interne en externe druk die ontstaat in een gegeven tijdgeest. Voornamelijk op het vlak de interne druk zijn er landenverschillen waarneembaar, bijvoorbeeld welke actor kan beschouwd worden als de 'promotor' van evaluatie. In enkele landen, bv. Duitsland en de Scandinavische landen, heeft het parlement een bepalende invloed gehad. In Nederland heeft de Algemene Rekenkamer initiatieven genomen tot een interne capaciteitsopbouw voor beleidsevaluaties, naast haar audittaak. In andere landen, bv. het Verenigd Koninkrijk en Italië, kwamen de impulsen tot evaluaties dan weer voornamelijk vanuit het ministerie van financiën.

Dergelijk traject en de aspiraties van de binnenlandse promotor hebben vanzelfsprekend een neerslag in de wijze waarop en de mate waarin evaluatie cultureel en structureel verankerd werd in de beleids- en managementcontext van dat land.. In het Verenigd Koninkrijk was besparen een belangrijk motief bij de lancering van het 'Financial Management Initiative' van Thatcher, terwijl enkele jaren later in Frankrijk het accent lag op de versterking en ondersteuning van het democratische debat wanneer premier Rocard een systeem van interministeriële evaluatie invoerde. Dit voorbeeld toont enerzijds aan hoe op uiteenlopende wijze twee systemen hebben bijgedragen tot het verspreiden van een evaluatiecultuur binnen het ambtelijke apparaat.

Anderzijds is er de vraag of dergelijk anders lopende trajecten op langere termijn de politiekambtelijke cultuur in elk van de betrokken landen zullen beïnvloeden. De diversiteit op dat vlak lijkt echter niet zo sterk dat men van verschillende modellen of culturen van evaluatie in de Europese context kan spreken. Het gaat eerder om variaties in stijl, waarbij verschillen aan het vervagen zijn naarmate de professionalisering van beleidsevaluatie toeneemt.

De verklaring achter deze tendens naar convergentie is dat beleidsevaluatie meer en meer evolueert naar een eigenstandig praktijkveld. Dat leert een aftoetsing op enkele criteria die in dit verband relevant worden geacht. Ten eerste is er een groeiende vraag naar evaluatiespecialisten binnen de overheid. Ten tweede is de literatuur rond beleidsevaluatie meer en meer zichtbaar geworden door middel van gespecialiseerde tijdschriften en boeken. Ten derde besteden universiteiten, bedrijven en agentschappen meer aandacht aan specifieke opleidingsprogramma's voor evaluatoren, al loopt Europa hier nog achter op de Verenigde Staten. Een vierde aspect is de graad van institutionalisering van evaluatie. Ook daarvoor bestaat er empirische bevestiging, zij het dat er belangrijke verschillen blijven bestaan tussen

landen en ook tussen bestuursniveaus in eenzelfde land. Daarenboven wordt de uitvoering van evaluatie vaak uitbesteed door de overheid. Het al dan niet bestaan van certificering is een volgend criterium om de professionalisering van een branche te beoordelen. Dit mechanisme met betrekking tot beleidsevaluatie vooralsnog niet ontwikkeld. Het debat loopt momenteel over de wenselijkheid en haalbaarheid van het systeem. Voorstanders zien hierin een kwaliteitsgarantie, tegenstanders vrezen een beknotting van de bestaande en toekomstige diversiteit aan alternatieve evaluatiebenaderingen. Ook de inherente politieke dimensie van beleidsevaluatie vormt mogelijk een verklaring voor de vooralsnog uitgebleven introductie van accreditatie voor evaluatoren.

Een andere graadmeter die voor beleidsevaluatie wel al toepasselijk is, betreft het stijgende aantal evaluatieassociaties. Zij vervullen een belangrijke forumfunctie voor de uitwisseling van ideeën en ervaringen en tevens een brugfunctie tussen theorie en praktijk. De meeste evaluatieassociaties brengen zowel ambtenaren, academici als private actoren uit het praktijkveld van beleidsevaluatie bij elkaar.

In de VS ontstond in 1986 de *American Evaluation Association* (AEA) uit het samengaan van twee bestaande organisaties, de *Evaluation Research Society* (ERS) en het *Evaluation Network* (ENet). In Europa zag de *European Evaluation Society* (EES) het levenslicht in 1995. De EES speelt sindsdien een belangrijke rol in de verspreiding van een evaluatiecultuur op het Europese continent, voornamelijk door de organisatie van een tweejaarlijkse conferentie. Meer recent werd in heel wat landen een nationale evaluatieassociatie opgericht: Duitsland, het Verenigd Koninkrijk, Zwitserland, Italië, Spanje, Zweden, Finland, Polen en Frankrijk. Ook in Portugal en Hongarije werden de eerste stappen daartoe gezet.

In België bestaat er een Waalse evaluatieassociatie, opgericht in 1999: de *Société Wallonne de l'Evaluation en de Prospective* (SWEP). Eind 2006 werd de eerste aanzet gegeven voor de oprichting van een Vlaams evaluatieplatform binnen de Vlaamse Vereniging voor Bestuur en Beleid (VVBB).

Tabel 1 bevat een overzicht van evaluatienetwerken in de VS, Canada en Europa.

Land	Naam netwerk	Opstartjaar	Leden
Verenigde Staten	(Evaluation Research Society)	1976	-
	(Evaluation Network)		-
	American Evaluation Association (AEA)	1986	+/- 4000
Canada	Canadian Evaluation Society (CES)	1981	1924
Verenigd Koninkrijk	United Kingdom Evaluation Society (UKES)	1994	300
Europa	European Evaluation Society (EES)	1995	450
Zwitserland	Schweizerische Evaluationsgesellschaft (SEVAL)	1996	350
Duitsland	Deutsche Gesellschaft für Evaluation (DeGEval)	1997	400
Italië	Associazione Italiana di Valutazione	1997	340
Frankrijk	Société française de l'Evaluation (SFE)	1999	250
Finland	Suomen arviointiyhdistykset	1999	n.b.
Polen	Polskie Towarzystwo Ewaluacyjne (PTE)	2000	n.b.
Denemarken	Dansk EvalueringsSelskab (DES)	2000	170
Spanje	Sociedad Española de Evaluación (SEE)	2001	96
Nederland	Beroepsvereniging VIDE	2002	211
Zweden	Svenska utvärderingsföreningen (SVUF)	2003	200

Tabel 1: Overzicht van evaluatienetwerken in de VS, Canada en Europa

Samenvattend kan gesteld worden dat beleidsevaluatie aan heel wat relevante criteria beantwoordt om te spreken van een eigenstandig praktijkveld en dat zowel in Noord- als Zuid-Europa beleidsevaluatie een belangrijk thema werd voor de overheid.

Deze net beschreven opgang hangt vanzelfsprekend samen met de groeiende vraag vanuit de overheid naar evaluatie in het algemeen en naar beleidsevaluatie in het bijzonder. Logischerwijze volgt dergelijke vraag naar evaluatie uit de zoektocht naar het antwoord op de 2 belangrijke kernvragen die elke overheid zich voortdurend moet stellen: “*Doen we de dingen goed?*” en “*Doen we de goede dingen?*”. Beide vraagstellingen zijn essentieel. De praktijk zal verschillen tussen landen vertonen naargelang welke vraagstelling de meeste aandacht krijgt en in het recente verleden werd de klemtoon wellicht vaak gelegd op de eerste vraag, in het kader van managementhervormingen en moderniseringstrends. Echter, de daaraan complementaire vraag “*Doen we de goede dingen?*” is vanuit beleids oogpunt misschien wel de meest fundamentele vraagstelling.

Vandaag wordt van de overheid verwacht dat ze voortdurend haar rol beargumenteert en verantwoordt. Tegelijk neemt de maatschappelijke context waarin deze overheid wordt gevraagd om op te treden in complexiteit toe. Daardoor wordt de vraagstelling “*Doen we de goede dingen?*” -zonder de andere op te geven - des te meer relevant. Deze vraag krijgt ook in België en Vlaanderen meer en meer aandacht. Beleidsevaluatie kan een unieke bijdrage leveren tot de antwoorden. De tendens van versterkte aandacht voor beleidsevaluatie bij de overheid vormde mee de aanleiding tot het schrijven van deze handleidingenset.

2.1.3 Beleidsevaluatie als theorieveld: stromingen, kenmerken en tendensen

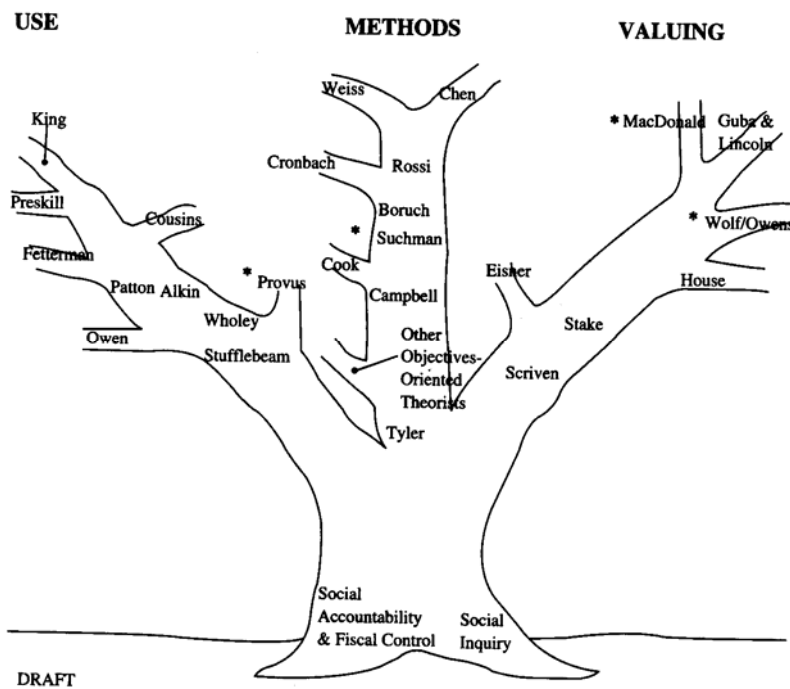
Beleidsevaluatie mag geleidelijk tot een identificeerbaar praktijkveld geëvolueerd zijn, daartegenover staat de verscheidenheid in theorieën en benaderingen van de evaluatiepraktijk en de brede waaier aan methoden en technieken die in de evaluatiepraktijk toegepast worden. Ook daarover bestaat inderdaad consensus: evaluatie is als theorieveld veelzijdig en op methodologisch vlak interdisciplinair. In wat volgt gaan we dieper in op beide aspecten.

Wat zijn de essentiële kenmerken van het 'acquis communautaire' in het evaluatiedenken? Typerend voor het geheel aan **evaluatietheorie** is allereerst de veelsoortigheid op het vlak van benaderingen, opvattingen en klemtonen. Dat maakt het niet eenvoudig om beleidsevaluatie als theorieveld overzichtelijk te maken. Alkin (2004) en Christie hebben met hun werk 'Evaluation Roots' een interessante en verdienstelijke bijdrage geleverd. Ze hanteerden daarvoor de metafoor van een boom.

In de wortels van de boom kan de oorsprong en aanleiding tot theorievorming over evaluatie geplaatst worden. Vanuit deze 'roots' is de boom verder gegroeid en heeft hij zich vertakt. Daarbij zijn meerdere deelstammen - theoretische stromingen of benaderingen- ontstaan. De deelstammen vertakken op hun beurt verder, naargelang andere accenten worden gelegd of verschillende perspectieven worden gehanteerd. Het vertakken van de deelstammen verloopt tegen ongelijke snelheid. De stammen kunnen verder uit elkaar of naar elkaar toe groeien, of vertakkingen van de ene stam kunnen bewegen naar vertakkingen van een andere stam.

Hoe hebben de auteurs de boom getekend en ingedeeld? De eerste stap bestond uit een stelselmatige screening en vergelijking van de bestaande theoretische bijdragen tot evaluatie van een hele reeks geselecteerde auteurs, die men door hun werk als theoretici kan beschouwen. Op basis daarvan hebben de auteurs een eerste draft van boom getekend. Vervolgens werd terugkoppeld naar de theoretici en gevraagd of zij zich konden vinden in de positionering van hetzij zichzelf, hetzij een andere (al overleden) auteur in de boom. Gebaseerd op hun feedback werd de boom door Alkin en Christie in een laatste stap nog verder verfijnd.

Het resultaat, weergegeven in figuur 1, laat zien dat beleidsevaluatie een dubbele 'roots' heeft: enerzijds in de vraag naar verantwoordelijkheid en anderzijds in het maatschappelijk onderzoek. Deze twee elementen hebben de ontplooiing van het veld op een verschillende manier beïnvloed. De vraag naar verantwoordelijkheid vormde de aanleiding tot evaluatie. Als oorsprong voor beleidsevaluatie wordt verantwoording zo breed mogelijk opgevat; het is bestemd om het beleid en de maatschappij waarin we leven te verbeteren. Het maatschappelijk onderzoek als tweede oorsprong komt voort uit het belang van het aanwenden van een systematische en verantwoorde set van methoden om verantwoordelijkheid te bepalen. Terwijl de vraag naar verantwoording de aanleiding vormt, werden modellen of benaderingen voor evaluatie voornamelijk ontleend aan het maatschappelijk onderzoek.



Figuur 1: De ‘evaluation theory tree’

Vanuit deze wortels is de theorievorming rond evaluatie als een boom met drie deelstammen gegroeid. Elk van deze deelstammen vertoont verdere vertakkingen naargelang bepaalde auteurs in hun werk specifieke invalshoeken hanteerden of welbepaalde aandachtspunten verder uitdiepten. Een eerste hoofdstroming legt verder de klemtoon op methoden voor beleidsevaluatie en is de voortzetting van de wortel van het maatschappelijk onderzoek. Bij de tweede hoofdstroming staat het vormen van oordelen en de rol van de evaluator daarin centraal. Het hechten van een ‘waarde’ op gegevens en informatie over het beleid wordt hierbij beschouwd als de meest essentiële stap in het evalueren. De verdere vertakkingen van deze stam zijn deels een vertaling van het debat rond de vraag of deze stap van waardebepaling een exclusieve taak is van de evaluator, dan wel dat deze eerder anderen (beleidsmaker, opdrachtgever) daarbij dient te begeleiden. De derde hoofdstroming focust op het gebruik van evaluaties in relatie tot beslissingen over beleid. De verdere vertakkingen vertegenwoordigen alternatieve benaderingen van evaluatie georiënteerd op het gebruik of de invloed van evaluatie in het kader van beleid.

De sterkte van de boommetafoer is dat men ermee door de bomen het bos kan zien. De veelheid aan theorieën en benaderingen wordt overzichtelijk gestructureerd. Het nadeel van deze werkwijze is dat het een statisch beeld oplevert over de (onderlinge) positionering van de verschillende theoretici. Auteurs kunnen echter ook evolueren in hun denken en klemtonen verleggen in hun oeuvre. Om dergelijke individuele ontwikkelingen in het evaluatiedenken te visualiseren lijkt bijvoorbeeld een riviermetafoer

meer aangewezen. Voor een verdere bespreking van de verschillende auteurs in elke hoofdstroming verwijzen we de lezer naar het werk van Alkin (2004).

Opvallend is dat het merendeel van de in de boom opgenomen theoretici uit de Verenigde Staten komen. Dit was enigszins te verwachten daar het evaluatiedenken daar eerst tot ontbolstering kwam. Anderzijds kan men vaststellen dat op dit vlak de Amerikaanse trekkersrol tot vandaag onveranderd bleef. Het is inderdaad niet zo eenvoudig om een lijst met Europese theoretici in evaluatie samen te stellen, laat staan een Europese toepassing van de boommetafoor uit te werken. In de zoektocht naar verklaringen kan op een tweetal factoren gewezen worden.

Ten eerste is de evaluatiepraktijk pas in een later stadium naar Europa overgewaaid, op een moment dat de fundamentele voor theorievorming al gelegd waren in de VS. In Europa kon men daarop verder bouwen.

Ten tweede, en wellicht fundamenteler, kan men toch een belangrijk verschilpunt tussen de VS en Europa onderkennen in de wijze waarop beleidsevaluatie als praktijk ingang vond. In Europa zijn de impulsen meer vanuit de overheid zelf gekomen, terwijl in de VS de vraag naar evaluaties meer vanuit het maatschappelijke veld is gekomen, met de zogenaamde foundations of community-based actoren als katalysator. Deze laatste soort van maatschappelijke constructies is minder prominent aanwezig in Europa. De verklaring voor dit verschil vergt nader onderzoek. Een hypothese vertrekt van een verschil in vertrouwen vis-à-vis de overheid; een hoger wantrouwen zou een kritisch bevragen of evalueren van de overheid door externen in de hand werken.

Het vastgestelde verschil in locus van de impulsen tot evaluatie van beleid kan wel deels de vraag beantwoorden waarom in Europa het evaluatiedenken moeilijker terug te brengen is naar individuele theoretici. Als evaluatie in Europa meer vanuit de overheid zelf is ontstaan, moet misschien meer belang worden gehecht aan de rol die instituties hebben gespeeld in de ontwikkeling van de evaluatiepraktijk. Daarmee zijn we terug aanbeland bij het diffusieverhaal van een evaluatiecultuur in de Europese landen. Zij verschillen op het vlak van de trekkende actor bij het ontstaan van een evaluatiecultuur: parlement, regering, departement, rekenkamer, enz. Bovendien werd in de meeste Europese landen beleidsevaluatie niet als een alleenstaand en alomvattend pakket van ideeën en tools ingevoerd, maar vaak incrementeel en ingebed in managementgeoriënteerde hervormingstrends. Het discours van de trekkende actor heeft dan in belangrijke mate de benadering van evaluatie vorm gegeven in een bepaald land.

Maar zoals eerder aangehaald heeft dit niet geleid tot verschillende evaluatieculturen in Europa. Bovendien is er meer en meer een trend naar uitwisseling van ideeën en ervaringen over beleidsevaluatie tussen de VS en Europa, maar ook met andere continenten. Internationale conferenties en structurele netwerken zijn daarbij de motor.

Tevens kenmerkend voor het geheel van evaluatietheorie is dat het voornamelijk bestaat uit *prescriptieve theorieën* ofwel theorieën van de evaluatiepraktijk. De verschillende benaderingen schrijven meestal voor, elk vanuit een bepaald perspectief of met eigen klemtoon, hoe een evaluatieproces moet verlopen. Het gaat dus om een brede waaier van zogenaamde middle range theorieën: ze verklaren niet alles, maar dragen kritieke slaagfactoren met een beperkte draagwijdte aan. Het debat loopt al een tijdje of het wenselijk en haalbaar is om te evolueren naar meer integratie binnen de grote diversiteit die vandaag zo kenmerkend is voor het evaluatiedenken.

Wie dieper graaft in de inhoudelijke diversiteit tussen evaluatiebenaderingen zal stoten op een fundamentele onderscheid tussen bepaalde paradigma's die eraan ten grondslag liggen. Er bestaat geen universele filosofie van de wetenschap. Hetzelfde geldt dan ook voor het evaluatiedenken, dat bijvoorbeeld niet ontsnapte o aan de courante tweedeling in onderzoeksparadigma's tussen positivisme en constructivisme.

Het positivistische paradigma stelt dat kennis (enkel) gebaseerd is op natuurlijke fenomenen, hun proporties en relaties, zoals dat empirisch kan worden vastgesteld. De beperkingen van deze (strakke) zienswijze zijn ondertussen wel onderkend: de nood aan interpretatie van wat men empirisch vaststelt aan de hand van referentiekaders of theorieën, onvermijdbare effecten door het meten zelf die het beeld kunnen vertekenen, contextgebonden factoren die verwachte patronen doorbreken, de moeilijkheid om causaliteit vast te stellen, en de waardegeladenheid van menselijke oordelen waardoor ervaren problemen even belangrijk zijn als 'feiten'.

Dergelijke kritieke aspecten werden beter in acht genomen bij denkstromingen die men groepeerd onder het zogenaamde postpositivisme. Het *constructivisme* is daarvan een radicale variant. Dit paradigma verwerpt de meeste assumpties van het positivisme. Kennis bestaat binnen deze zienswijze niet objectief maar wordt opgebouwd door mensen. De realiteit is een sociale constructie van een individuele of groepsgebonden omschrijving van de waargenomen situaties.

Het *realisme* wordt nog als een derde perspectief geduid. Centraal daarbij staat het begrijpen van de verschillende contexten en referentiekaders die verklaring en interpretatie begrenzen. Het tracht, toegepast op evaluatie van beleid, de 'black box' te openen in beleid en programma's om de mechanismen te ontdekken die verandering meebrengen. Systemen die men vanuit deze invalshoek bestudeert, beschouwt men als open en de beleidsprogramma's als ingebed in gelaagde sociale en organisatorische processen waar mensen kunnen beslissen, handelen en leren.

Het verschil tussen de eerste twee paradigma's was in het verleden bron van debat en zelfs van rivaliteit waarbij men eerder in termen van een dichotomie dacht dan van continuüm. Nochtans zijn beide paradigma's niet incompatibel. Meer recent is het debat ook geluwd en is er een tendens naar een integratie van beide tradities.

Eén en ander heeft ook gevolgen gehad voor de evaluatiemethodologie. Methoden en technieken werden geassocieerd met (maar niet beperkt door) de respectieve paradigma's. Een voornaam onderscheid is dat tussen *kwantitatieve en kwalitatieve methoden*. Kwantitatieve methoden zijn gericht op het testen van specifieke hypothesen als deel van een breder theoretisch perspectief en staan daarmee dicht bij het traditionele model van sociale wetenschappen en statistische analysemethoden, waarbij de klemtoon ligt op standaardisering, precisie, objectiviteit en betrouwbaarheid, herneembaarheid en generaliseerbaarheid van de bevindingen. De kwalitatieve methoden hechten belang aan de onderzoeker zelf als belangrijk 'instrument' bij de verzameling en analyse van data en zijn geschikt voor zogenaamde 'thick description': het waarnemen van patronen van actie en betekenis vanuit het perspectief van de bestudeerde personen. De focus ligt op sociale processen in plaats van vooral of uitsluitend op resultaten. De analyse gebeurt vaker op een inductieve manier en het geldigheidsbegrip (validiteit) is dan sterker gebaseerd op de ervaring, expertise en vaardigheden van de individuele evaluator.

Gedurende drie decennia was er debat tussen de aanhangers van kwalitatieve en kwantitatieve methoden. Recent verwaterde eveneens de aandacht voor deze controverse. De meeste evaluatoren erkennen dat kwalitatieve methoden methodologische standaarden hebben die verschillend zijn, maar niet gebrekkig ten overstaan van kwantitatieve methoden. Vaak wordt de complementariteit en compatibiliteit van beide methoden in de praktijk bewezen. Het blijft uiteraard wel van belang dat men de uitgangspunten en beperkingen kent van de diverse paradigma's en methoden.

Verder is evaluatie op methodisch vlak *interdisciplinair*. Er is dan ook een permanente nood aan nieuwe opvattingen en inzichten over hoe bestaande methoden uit andere disciplines adequaat aangewend kunnen worden in functie van de vraagstellingen die centraal staan bij het evalueren van beleid.

2.1.4 Recente trends in evaluatie

Grote fundamentele debatten mogen dan al geluwd zijn, beleidsevaluatie blijft veelsoortig en veelzijdig op het vlak van benaderingen en methodologieën, ofwel theorie en praktijk. Daarenboven zijn er een aantal externe recente trends aanwijsbaar die een impact hebben op diverse facetten van evaluatie als theorie- en praktijkveld.

- *Nieuwe beleidsdimensies, beleidsvelden en beleidskwestie*

De veranderlijke maatschappelijke context waarin politiek en beleid tot stand komen heeft gevolgen voor evaluatie. Maatschappelijke veranderingsprocessen aan en de trend naar globalisering gaan samen met een toenemende verwevenheid en complexiteit op politiek, economisch en cultureel vlak. Dergelijke veranderingen dragen extra dimensies aan met betrekking tot gepercipieerde beleidsproblemen en verwachte oplossingen. Zo heeft beleid meer en meer een grensoverschrijdend karakter gekregen. Problemen en noden overstijgen geografische en administratieve grenzen. Accurate oplossingen vragen om samenwerking over grenzen heen. De vaak gehanteerde term governance duidt op het samenspel

van overheden en bestuurslagen, non-profit en private actoren. Een integrale of transversale aanpak overstijgt de grenzen van beleidssectoren of -domeinen. Bijgevolg verandert ook de beleidscontext voor evaluatie. De horizontale dimensie van beleid wordt belangrijker, bijvoorbeeld via thema's als duurzame ontwikkeling, gelijke kansen of gebiedsgericht beleid. De verticale dimensie wordt belangrijker: verschillende bestuurlagen vormen elk een onmisbare schakel in een langere beleidsketen. Er ontstaan dus andere schalen en setting waarbinnen beleid tot stand komt en geëvalueerd moet worden. Daarnaast ontstaan er ook nieuwe beleidsvelden en beleidskwesties.

- *Proliferatie van actoren en stakeholders*

Tevens verruimt het aantal actoren en stakeholders die betrokken zijn bij beleidsevaluatie. Dat leidt tot nieuwe spanningsvelden en uitdagingen voor het verloop van het evaluatieproces. Mogelijk bestonden ze al langer, maar in de geschetste veranderende context krijgen ze een versterkte betekenis of nieuwe invulling. Het kunnen omgaan van een veelheid aan belangen en perspectieven is daarvan een voorbeeld.

- *De rol van de evaluator*

Aansluitend bij het voorgaande punt wordt de explicitering van de rol van evaluator ook belangrijker. Wat is het mandaat van de evaluator? Welke rol wordt er verwacht? Hoe worden de relaties tussen de evaluator en de stakeholders georganiseerd? Gegeven de diversiteit aan evaluatiebenaderingen maar ook de verschillende argumenten om te evalueren, is het belangrijk om de status en de rol van de evaluator te expliciteren. Het motief om te evalueren en de evaluatiebenadering met daarbij de rol van de evaluator dienen goed doordacht op elkaar afgestemd te worden. Zo ontstaan er bijvoorbeeld raakvlakken beleidsevaluatie en audit raakvlakken. Het gaat telkens om een vorm van beoordelen, maar in essentie wordt vertrokken van verschillende vraagstellingen. Dit heeft gevolgen voor de rol, mogelijkheden en beperkingen van de evaluator.

- *Moderne versies van oude argumenten voor evaluatie*

Er ontstaan nieuwe aandachtspunten bij het evalueren van beleid. Meestal gaat het echter om moderne versies van oude argumenten ofwel motieven om beleid te evalueren. Het motief of de opzet voor een beleidsevaluatie heeft een impact op hoe de evaluatie gebruikt kan worden. Bovendien is het motief van belang met betrekking tot de methodologie en de rol van de evaluator.

Doorheen de diversiteit in evaluatietheorie en -praktijk (cf. supra) kan men de (brede range aan) beweegredenen voor hedendaagse evaluatie onderbrengen in een *kwartet van terugkerende algemene evaluatiemotieven*:

- Ondersteuning van beleidsplanning

- Opvolging van de beleidsuitvoering en institutionele ontwikkeling
- Verantwoording
- Beleidsleren

Deze motieven voor beleidsevaluatie worden in een volgende paragraaf nader besproken.

- *Evaluatie als theorie- en praktijkveld*

De bovenstaande factoren dragen bij tot het feit dat evaluatie als theorie- en praktijkveld in ontwikkeling is, zowel in de diepte als in de breedte. Bestaande theoretische benaderingen van de evaluatiepraktijk worden verder verfijnd, er ontstaan varianten en er worden alternatieve perspectieven uitgewerkt. Anderzijds stellen bepaalde auteurs dat er nood is aan het evolueren naar meer synthese binnen de veelsoortigheid. Daarbij komt het vraagstuk hoe het geheel aan evaluatietheorie verder gekoppeld en gefundeerd kan worden met de bredere beleids- en bestuurskundige theorieën. Andere auteurs wijzen net op de rijkheid die de diversiteit aan benaderingen aanbiedt. Ze zien daarin een mogelijkheid tot kruisbestuiving als mechanisme voor verdere theorieontwikkeling. Ze staven hun argumenten door te verwijzen naar de evaluatiepraktijk waar vaak verschillende methoden en technieken gecombineerd worden omdat ze als complementair worden beschouwd.

Het praktijkveld laat inderdaad ook een grote variëteit zien. We vermeldden al het interdisciplinaire karakter op methodologisch vlak. Auteurs zoals Scriven en Donaldson zien daarin een kans voor beleidsevaluatie op te evolueren naar zij noemen een 'transdiscipline'. Evaluatie heeft een zelfstandige *raison d'être* ter ondersteuning van het beleid. Maar daarnaast, zo stellen deze auteurs, kunnen evaluaties bijdragen aan de ontwikkeling van verschillende wetenschappelijke disciplines.

Uit de groeiende vraag naar evaluatie in de praktijk in combinatie met de bestaande veelzijdigheid volgt ook een stijgende nood aan mogelijkheden voor opleiding en vorming van evaluatoren. Veel personen krijgen er binnen hun dagdagelijkse functie op een maandagmorgen ook evaluatietaken bij. Vaak zijn ze echter niet op de hoogte van de verschillende benaderingen en technieken die er bestaan en hun mogelijkheden en beperkingen. Deze handleiding wil deze lacunes mee opvullen.

2.2 Definities van beleidsevaluatie

De lezer die op zoek is naar de ultieme of alleszeggende definitie van beleidsevaluatie moeten we meteen teleurstellen; er bestaan haast evenveel definities van evaluatie als er evaluatoren zijn, om het met een boutade te zeggen. De grote diversiteit die kenmerkend voor zowel het theorie- als praktijkveld van evaluatie is daar niet vreemd aan.

Een handleiding kan echter moeilijk zonder definitie. Vandaar dat we een tweetal definities geselecteerd hebben. Ze reflecteren meteen ook de variëteit die er bestaat: sommige auteurs proberen de essentie te vatten, terwijl andere auteurs in hun omschrijving een reeks van basiskenmerken willen meegeven. Geen van beide types is perfect. De kernachtige bepaling vertelt misschien te weinig, terwijl de aspiratie om volledig te zijn toch net te hoog gegrepen is voor de uitgebreide varianten, die bovendien misschien meer vragen oproepen dan dat ze er beantwoorden. Met de twee onderstaande definities willen we een combinatie van de essentie en enkele voorname kenmerken van beleidsevaluatie benadrukken, die we als relevant en fundamenteel beschouwen. Ze vormen het vertrekpunt voor onze benadering van beleidsevaluatie doorheen de verdere handleiding. Een handleiding staat immers niet los van een bepaalde perceptie, interpretatie en benadering van het 'studieonderwerp'. Beide definities overspannen ook een aanzienlijke tijdsperiode.

De eerste definitie dateert van 1980 en komt van Michael Scriven, die als één van de grondleggers kan worden beschouwd van beleidsevaluatie als theorieveld. Deze definitie geeft de essentie van evaluatie weer.

“Evaluation is the systematic and objective determination of the worth or merit of an object.”

Scriven stelt dat evaluatie de systematische en objectieve bepaling is van de (meer)waarde of verdienste van evaluatieobject of het evaluandum. Daarmee verwijst hij naar wat men kan beschouwen als de essentie van evaluatie: het toekennen van waarde. Verder van belang is de wijze waarop dit gebeurt: systematisch en objectief. Dit veronderstelt een wetenschappelijk verantwoorde methode.

De tweede definitie dateert van 2006 en komt van James McDavid en Laura Hawthorn, twee auteurs die de recente trends in het theorie- en praktijkveld van nabij hebben opgevolgd. Hun definitie omvat een aantal sleutelkenmerken en kritische aandachtspunten op het vlak van beleidsevaluatie.

“Evaluation can be viewed as a structured process that creates and synthesizes information intended to reduce the level of uncertainty for stakeholders about a given program or policy. It is intended to answer questions or test hypotheses, the results of which are then incorporated into the information bases used by those who have a stake in the program or policy.”

Deze uitgebreidere definitie reikt ons indicaties aan over verschillende aspecten:

- het motief om beleid te evalueren en de meerwaarde van evaluatie voor het beleid: onzekerheid verminderen

- het verloop en het resultaat van het evaluatieproces: gestructureerd, creëren en samenvatten van informatie
- de methode van evaluatie: vragen beantwoorden, assumpties testen
- het gebruik van het evaluatieresultaat: de belanghebbenden van het beleid gebruiken de informatie verkregen uit evaluaties naast andere informatie uit andere bronnen

Deze elementen vormen de aanknopingspunten voor de volgende secties.

2.3 Motieven om beleid te evalueren

Achter de groeiende aandacht voor beleidsevaluatie zitten verschillende beweegredenen. In de Europese evaluatiebeweging staan drie zienswijzen op de rol en meerwaarde van evaluatie centraal:

- evaluatie ter verbetering van het management;
- evaluatie ter versterking van de democratische principes;
- evaluatie vanuit een pluralistisch perspectief gericht op het bereiken van overeenstemming tussen verschillende stakeholders betrokken bij het publiek beleid.

Onder deze zienswijzen kan een hele range van motieven en principes geplaatst worden van waaruit belang wordt gehecht aan beleidsevaluatie.

Deze brede waaier kan, zoals gesteld in voorgaande paragraaf, teruggebracht worden tot een kwartet van voornaamste motieven om te evalueren:

1. Ondersteuning van beleidsplanning
2. Opvolging van beleidsuitvoering en bijdragen tot institutionele ontwikkeling van de betrokken organisaties
3. Verantwoording afleggen
4. Beleidsleren

In wat volgt staan we stil bij elk van deze motieven. Ten slotte gaan we in op het belang van het expliciteren van het motief voor een gegeven evaluatie.

1. Ondersteuning van beleidsplanning

Ten eerste kan evaluatie een ondersteunende rol bieden in het kader van de voorbereiding en bepaling van het beleid. Daarbij kan allereerst de beleidsproblematiek beoordeeld worden, bijvoorbeeld op draagwijdte of op dringendheid. De raison d'être van beleid ligt in het antwoord dat het biedt op gepercipieerde noden en problemen. De beleidsrelevantie kan dan ook het voorwerp zijn van een evaluatie vooraleer er beslist wordt tot de uitvoering ervan. Vaak zal men ook de keuze hebben tussen alternatieve beleidsinitiatieven, of varianten van eenzelfde beleidsinitiatief. Evaluatie en onderlinge vergelijking van deze alternatieven kan het keuzeprocess vergemakkelijken en de uiteindelijke beslissing staven. Wanneer beleid in de planningsfase zit, kan het ook geëvalueerd worden op interne coherentie: vormen de doelstellingen en maatregelen een samenhangend geheel. Dezelfde vraag kan ook worden gesteld met een externe blik: is het voorgestelde beleid consistent met initiatieven binnen andere beleidsdomeinen? Zijn er eventuele tegenstrijdigheden of synergieën?

2. Verbeteren van de beleidsuitvoering en bijdragen tot institutionele ontwikkeling van de betrokken organisaties

Een tweede motivatie om te evalueren ligt in de ondersteuning van de uitvoering van het beleid en van het management van de organisaties die daarbij betrokken zijn. Op basis van evaluaties kan beleid dat in uitvoering is nog bijgestuurd worden. De betrokken organisaties kunnen daarenboven hun werking afstemmen en verbeteren in functie van de geformuleerde beleidsdoelstellingen. Evaluatie heeft dan als functie het aanbrengen van alle elementen die beleidsmakers in staat stellen het bestaande beleid te verbeteren. Verbeteringen kunnen ondermeer gericht zijn op een meer rationele distributie van de menselijke en financiële middelen, of een verbeterd dienstenmanagement. Het uiteindelijke doel is de identificatie van succes –en faalfactoren, ter ondersteuning van de interventies en de keuzes van het management tijdens de eigenlijke beleidsimplementatie. De belangrijke meerwaarde ligt dan in het tijdig kunnen inspelen op veranderingen. Vaak zullen bij evaluaties vanuit dit motief processen centraal staan. Verloopt de beleidsuitvoering op een efficiënte manier? Zijn de stakeholders voldoende betrokken bij het implementatieproces? Wordt de tijdslijn gerespecteerd? Zit men op schema om de doelstellingen te halen?

3. Verantwoording afleggen

De verantwoording voor gemaakte beleidskeuzes en, in functie daarvan, de inzet van middelen vormt een derde reden om beleid te evalueren. 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.

Het afleggen van verantwoording kan zowel vanuit democratisch als managementoogpunt worden benaderd. Overheidsorganisaties hebben de plicht om feedback te leveren over de werking van de diensten enerzijds en de resultaten van het beleid anderzijds. De burger is voor een aantal publieke goederen en dienstverleningen geheel afhankelijk van de overheid. Vaak heeft hij geen keuzevrijheid zoals deze bestaat op de private markt. Evaluaties maken het de burger mogelijk inzicht te verwerven in de efficiënte en effectieve werking van de overheid en in de kwaliteit en meerwaarde van de dienstverlening. De feedback van de burger in een representatieve democratie kan georganiseerd worden via een aantal inspraak- en participatiekanalen, die echter een beperkte aanvulling vormen op de algemene goed- of afkeuring van het beleid via periodieke verkiezingen. Alleen vormen van directe democratie via referenda, zoals deze sterk zijn geïnstitutionaliseerd in Zwitserland, bieden de burger een mogelijkheid tot de directe goed- of afkeuring van één enkele beleidsmaatregel. Bijgevolg is het parlement, de provincieraad of de gemeenteraad een belangrijke stakeholder bij beleidsevaluatie vanuit het democratisch verantwoordingsperspectief. Maar ook binnen de uitvoerende macht, regering en administratie, vormt verantwoording een belangrijk en vaak terugkerend evaluatiemotief.

Werden de doelstellingen gehaald? Werden de middelen efficiënt ingezet? In welke mate zijn de effecten of de vastgestelde veranderingen daadwerkelijk toe te schrijven aan het beleid? In welke mate zijn de beleidseffecten duurzaam? Wat is de beleidsimpact op langere termijn?

4. Beleidsleren

Het trekken van lessen uit ervaringen met beleid is een vierde evaluatiemotief. Het uiteindelijke doel van dit kennismotief is een groter inzicht verwerven in de factoren die aan de basis liggen van maatschappelijke problemen en hoe beleid daarop op een adequate wijze kan inspelen. Evaluaties vanuit dit motief trachten het verband te achterhalen tussen problematiek, de beleidslogica en beleidsresultaten. Waarom heeft beleid wel of niet effect? Hoe is het beleid er al dan niet in geslaagd om de beoogde verandering tot stand te brengen? Vanuit de antwoorden op dergelijke vragen kan de overheid lessen trekken over de doeltreffendheid van beleidsinstrumenten. Ze kan rekening houden met contextgebonden factoren die mee bepalend zijn om tot een geslaagd beleid te kunnen komen. Dit is van belang voor de overdracht van beleidsmaatregelen naar andere settings en voor de afweging van alternatieve werkwijzen.

Beleidsleren kan op twee niveaus gebeuren. Bij “één cirkel leren” (single-loop learning) vormen de doelstellingen, waarden en referentiekaders een gegeven. Het accent ligt daarbij op het verbeteren van de instrumenten in functie van de gekozen strategie. Budgettaire middelen worden dan bijvoorbeeld gebundeld voor de meest succesvolle beleidsacties. Bij “dubbel cirkel leren” (double-loop learning) stelt men ook de strategie en referentiekaders in vraag. De logica achter het beleid of de beleidstheorie wordt

dan ook het voorwerp van evaluatie. Dit kan resulteren in veranderingen in de strategische beleidsbeslissingen.

Het motief om beleid te evalueren is een belangrijk vertrekpunt. Zoals Eleanor Chelimsky (1997) stelt: "The purpose of an evaluation, conditions the use that can be expected of it". Met andere woorden: het motief achter een evaluatie bepaalt mee op welke wijze de resultaten en conclusies van deze evaluatie gebruikt zullen kunnen worden. Want afhankelijk van het motief zal het evaluatieproces op een verschillende wijze gestalte krijgen. De vier bovenstaande motieven voor evaluatie leiden tot een andere 'soort' evaluatie: met een ander design, een ander tijdsperspectief en een andere rol voor de evaluator en bijgevolg een andere relatie met de stakeholder die de evaluatie vraagt.

Verder in deze handleiding diepen we de samenhang tussen motieven om te evalueren en types van evaluatie verder uit. Tot slot van deze paragraaf wijzen we de lezer op een beperking die vervat zit in ons overzicht van evaluatiemotieven. De hoger vermelde motieven kunnen worden gekoppeld aan bepaalde maar niet alle mogelijk gebruikswijzen. Balthasar (2006) onderscheidt een viertal vormen van gebruik van evaluaties:

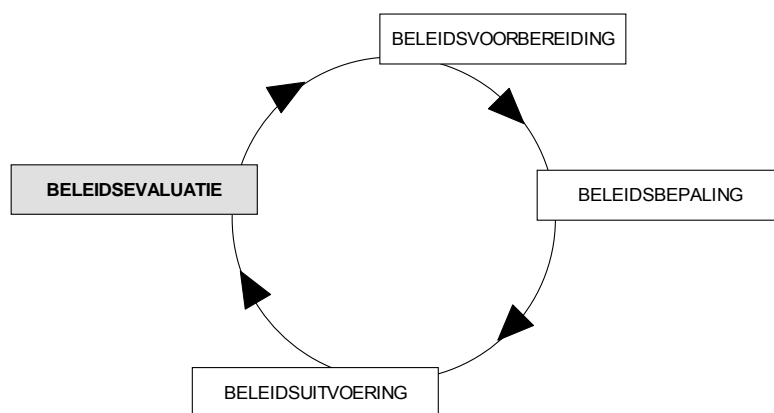
- Instrumenteel gebruik: evaluaties verschaffen een impuls die gebruikt wordt om het beleidsproces te informeren op een aantoonbare wijze. Conclusies van de evaluatie worden rechtstreeks toegepast. Deze gebruikswijze is minder waarschijnlijk wanneer het voorwerp van evaluatie sterk gepolitiseerd is.
- Conceptueel gebruik: evaluatieresultaten beïnvloeden de attitudes en overtuigingen van beleidsmakers met betrekking tot het voorwerp van de evaluatie. Dit is meestal het geval wanneer het om een externe evaluatie gaat.
- Procesgerelateerd gebruik: het uitvoeren van de evaluatie op zich brengt voordelen mee voor de betrokken stakeholders, bijvoorbeeld het creëren van een gemeenschappelijk begrip, de ontwikkeling van een netwerk, of leerprocessen.
- Symbolisch gebruik: de evaluatie wordt gehanteerd om bijkomende legitimatie van een beleidsinitiatief te verschaffen. Verbetering van beleid door middel van de evaluatie wordt daarbij niet beoogd.

De hoger vermelde evaluatiemotieven kunnen voornamelijk aan de eerste drie gebruikswijzen gekoppeld worden. Bij instrumenteel gebruik kan men spreken van *direct of daadwerkelijk gebruik*. Bij conceptueel en procesgerelateerd gebruik is *invloed* wellicht de meest geëigende beschrijving van de doorwerking van een evaluatie.

Symbolisch gebruik bestaat en mag niet genegeerd worden. Inherent aan symbolisch gebruik is dat er vooraf de intentie bestond om de resultaten en conclusies niet te gebruiken. Het uitvoeren van de evaluatie is in dit geval het doel, niet het middel. Het fundamentele verschil met de andere gebruikswijzen is dat de evaluatie in se géén meerwaarde verschaft voor het beleid. Om deze reden gaan we in deze handleiding er ook niet verder op in.

2.4 Positionering van evaluatie in de beleidscyclus en typologieën van beleidsevaluatie

Eén van basisideeën rond beleidsevaluatie waarop deze handleiding steunt, heeft betrekking op de situering van evaluatie in de beleidscirkel of -cyclus. De traditionele voorstelling van de beleidscyclus deelt deze cyclus in een aantal fasen in. Afhankelijk van de auteur verschilt het aantal fasen dat onderscheiden wordt. Zo onderkent men bijvoorbeeld agendavorming, beleidsontwerp, beleidsinvoering, beleidstoepassing, beleidsopvolging en beleidsevaluatie. De meest compacte variant is beperkt tot de fasen van beleidsplanning, beleidsuitvoering en beleidsevaluatie. Gemeenschappelijk element bij deze indelingen is dat beleidsevaluatie steeds als de laatste fase in de beleidscyclus wordt voorgesteld. Dit is weergegeven in onderstaande figuur.



Figuur 2: Het traditionele schema van de beleidscyclus

Het kwartet van evaluatiemotieven dat we uit de theorie en praktijk hebben afgeleid, geeft echter aan dat evaluatie van beleid in verschillende fasen van de beleidscyclus een ondersteunende rol kan spelen. Onze alternatieve opvatting stelt bijgevolg dat beleidsevaluatie moet gedacht worden als een noodzakelijk, nuttig en mogelijk onderdeel of proces doorheen de ganse beleidscirkel. Met andere woorden, evaluatie is niet beperkt tot de laatste fase van beleidscyclus. Om terminologische verwarring te vermijden en de klassieke beperkende associatie van evaluatie met de laatste fase op te heffen, duiden we deze laatste fase verder in de teksten aan met beleidsbeoordeling.

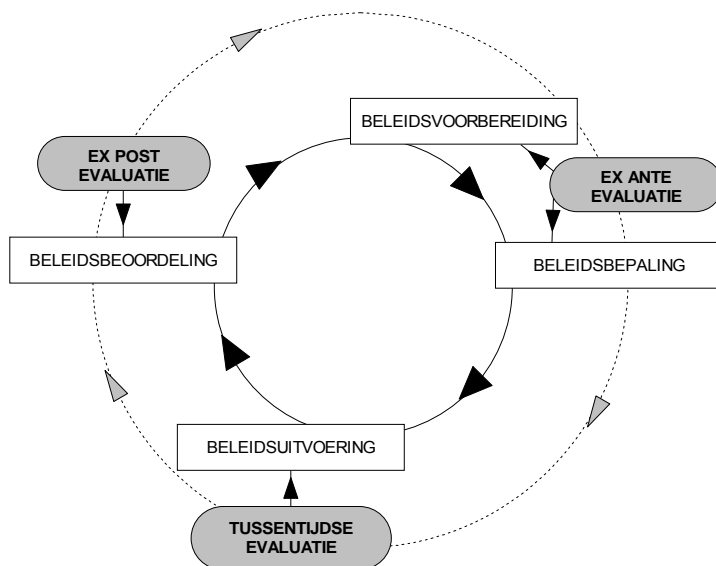
2.4.1 Evaluatietypes vanuit de beleidscyclus

De beleidscyclus is dan ook een belangrijk referentiekader om soorten evaluaties te onderscheiden en een basistypologie op te stellen. Evaluaties die in een bepaalde fase gebeuren kunnen gekoppeld worden aan een tijdsperspectief. Zo kunnen we evaluaties opdelen in ex ante evaluaties, tussentijdse evaluaties en ex post evaluaties. De fase van beleidsuitvoering wordt als spil genomen om het tijdslabel toe te kennen.

Fase in beleidscyclus	Type evaluatie
beleidsvoorbereiding	<i>ex ante</i> evaluatie
beleidsuitvoering	<i>tussentijdse</i> evaluatie
beleidsbeoordeling	<i>ex post</i> evaluatie

Tabel 2: Types van evaluatie gebaseerd op fase in de beleidscyclus

De onderstaande figuur geeft de koppeling van evaluatie met de verschillende elementaire fasen van de beleidscyclus weer.



Figuur 3: Evaluatietypes gesitueerd in de beleidscyclus

De tijdsgebonden typologie hangt verder samen de eerder aangehaalde evaluatiemotieven. De opzet van een evaluatie verschilt namelijk naargelang de beleidsfase waarin de evaluatie wordt uitgevoerd. Het onderstaande schema geeft het verband aan tussen de fasen van de beleidscyclus, het bijhorende type van evaluatie, en deze motieven.

Fase in beleidscyclus	Type evaluatie	Evaluatiemotief
beleidsvoorbereiding	<i>ex ante</i>	Ondersteuning van de beleidsplanning
beleidsuitvoering	<i>tussentijds</i>	Verbeteren van de beleidsuitvoering Bijdragen aan institutionele ontwikkeling
beleidsbeoordeling	<i>ex post</i>	Verantwoording Beleidsleren

Tabel 3: Het verband tussen beleidsfase, evaluatietype en evaluatiemotief

2.4.2 Evaluatietypes op basis van de inhoudelijke focus

Een tweede basistypologie van evaluaties kan worden opgesteld op basis van de algemene inhoudelijke focus van een evaluatie. Daarbij zijn twee categorieën belangrijk:

- evaluatie van de *beleidsinhoud*: het concept, de structuur, de instrumenten, processen en acties van het beleid
- evaluatie van de *beleidsimpact*: de veranderingen die het beleid teweeg brengt in de maatschappelijke omgeving

Beide classificaties kunnen we met elkaar vergelijken in hun relatie met de evaluatiemotieven. Er is geen 1-op-1 relatie mogelijk tussen de types uit beide typologieën. De reden daarvoor is niet alleen het ongelijke aantal categorieën. De inhoudelijke focus van een evaluatie hangt ook niet exclusief samen met een bepaalde fase in de beleidscyclus en bijgevolg een tijdsperspectief. Toch kunnen we een zekere clustering in de onderlinge samenhang aanbrengen. Deze is weergegeven in onderstaand schema.

ex ante	ondersteunen van de beleidsplanning	beleidsinhoud ----- beleidsimpact
tussentijds	verbeteren van de uitvoering institutionele ontwikkeling	
ex post	verantwoording beleidsleren	

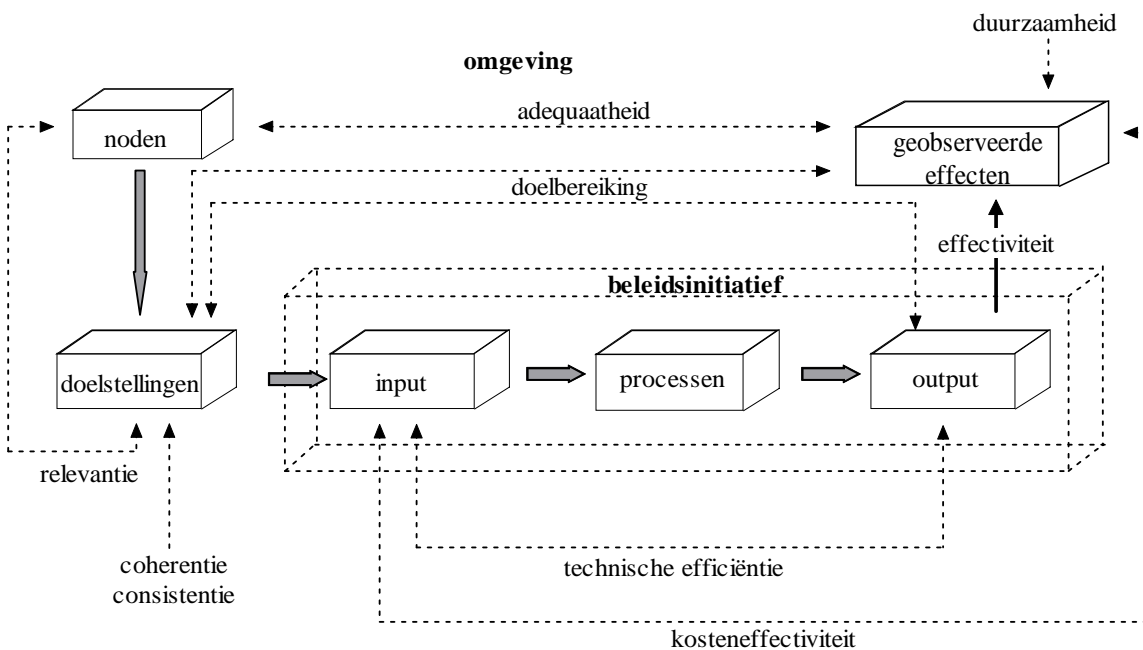
Tabel 4: Clustering van evaluatietypes en motieven

In *ex ante* evaluaties staat doorgaans op de beleidsinhoud centraal, terwijl *ex post* evaluaties de beleidsimpact als focus hebben. Een tussentijdse evaluatie kan zowel de beleidsinhoud, de beleidsimpact of een combinatie van deze twee als voorwerp nemen. Anderzijds kan in een *ex ante* evaluatie ook de (verwachte) impact van het beleid geëvalueerd worden, bijvoorbeeld wanneer beleidsalternatieven onderling worden afgewogen (cf. infra). Ingeval van een *ex post* evaluatie kan ook het uitklaren van de beleidsinhoud en het beschrijven van het beleidsproces een eerste noodzakelijke stap vormen.

2.4.3 Evaluatietypes op basis van het evaluatiecriterium

Een derde basistypologie van beleidsevaluaties is gebaseerd op het evaluatiecriterium dat centraal staat in een evaluatie. Er zijn heel wat evaluatiecriteria denkbaar die op verschillende wijze geïnclassificeerd kunnen worden. We hanteren hiervoor het schema van de beleidscyclus als open systeemmodel. Dit schema is weergegeven in de volgende figuur.

Het open systeemmodel positioneert de beleids- en beheerscyclus van de overheid in hun maatschappelijke omgeving. Vanuit deze omgeving ontstaan er noden met daaraan een gekoppelde beleidsproblematiek. Als antwoord op deze noden formuleert de overheid beleidsdoelstellingen, zowel op strategisch als operationeel niveau. Daarna komt een operationeel proces binnen de overheidsorganisaties op gang waarbij middelen (input) worden aangewend in processen die vervolgens resulteren in bepaalde producten en dienstverlening (output). Het is binnen dat proces dat een beleidsinitiatief concreet gestalte krijgt, binnen de organisatiegrens van de overheid. De output die de overheid aflevert, vormt echter geen eindpunt. Het einddoel is het beïnvloeden van maatschappelijke factoren of, anders gezegd, het creëren van effecten of een verandering.



Figuur 4: Evaluatiecriteria afgeleid uit de beleids- en beheerscyclus als open systeemmodel

Evaluatiecriteria hebben betrekking op enerzijds kenmerken van de componenten uit het open systeemmodel en anderzijds op relaties tussen deze componenten. In het bijzonder zijn in het bovenstaande schema volgende evaluatiecriteria opgenomen:

- beleidsrelevantie
- coherentie en consistentie van beleid

- technische efficiëntie
- doelbereiking
- effectiviteit
- kosteneffectiviteit
- duurzaamheid van effecten
- adequaatheid van effecten

Verderop in dit boekdeel geven we een meer uitgebreide beschrijving van deze evaluatiecriteria. We geven daar ook aan in welke mate elk van deze criteria aan bod kan komen in een ex ante, tussentijdse of ex post evaluatie.

In eenzelfde evaluatie kunnen meerdere evaluatiecriteria centraal staan. Echter, hoe meer criteria, hoe uitgebreider en complexer de evaluatie zal zijn. Bovenstaande lijst is niet exhaustief. Zo kan een evaluatie ook het maatschappelijke draagvlak voor een beleidsinitiatief als voorwerp hebben. De opgesomde criteria zijn generieke criteria. Daarnaast kan een evaluatie ook meer concrete criteria hanteren die bijvoorbeeld betrekking hebben op specifieke kenmerken van een beleidsinitiatief. In het bijzonder is dit het geval wanneer men meerdere beleidsalternatieven tegenover elkaar wil afwegen in een evaluatie.

2.4.4 Interne en externe evaluatie

Een vierde en laatste basisindeling van evaluaties is gebaseerd op de **organisatorische positie van de evaluator of het evaluatieteam**. Afhankelijk van de situering van de evaluator vis-à-vis de opdrachtgever of gebruiker van de evaluatie, kan men spreken van:

- een *interne* evaluatie
- een *externe* evaluatie.

Wanneer de evaluator niet tot de organisatie behoort waarvan het beleidsinitiatief wordt geëvalueerd, spreken we van een externe evaluatie. Wanneer de evaluator wel deel uitmaakt van de betrokken beleidsvoorbereidende of -uitvoerende organisatie, gaat het om een interne evaluatie.

Gegeven de 'roots' van evaluatie (cf. supra) met een nadruk op verantwoordelijkheid heeft men in het theorie- en praktijkveld van evaluatie lange tijd voornamelijk aandacht gehad voor externe evaluatie. Meer recent is op dat vlak de balans meer in evenwicht gekomen tussen interne en externe evaluatie. Deze evolutie houdt verband met de gegroeide diversiteit aan evaluatiemotieven.

Men kan inderdaad stellen dat afhankelijk van het evaluatiemotief, interne dan wel externe evaluatie méér aangewezen is. De redenen waarom dit zo is, variëren echter eveneens naargelang het evaluatiemotief. Ze hebben te maken met de respectieve sterktes en zwaktes van interne en externe evaluatie. Let wel, het gaat niet telkens om de keuze tussen ofwel interne ofwel externe evaluatie. Hybride vormen of het combineren van interne en externe evaluaties kunnen de sterktes en zwaktes van beide types respectievelijk benutten en vermijden.

Onderstaand schema geeft in hoofdlijnen de relatie tussen het evaluatiemotief en de twee types van evaluatie weer. In wat volgt gaan we in op de sterktes en zwaktes van interne en externe evaluatie en we koppelen terug naar de evaluatiemotieven.

Evaluatiemotief	INTERN	EXTERN
<i>Ondersteuning van beleidsplanning</i>		
<i>Verbeteren van de beleidsuitvoering en organisatorische ontwikkeling</i>		
<i>Verantwoording afleggen</i>		
<i>Beleidsleren</i>		

Figuur 5: Het evaluatiemotief en de keuze tussen interne of externe evaluatie

Interne evaluaties hebben als voordeel dat de evaluator doorgaans beschikt over een gedetailleerde en toegankelijke kennis van het evaluandum. De evaluator is dan immers betrokken bij de planning en/of uitvoering van het beleid. Bovendien beschikt een interne evaluator over een hoge contextgevoeligheid: men kent de bestaande structuren, de verhoudingen tussen de personen betrokken bij het beleid, de standpunten van de verschillende stakeholders, de spanningsvelden enz. Interne evaluaties worden als minder dreigend gepercipieerd en zijn vaak opgevat als een constructieve aanpak tot verbetering. Daardoor kan een lange termijn engagement of een evaluatiecultuur gecreëerd worden in de organisatie. Interne evaluaties kunnen daarenboven sneller leiden tot een institutionalisering van een evaluatiefunctie in de organisatie die een breed mandaat krijgt waardoor interne expertise en capaciteit kan worden opgebouwd.

Interne evaluatie heeft als eerste mogelijke zwakte dat de interne evaluator wordt geconfronteerd met rolcomplexiteit. Afhankelijk van hoe de interne evaluatiefunctie concreet is georganiseerd, kan een interne evaluator, naast lid van de organisatie, ook rechtstreeks medeverantwoordelijk zijn voor het beleid. Zo kan er belangenvermenging ontstaan in hoofde van de interne evaluator. Door de permanente financiële afhankelijkheid van de organisatie kan het in bepaalde gevallen moeilijk zijn om de balans te vinden tussen werkethiek en loyaliteit aan het management. Conclusies van interne evaluaties kunnen onder een bepaalde drempel op het vlak van kritische beoordeling blijven steken. Negatieve bevindingen

kunnen systematisch genegeerd of zelfs ‘vermeden’ worden. Bij een interne evaluatiefunctie is het ook de vraag in welke mate er daadwerkelijk tijd gegarandeerd wordt om evaluaties uit te voeren. Voorziet men in een aparte evaluatiefunctie, of werkt men met zogenaamde ‘maandagmorgen evaluatoren’ die naast hun dagdagelijkse functie en bovenop hun standaard takenpakket er plots een evaluatieopdracht bij krijgen? In het laatste geval zal ook de methodische kennis en kwaliteit een zwakte vormen.

Externe evaluatie heeft als voordeel dat de evaluator een hoge graad van onafhankelijkheid heeft, of althans een gepercipieerde grote onafhankelijkheid. Onafhankelijkheid wordt meestal terecht geassocieerd met objectiviteit. Naar de buitenwereld komen externe evaluaties geloofwaardiger over. De facto zal dit afhangen van hoe de relaties tussen opdrachtgever en externe evaluator georganiseerd zijn. De inzet van externe evaluatoren kan een nieuwe wind doen waaien; nieuwe ideeën, alternatieve visies worden in de organisatie binnengebracht. Zij kunnen een helikopterperspectief creëren en fundamentele verbeterpunten aanwijzen. Een externe evaluator is beter geplaatst om negatieve bevindingen te communiceren. Daarenboven kunnen externe evaluatieteams meestal bogen op een gespecialiseerde kennis en vaardigheden op het vlak van specifieke evaluatietechnieken. Ook genieten ze vaak een hogere flexibiliteit op het vlak van inzetbare personeelscapaciteit. Een externe evaluatie heeft ook een financiële transparantie: de evaluatieopdracht wordt op voorhand genegotieerd en het prijskaartje is bekend.

Externe evaluatie heeft ook een aantal mogelijke zwaktes. De toegang tot relevante informatie kan in de praktijk moeizaam verlopen. Bepaalde informatie wordt achter gehouden, of interne conflicten tussen personen of afdelingen maken bruikbare dataverzameling onmogelijk. Daarmee samenhangend kan een externe evaluator ook worden geconfronteerd met minder gelegenheid tot interactie met de beleidsverantwoordelijken en het management. Dergelijke contacten vragen een tijdsinvestering maar zijn belangrijk om een accuraat evaluatiedesign op te stellen. Onvoldoende interactie kan ertoe leiden dat bepaalde stakeholders zich maar moeilijk of niet kunnen identificeren met het evaluatiedesign. De concurrerende markt van externe evaluatoren kan ook disfuncties hebben op het vlak van objectiviteit: een externe evaluator kan met het oog op eventuele vervolgontracten bij dezelfde opdrachtgever de gepercipieerde sterke objectiviteit laten verwateren.

De volgende tabellen geven een overzicht van de sterktes en zwaktes van beide types van evaluatie.

Interne evaluatie

+	-
Gedetailleerde kennis over het evaluandum Toegang tot informatie Contextgevoeligheid Minder gepercipieerd als bedreigend Lange termijn engagement Interne capaciteitsopbouw	Rolcomplexiteit Mogelijke belangenvermenging Afhankelijk van management voor organisatorische uitwerking en mogelijkheden voor methodische kennisopbouw

Figuur 6: Sterktes en zwaktes van interne evaluatie

Externe evaluatie

+	-
Sterke (gepercipieerde) onafhankelijkheid Sterke (gepercipieerde) objectiviteit Aanbreng van vernieuwende perspectieven Specialisatiekennis op methodische vlak Flexibele personeelscapaciteit Financiële transparantie	Toegang tot informatie Minder mogelijkheden tot interactie Disfuncties concurrerende markt

Figuur 7: Sterktes en zwaktes van externe evaluatie

Bovenstaande vergelijking leert dat de keuze tussen interne en externe evaluatie vaak neerkomt op een trade-off: het voordeel van de ene vorm is een nadeel van het andere en vice versa. De voor- en nadelen zijn niet automatisch in alle omstandigheden en in alle organisaties terug te vinden. Heel wat factoren spelen een rol, onder meer de grootte en 'maturiteit' van een organisatie, de organisatiestructuur en -cultuur.

Naargelang het evaluatiemotief kan interne of externe evaluatie meer aangewezen zijn, om uiteenlopende redenen.

Ingeval van evaluaties ter ondersteuning van de beleidsplanning kunnen zowel argumenten voor interne als voor externe evaluatie de doorslag geven. Dit hangt samen het concrete evaluandum. Wanneer een grote dataverzameling onder de doelgroep van het geplande beleid is gewenst, is externe evaluatie eerder aangewezen. Anderzijds kan de noodzaak van gedetailleerde kennis over het beleid de pendel doen doorslaan naar interne evaluatie. Voor het verfijnen van de beleidslogica en het implementatieplan in de planningsfase lijkt een interne toetsing in nauwe samenwerking met de planners de meeste voordelen te bieden.

Voor evaluaties ter verbetering van de beleidsuitvoering en de werking van de organisatie in functie daarvan, zijn contextgevoeligheid en toegang tot informatie belangrijk. Dergelijke evaluaties zullen dus vaak intern worden uitgevoerd. Dit biedt de meeste garantie op het rechtstreeks doorgeven en vertalen van leerpunten naar het management van de organisatie.

Daarentegen zal voor evaluaties in het kader van het verantwoordingsmotief de externe evaluatievorm in meer gevallen als de beste optie worden beschouwd, omwille van de afstandelijkheid van de evaluator en de bijhorende assumptie van een grotere garantie op onafhankelijkheid en objectiviteit.

Met het oog op evaluaties in functie van beleidsleren is de keuze tussen interne en externe evaluatie minder evident en opnieuw afhankelijk van verschillende factoren, onder meer de specificiteit van het evaluandum, de aanwezige interne capaciteit en methodische vereisten.

Uit deze clustering volgt dat interne evaluaties meestal ex ante evaluaties en in mindere mate tussentijdse evaluaties zijn, terwijl externe evaluaties vaak ex post evaluaties zijn. De keuze tussen beide vormen hangt echter steeds af van een combinatie van elementen: het evaluatiemotief en de vooropgestelde gebruikswijze, de beschikbare capaciteit binnen de organisatie die de evaluatie vraagt en de financiële middelen die vrijgemaakt kunnen worden.

In sommige situaties biedt de combinatie van beide vormen voordelen. Het evaluatieproces volledig uitbesteden kan omwille van verschillende redenen zeer onpraktisch blijken. Voorbeeld hiervan is het gebrek aan directe feedback over het lopende beleid. Dit is zeker het geval voor ex ante evaluaties. Toch kunnen, zelfs wanneer de coördinatie en verschillende deelaspecten van een evaluatie intern worden uitgevoerd, bepaalde stappen die gespecialiseerde kennis of een grote capaciteit vereisen toevertrouwd worden aan externen, bijvoorbeeld op het vlak van dataverzameling.

2.4.5 Besluit in verband met evaluatietypologieën

In principe kan elke evaluatie ondergebracht worden bij een categorie binnen elk van de zonet beschreven vier basistypologieën (tijdsperspectief, focus, evaluatiecriterium, positie van de evaluator). De keuze van het soort evaluatie dat men wenst te ondernemen, is dan ook een belangrijk vertrekpunt om diverse inhoudelijke en organisatorische aspecten van de evaluatie uit te werken. Idealiter volgt deze keuze uit het evaluatiemotief en de vooropgestelde gebruikswijze, de gebruikers of het doelpubliek, de evaluatievragen én een vergelijkende inschatting van de uitvoeringsopdracht en de interne evaluatiecapaciteit.

2.5 Het evaluatieproces

Ongeacht het type of de benadering kan men in elk evaluatieproces een viertal terugkerende standaardstappen onderkennen: structurering, dataverzameling, data-analyse en beoordeling.

(1) Structurering van de evaluatie

In deze eerste stap staat het ontwerpen van het evaluatiedesign centraal. Dit omvat twee kerntaken: enerzijds het afbakenen van de inhoudelijke scope van de evaluatie en anderzijds het plannen van de volgende stappen in het evaluatieproces.

Het afbakenen van de inhoudelijke scope van de evaluatie omvat op zijn beurt een beschrijving van het evaluandum en het beantwoorden van de vraag: "Wat wil men weten over het evaluandum". Het antwoord op deze laatste vraag wordt gegeven met de selectie van de concrete evaluatievragen en -criteria.

Het plannen van de volgende stappen in het evaluatieproces komt neer op het maken van keuzes in verband met gegevensbronnen, en technieken om gegevens te verzamelen, te analyseren en om te komen tot conclusies.

(2) Dataverzameling

In deze stap wordt de vereiste informatie verzameld om de analyse en beoordeling van het evaluandum te kunnen maken. Afhankelijk van de evaluatievragen en de kwaliteit en beschikbaarheid van gegevens, kan men beroep doen op bestaande gegevens. Gegevens uit monitoring of meetsystemen vormen vaak relevante informatie. Vaak zal men echter voor een evaluatie bijkomend nog nieuwe gegevens moeten verzamelen.

(3) Data-analyse

In deze stap worden de verzamelde gegevens samengebracht en geanalyseerd in functie van de evaluatievragen en -criteria. Vaak omvat dit een vergelijking van gegevens, in de tijd en/of ruimte. Ook kunnen in deze stap waarden op bepaalde onderzoeksvariabelen geschat of berekend worden. Onderzoeksvariabelen zijn omschreven factoren die verwijzen naar kenmerken, oorzaken of effecten.

(4) Beoordeling

In de stap van beoordeling worden de conclusies van de evaluatie geformuleerd. Deze stap verwijst naar de essentie van evalueren: het toekennen van waarde aan het evaluandum of evaluatieobject. De beoordeling gebeurt door de resultaten van de analyse af te zetten tegen

een bepaald referentiekader. Het referentiekader verwijst naar normen en waarden die vanuit de politiek of het beleid worden aangereikt. Aan elk evaluatiecriterium is een referentiekader verbonden. Het formuleren van conclusies kan in bepaalde gevallen in meerdere stappen verlopen wanneer verschillende criteria als toetssteen worden gehanteerd. In een eerste fase kan men het evaluandum beoordelen volgens elk criterium apart. In een tweede fase tracht men dan tot een eind- of synthesebeoordeling te komen.

Bij het doorlopen van elke stap kunnen één of meerdere technieken of instrumenten ingezet worden. In dat opzicht kan een evaluatie zelden of nooit gebaseerd zijn op één enkele techniek. In boekdeel 3 komt een brede waaier aan technieken aan bod, waarbij voor elke stap in het evaluatieproces een aantal technieken besproken worden. Hoofdstuk 4 van dit eerste boekdeel besteedt uitgebreide aandacht aan de eerste stap in het evaluatieproces: het structureren van de evaluatie of het ontwerpen van een evaluatiedesign.

2.6 Bijzondere vraagstukken

Beleidsevaluatie is, zoals uit het bovenstaande blijkt, een vrij jong theorie- en praktijkveld dat geconfronteerd wordt met een aantal bijzondere vraagstukken. In deze paragraaf belichten we kort een aantal belangrijke uitdagingen: de capaciteitsopbouw, de horizontale en verticale beleidsevaluatie. Voor elk bijzonder vraagstuk schetsen we kort het belang en de problematiek. Deze handleiding wil bijdragen aan de antwoorden op de geschetste uitdagingen.

2.6.1 Capaciteitsopbouw

Evaluatiecapaciteit en evaluatiepraktijk zijn zeer nauw met elkaar verbonden. De evaluatiecapaciteit verwijst naar het menselijk kapitaal, de financiële en materiële bronnen die noodzakelijk zijn voor de uitvoering van een evaluatie. Evaluatie van beleid is in de praktijk enkel mogelijk indien de organisatie over de nodige capaciteit beschikt om evaluaties te plannen, te ontwikkelen en te implementeren of op te volgen. In deze sectie belichten we de noodzakelijkheid van een adequate evaluatiecapaciteit verder toe. Vervolgens behandelen we een drietal spanningsvelden en de kritische succesfactoren voor de opbouw van een evaluatiecapaciteit. Tot slot formuleren we enkele uitdagingen waarvoor organisaties en personen staan tijdens de ontwikkeling van een evaluatiecapaciteit.

De opbouw van evaluatiecapaciteit is noodzakelijk om evaluaties op een kwaliteitsvolle manier uit te voeren. De afwezigheid van een degelijke ondersteuning leidt tot beoordelingen en nieuwe initiatieven op

basis van weinig objectieve en onderbouwde vaststellingen. Tevens kan een organisatie, op dat ogenblik geen afweging maken tussen verschillende argumenten of keuzemogelijkheden. Hiervoor ontbreekt het de organisatie vaak aan voldoende en correcte informatie. Een laatste reden voor de ontwikkeling van een kwaliteitsvol evaluatiesysteem is de vaststelling dat overheidsorganisaties voortdurend de confrontatie aangaan met vragen die een antwoord vereisen. De ontwikkeling van een evaluatiesysteem maakt het beantwoorden van deze vragen mogelijk. Een goed functionerend evaluatiesysteem kan de onzekerheid bij beleidsbeslissingen verlagen. Dit zal op zijn beurt de kwaliteit en de performantie van het overheidsoptreden ten goede komen.

De koppeling van evaluatie met het beslissingsproces is afhankelijk van de mate waarin evaluatie geïnstitutionaliseerd is, ofwel verankerd is in het beleidsproces. Deze institutionele verankering speelt een zeer belangrijke rol in de ontwikkeling van een praktijk en cultuur van beleidsevaluatie.

In de literatuur worden drie belangrijke spanningsvelden aangehaald die zich voordoen bij de opbouw van een evaluatiecapaciteit:

- (1) interne vs. externe capaciteit
- (2) gecentraliseerde vs. gedecentraliseerde capaciteit
- (3) het democratische perspectief vs. het managementperspectief

Het eerste spanningsveld heeft betrekking op de keuze tussen de evaluatie zelf uitvoeren ofwel uitbesteden. Deze keuze is niet altijd eenvoudig te maken; interne en externe evaluaties hebben elk voor- en nadelen (cf. supra). Essentieel is echter dat zelfs indien men ervoor kiest om alle evaluaties uit te besteden, ook deze optie een interne capaciteit vereist. Dergelijke capaciteit neemt dan de vorm aan van een kleine staf die de externe evaluatoren de opdracht geeft, opvolgt tijdens het evaluatieproces en beoordeelt na de uitvoering van de evaluatie. Zoals al gesteld zijn verschillende combinaties van interne en externe evaluatie denkbaar. Het spanningsveld tussen interne en externe capaciteit dient dan ook eerder gezien te worden als een continuüm. De vraag is niet of er een interne capaciteit nodig is, maar wel hoe uitgebreid men de interne capaciteit ontwikkelt en in welke mate men (daarnaast) beroep doet op een externe evaluatiecapaciteit.

Het tweede spanningsveld betreft plaatsing van de (minimale of uitgebreide) evaluatiefunctie binnen de betrokken organisatie. Hierin spelen twee aspecten een belangrijke rol: de opzet van de evaluatie en de geldende randvoorwaarden vanuit wetgeving en andere richtlijnen.

Een organisatie heeft hierbij de keuze uit drie opties: een gecentraliseerde evaluatiefunctie, een gedecentraliseerd systeem of een combinatie van beiden. De initiatiefnemers dienen zich zeer goed bewust te zijn van de verschillende mogelijkheden en beperkingen en de weerslag van hun keuze op de noodzakelijke evaluatiecapaciteit. De oprichting van een centrale dienst is vanuit het standpunt van de capaciteitsopbouw eenvoudiger dan de oprichting van decentrale diensten. Toch kan een combinatie van beide benaderingen in bepaalde omstandigheden aangewezen zijn. Zo blijkt uit de praktijk dat evaluaties gericht op de effectiviteit van beleid beter door gecentraliseerde diensten kunnen worden uitgevoerd. Dit in tegenstelling tot evaluaties gericht op het implementatieproces waarbij een gedecentraliseerde aanpak meer voordelen oplevert.

Het derde en laatste spanningsveld heeft te maken met het globale perspectief van waaruit men beslist om een evaluatiecapaciteit op te bouwen. Zo kan men de functie van evaluatie eerder als versterking van de democratie beschouwen, ofwel ter ondersteuning van beleidsmakers en het management. Evaluaties ter versterking van de democratie leggen de nadruk op een uitbreiding en versterking van de institutionalisering in de wetgevende macht. De klemtoon ligt hierbij op de verantwoording voor de beleidsbeslissingen en de besteding van de publieke middelen. Voorstanders van de versterking van evaluatie als beleidsondersteunend en managementinstrument pleiten voor de verankering in de uitvoerende macht. De nadruk hierbij, ligt op de verbetering van beleidsinitiatieven en de uitwerking van een systeem van beleidsopvolging via monitoringsystemen.

Ook dit spanningsveld is terug te voeren naar de beginvraag: wat is de algemene opzet of beweegredenen van beleidsevaluatie en welke gebruikswijze(n) stelt men voorop? In functie daarvan dient men keuzes te maken over de concrete verankeringspunten van beleidsevaluatie binnen het ruime politiekambtelijke systeem.

Naast een aantal spanningsvelden is de ontwikkeling van de vereiste evaluatiecapaciteit afhankelijk van een aantal kritische succesfactoren.

- Evaluatiecultuur

Een evaluatiecapaciteit kan maar opgebouwd worden vanuit een overtuiging van de meerwaarde die evaluatie voor het beleid kan hebben. Er dienen in de eerste plaats personen te zijn die als trekker kunnen optreden bij de uitbouw van een evaluatiecapaciteit in de organisatie. De verantwoordelijken dienen een duidelijk mandaat te krijgen. De mogelijkheden die deze verantwoordelijken krijgen zal in grote mate bepaald worden door de mate waarin er een fundament van een evaluatiecultuur aanwezig is. Op langere termijn zal er een wisselwerking kunnen ontstaan tussen een werkende evaluatiecapaciteit en een evoluerende evaluatiecultuur. Bij aanvang is echter leiderschap noodzakelijk.

- Informatie- en feedbacksystemen

Evaluaties kunnen maar uitgevoerd worden wanneer de organisatie de nodige gegevens en informatie kan verzamelen. Evaluaties kunnen steunen op diverse informatiebronnen (cf. infra). Een belangrijke soort vormen de informatie- en feedbacksystemen in de organisatie zelf. Monitoring- en rapporteringssystemen en daaraan gekoppelde databanken vormen dan ook mee de ruggengraat voor een performante en duurzame evaluatiecapaciteit. Evaluatie en beleidsmaken zijn, in een optimaal systeem, intrinsiek met elkaar verbonden waarbij via verschillende kanalen de uitwisseling van informatie plaatsvindt.

- Vorming en opleiding

Het plannen, zelf uitvoeren of opvolgen en beoordelen van evaluaties vergt ook een zekere know-how op het vlak van methoden en technieken. Het merendeel van de methoden en technieken die in evaluaties van toepassing zijn vindt zijn oorsprong in de sociale wetenschappen. Het is belangrijk dat beleidsmakers in evaluatiefuncties kennis hebben van welke evaluatietechnieken er bestaan, in welke context ze toegepast kunnen worden en welke mogelijkheden en beperkingen of voor- en nadelen eraan verbonden zijn. Professionele training in evaluatie dan een noodzakelijke bouwsteen voor de ontwikkeling van een interne evaluatiecapaciteit.

- Openheid voor de verschillende evaluatiemotieven

Op de diversiteit van beleidsevaluatie is al herhaaldelijk gewezen. Afhankelijk van de concrete organisatiecontext en achtergronden, zal men bepaalde ideeën hebben over de mogelijke meerwaarde van beleidsevaluatie. Het is belangrijk dat men zicht heeft op de verschillende algemene beweegredenen en de meer concrete evaluatiemotieven en gebruikswijzen, vooraleer men de belangrijke eerste keuzes maakt bij de opbouw van een evaluatiecapaciteit.

Zo wordt evaluatie door velen in de eerste plaats geassocieerd met verantwoording en controle. Dit is echter niet het enige motief om te evalueren. Voor de onderbouwing en verbetering van beleid in de planningsfase en de bijsturing van beleid in uitvoering kan evaluatie eveneens een unieke bijdrage leveren. Hetzelfde geldt voor beleidsleren: het verwerven van inzicht waarom en hoe beleid al dan niet succesvol is. Om een adequate evaluatiecapaciteit uit te bouwen die tegemoet komt aan de noden, dient men voldoende aandacht te besteden al de verschillende mogelijkheden van beleidsevaluatie.

De ontwikkeling van de noodzakelijke evaluatiecapaciteit wordt geconfronteerd met enkele uitdagingen.

- Politisering van de evaluatiecontext

Evaluatie heeft inherent een politieke dimensie. Het gaat immers om het kritisch bekijken van gepland beleid, beleid in uitvoering of uitgevoerd beleid. Evaluatie stelt even wel niet de gepercipieerde noden en behoeften in vraag waarop de overheid wil inspelen. De onderkenning en prioriteitsbepaling daarvan is het exclusieve werkveld van politici. Evaluatie richt zich op de antwoorden die de overheid op een maatschappelijke problematiek formuleert. Bij de uitbouw van een evaluatiecapaciteit is het niet voldoende dat beleidsambtenaren overtuigd zijn van de meerwaarde van evaluatie voor het beleid. Politici zijn een voorname stakeholder bij de ontwikkeling van een evaluatiecultuur en een adequate evaluatiecapaciteit. Het creëren van synergieën tussen de verschillende belangen van stakeholders is een belangrijke voorwaarde in tweeërlei opzicht. Enerzijds om tot een breed gesteunde ontwikkeling van evaluatiecapaciteit te komen om evaluaties te kunnen (laten) uitvoeren. Anderzijds om te komen tot een maximaal gebruik van geproduceerde evaluaties ter ondersteuning van het beleid vanuit de eerder aangehaalde motieven.

- Naar een duurzame evaluatiecapaciteit

Het proces van de capaciteitsopbouw vereist permanent aandacht, bijsturing en investeringen. Expertise en ervaring met beleidsevaluatie opbouwen vraagt noodzakelijk tijd en middelen. Tal van auteurs concluderen dat de opbouw van evaluatiecapaciteit lang en vaak moeizaam proces is. De creatie van een 'evaluatiecultuur' is een proces van lange adem, veranderingen treden slechts zeer traag op. De inspanningen brengen slechts op wanneer ze gericht zijn op het installeren van een duurzame evaluatiecapaciteit. Het creëren van structurele koppelingen van evaluatie met het beleidsproces is dan ook een essentieel aandachtspunt.

- Afstemming tussen informatie- en feedbacksystemen (monitoring) en evaluatie

De strategische koppeling van evaluatie met het beleidsproces kan deels bereikt worden door een andere, meer operationele afstemming. Concrete informatie- en feedbacksystemen dienen afgestemd te worden in functie van terugkerende evaluatievragen. Monitoring vormt een potentieel erg belangrijke informatiebron voor beleidsevaluatie. Het is zelfs een noodzakelijke (maar onvoldoende) voorwaarde om bepaalde evaluatievragen te kunnen beantwoorden, bijvoorbeeld naar de mate van doelbereiking en de effectiviteit van beleid.

2.6.2 Evaluatie van domeinoverschrijdend beleid

Een tweede bijzonder vraagstuk voor beleidsevaluatie vormt de evaluatie van domeinoverschrijdend beleid. De horizontale dimensie van beleid kreeg de laatste jaren meer en meer aandacht van beleidsmakers. Overheden hebben in toenemende mate te maken met beleidsvragen en oplossingen die de klassieke bestuurlijke en administratieve grenzen overstijgen. Brede maatschappelijke tendensen zoals globalisering zorgen voor een toenemende verwevenheid en complexiteit van beleidsproblemen.

Anderzijds wordt de benadering van een probleem vanuit meerdere perspectieven meer en meer als een noodzakelijke kwaliteit van hedendaags bestuur beschouwd.

Horizontal governance of het gezamenlijke optreden van verschillende overheidsorganisaties is een noodzaak om dergelijk horizontaal beleid gestalte te geven. Dit vergt coördinatie om in het 'worst case scenario' tegenstrijdige maatregelen te vermijden. Domeinoverschrijdend beleid bestaat in een aantal varianten, hoewel de termen soms door elkaar worden gebruikt.

Integraal beleid verwijst naar inspanningen binnen verschillende beleidssectoren met het oog op het bereiken van gemeenschappelijke, overkoepelende doelstellingen. Duurzame ontwikkeling is daarvan een voorbeeld. Dit beleid stoelt op synergieën tussen economische, sociale en ecologische doelstellingen.

Transversaal beleid heeft doelstellingen die gericht zijn op een bepaalde doelgroep en waarbij binnen diverse beleidssectoren deze doelgroep specifieke aandacht krijgt. Voorbeelden zijn het gelijke kansenbeleid, jeugdbeleid, seniorenbeleid enz.

De complexiteit van horizontaal beleid heeft ook gevolgen voor de evaluatie ervan. Evalueren van domeinoverschrijdend beleid is niet vanzelfsprekend. Er zijn echter verschillende argumenten om dergelijke evaluaties uit te voeren. Een eerste reden is de vraag naar de algemene effecten van het gevoerde beleid over de verschillende organisatiegrenzen heen. Evaluatie van domeinoverschrijdend beleid kan voor beleidsmakers en burgers een totaalbeeld opleveren van een brede waaier van uiteenlopende maatregelen in functie van een gemeenschappelijk thema of bepaalde doelgroep. Ten tweede kan evaluatie van horizontaal beleid bijdragen tot duidelijkere doelstellingen, een betere agendasetting en een grotere verantwoording. De vraag of het uitgevoerde beleid dezelfde effecten bereikt op meerdere plaatsen of in meerdere contexten. Men probeert dan mogelijke variabelen te identificeren die eventuele verschillen verklaren.

Evalueren van beleid op niveau van een beleidsdomein is complexer dan beleidsevaluatie binnen het beleidsveld. Op analoge wijze is de evaluatie van beleidsdomeinoverschrijdend beleid een stuk gecompliceerder dan de evaluatie op het niveau van het beleidsdomein. We zetten kort enkele uitdagingen en problemen op een rij.

Een eerste uitdaging betreft de verzameling van informatie die toelaat het horizontale beleid te evalueren. De formulering van indicatoren om domeinoverschrijdend beleid op te volgen is niet eenvoudig. Vaak zijn er heel wat indicatoren vereist om het hele beleid af te dekken. De invulling van de indicatoren met meetgegevens wordt niet steeds gerealiseerd wat de evaluatie hypothekeert. Naast de beschikbaarheid van data vormt ook de kwaliteit of vergelijkbaarheid van gegevens vaak een probleem.

Een tweede uitdaging heeft betrekking op de inschatting van de beleidsimpact. Indicatoren leveren enkel informatie over de bruto impact de verschillende maatregelen en niet over hun nettobijdrage. De complexiteit van dergelijk beleid maakt dat men veel moeilijker kan vergelijken met een nulscenario ofwel de verwachte situatie zonder beleid.

Een derde uitdaging ligt in de identificatie van ongewenste neveneffecten veroorzaakt door maatregelen in één beleidsveld die interfereren met het beleid in een ander beleidsveld. Dit is nodig om tegenstrijdheden te verhelpen, bijvoorbeeld door maatregelen aan te passen of bijkomende, zogenaamde flankerende maatregelen te treffen.

Beschikken over voldoende capaciteit is een vierde uitdaging. Evaluatie van horizontaal veronderstelt een zeker helikopterperspectief om te komen tot syntheseconclusies. Deze voorwaarde contrasteert echter met de tendens van specialisme. Daarnaast zijn er vaak organisatorische problemen: is er een mandaat voor een evaluatiefunctie om een overkoepelende evaluatie uit te voeren, of is er voldoende samenwerking tussen de betrokken organisaties om de evaluatie uit te voeren?

Voor de verdere uitwerking van deze problematiek verwijzen we naar deel 4 van deze handleiding. Dit deel bevat een voorbeeld van een evaluatie van domeinoverschrijdend beleid.

2.6.3 Evaluatie van interbestuurlijk beleid

Naast een toenemende horizontale verwevenheid van beleidssectoren, komt ook de verticale dimensie van beleid de laatste decennia meer en meer op de voorgrond. Binnen veel beleidsvelden komt beleid tot stand in een multi-level governance context. De trend naar interbestuurlijk beleid beperkt zich overigens niet tot federale landen. De verdeling van taken binnen de bestuurlijke keten tussen de beleidsniveaus valt echter zelden samen met een duidelijk onderscheid tussen bijvoorbeeld centrale beleidsvoorbereiding en lokale beleidsuitvoering. Meerdere bestuursniveaus delen bevoegdheden en verantwoordelijkheden doorheen de verschillende fasen van de beleidscyclus. Binnen de Europese bestuurlijke context vinden daarvan overvloedige voorbeelden. Voor heel wat thema's heeft het Europese beleid een belangrijke impact op het beleid van de lidstaten. Voor België zien we dan vaak een cascade-effect van de federale overheid naar de Gemeenschappen en Gewesten en verder naar de provincies en gemeenten. De werking van de overheid is gekenmerkt door gedeelde verantwoordelijkheden en kostendeling. Dit keert zelfs terug binnen eenzelfde bestuurslaag, denk bijvoorbeeld aan de departementen en agentschappen binnen de Vlaamse overheid.

Gegeven deze wijdverspreide trend, is evaluatie van beleid dat in een interbestuurlijke context tot stand komt zowel relevant als noodzakelijk zonet onvermijdelijk om lessen te kunnen trekken. Het mag dan ook verwonderen dat dit thema in de literatuur nog maar erg beperkt aan bod is gekomen.

Rieper en Toulemonde hebben interbestuurlijke evaluatie omschreven als een evaluatie waarbij twee of meer bestuursniveaus interageren tijdens minimum één stap binnen het evaluatieproces. De stappen hebben betrekking op het afbakenen van de evaluatie, de dataverzameling en analyse, en het formuleren van conclusies. De auteurs bestudeerden praktijkvoorbeelden uit onder meer Canada, Zwitserland, de Verenigde Staten, het Verenigd Koninkrijk en de Europese Unie. Ze keken daarvoor naar de onderlinge positie van de betrokken bestuursniveaus en de mate van interactie tijdens het evaluatieproces. Ze kwamen tot besluit dat evaluatie in een interbestuurlijke context volgende patronen kan aannemen:

- Partnerschapmodus: de betrokken bestuursniveaus werken op gelijke voet samen doorheen het evaluatieproces, vanuit een partnerschapsrelatie.
- Coöperatieve modus: één bestuurslaag fungeert als trekkers of pivot in de evaluatie terwijl de andere betrokken bestuurslagen een secundaire rol spelen, bv. op vlak van dataverzameling.
- Asymmetrische capaciteit modus: tengevolge van capaciteitsbeperkingen van een lager bestuursniveau zal een hoger bestuursniveau de evaluatie uitvoeren. Het lagere niveau wordt voornamelijk betrokken in de eindfase van het proces, bij de formulering van conclusies.
- Interactiemijdende modus: hoewel verschillende bestuursniveaus op elkaar aangewezen bij de planning en uitvoering van het beleid, werken ze bewust niet samen bij het evalueren van het beleid. Daarmee wordt bijvoorbeeld een debat over bevoegdheidsverdeling vermeden.
- Competitieve modus: onder deze variant voert elke bestuurslaag zelf evaluaties uit over hetzelfde onderwerp, omwille van een conflictsituatie over bepaalde aspecten van het beleid en door een gebrek aan wederzijds vertrouwen. Van interactie tijdens het evaluatieproces is hier helemaal geen sprake.

De verschillende modi tonen ook aan dat beleidsevaluatie in een interbestuurlijke context volgens diverse scenario's kan verkopen. Slechts de eerste drie patronen beantwoorden aan de opgegeven definitie van interbestuurlijke evaluatie.

Interbestuurlijke evaluaties worden echter geconfronteerd met specifieke moeilijkheden. We bespreken kort de voornaamste uitdagingen.

Bestuurslagen hebben een eigen bevoegdheidssfeer, maar ook gedeelde bevoegdheden. Evaluaties kunnen gepercipieerd worden als een bedreiging of hefboom met betrekking tot de bestaande machtsverhoudingen. De mate waarin de evaluatieagenda kan bepaald worden door een gegeven bestuursniveau, zal vaak mee afhangen van diens gewicht in de bevoegdheidsverdeling, de financiële middelen en mogelijke schaalvoordelen om een evaluatiecapaciteit uit te bouwen.

Ten tweede dienen bestuurlijke niveaus verantwoording af te leggen tegenover een eigen electoraat. Verschillende bestuurslagen kunnen legitieme maar tegenstrijdige doelstellingen nastreven in eenzelfde beleidsveld. Dit maakt het moeilijk om een gemeenschappelijke focus in een evaluatie te bereiken.

Ten derde kan het bereiken van consensus rond de gemeenschappelijke effecten van de beleidsinspanningen problematisch zijn. Door de complexiteit en lengte van de interbestuurlijke keten is

het vaak zeer moeilijk om outputs van effecten te onderscheiden. Zo kan een centrale bestuurslaag op een grote hoogte boven het 'straatniveau' een gedragsverandering van een decentrale bestuurslaag als een effect van het eigen beleid beschouwen. Het decentrale bestuursniveau zal echter, vanuit zijn perspectief, beleidseffecten in de maatschappelijke omgeving definiëren. Zo kan bijvoorbeeld, in het kader van subsidieregeling ter ondersteuning van de dienstverlening in lokale besturen, de centrale overheid het succes van het beleidsinstrument afwegen aan het aantal gemeenten dat erop inschrijft. De gemeenten zelf zullen echter andere parameters voor succes hanteren, bijvoorbeeld in termen van klantentevredenheid over de lokale dienstverlening.

Wat is de specifieke bijdrage van elke bestuurslaag aan het beleidsresultaat? Het ontwarren van interbestuurlijk beleid en het toewijzen van effecten aan de ene of andere bestuurslaag is bijzonder moeilijk. Het gaat hier niet louter om een technisch probleem; de rol van elk betrokken bestuursniveau is immers bepaald vanuit een specifieke rationaliteit. Dezelfde evaluatiemethode leidt daarom niet noodzakelijk tot dezelfde conclusies voor meerdere bestuurslagen. Het in rekening brengen van mogelijk tegenstrijdige elementen in de beleidsinspanningen van diverse bestuurslagen is een bijkomende uitdaging.

Een vijfde uitdaging ligt in het definiëren van gemeenschappelijke ambitieniveaus wanneer er consensus bestaat over de evaluatievragen, -criteria en indicatoren. Dergelijke consensus is echter noodzakelijk om te kunnen komen tot syntheseconclusies over de beleidsprestaties.

Een meer technische uitdaging ligt op het vlak van de verzameling en het beheer van data. De evaluatie van interbestuurlijk beleid vindt vaak plaats op meerdere locaties en settings. Een ééndrachtige en kwaliteitsvolle verzameling van vergelijkbare data is dan een cruciaal aandachtspunt. Tevens varieert het niveau waarop gegevens geregistreerd en verzameld worden. Dit heeft op zijn beurt gevolgen voor de mogelijkheden op het vlak van aggregatie. Hoe lager de detailgraad van de registratie, des te meer mogelijkheden zijn er om te aggregeren. Een centrale bestuurslaag die een subsidiesysteem uitwerkt voor lagere bestuurslagen vraagt vaak feedback betreffende de besteding van de middelen.

De problematiek van het evalueren van interbestuurlijk beleid komt in boek 4 van deze handleiding nog verder aan bod.

Literatuur

- Alkin M.C. (ed.) (2004). *Evaluation roots : tracing theorists' views and influences*. Thousand Oaks: Sage.
- Balthasar A. (2006). "The effects of the institutional design on the utilization of evaluation: Evidenced using qualitative comparative analysis (QCA)"in: *The International Journal of Theory, Research and Practice*, 12 (3).

- Chelimsky E. and Shadish W.R. (eds.) (1997). *Evaluation for the 21st Century – A Handbook*. Thousand Oaks CA: Sage.
- Furubo J.-E., Rist R.C. and Sandahl R. (eds.) (2002). *International Atlas of Evaluation*, New Jersey: Transaction Publishers.
- Mathison S. (ed.) (2005). *Encyclopedia of Evaluation*, Thousand Oaks (Ca): Sage.
- McDavid J.C. & Hawthorn L.R.L. (2005). *Program Evaluation and Performance Measurement, an Introduction to Practice*, Thousand Oaks, CA: Sage.
- Rieper O. and Toulemonde J. (eds.) (1997). *Politics and Practices of Intergovernmental Evaluation*. London: Transaction Publishers.
- Scriven M. (1980). *The Logic of Evaluation*, Inverness CA: Edgepress.
- Worthen B.R., Sanders J.R. and Fitzpatrick J.L. (2003). *Program Evaluation – Alternative Approaches and Practical Guidelines*. New York: Addison Wesley Longman.

3. Evaluatie in België en Vlaanderen

3.1 Impulsen

Beleidsvaluatie in België en Vlaanderen is als praktijkveld nog vrij jong. In de Europese context dient men België te positioneren in het tweede peloton van landen die eind jaren '90 een evaluatiecultuur -en praktijk begonnen te ontwikkelen. Men kan zelfs stellen dat voor België het momentum naar een versterkte aandacht voor beleidsvaluatie er kwam bij de uitdijning van deze tweede golf waarmee de evaluatiecultuur zich in Europa verspreidde. Men kan een viertal impulsen onderscheiden waardoor beleidsvaluatie bij de binnenlandse overheden meer aandacht kreeg.

Een eerste impuls kwam er met de verspreiding van het ideeëngoed rond New Public Management. In het bijzonder het verantwoordingsmotief om te evalueren werd daarbij naar voor geschoven. Evaluatie werd benaderd vanuit een managementperspectief of sterk gekoppeld aan het organisatieniveau.

Een tweede belangrijke impuls kwam er vanuit de Europese Unie. Tal van Europese beleidsprogramma's en initiatieven vereisen een actieve participatie van verschillende bestuursniveaus in de lidstaten. De Europese Unie heeft een sterke evaluatiefunctie en -cultuur ontwikkeld binnen de Commissie, die verder uitstraalt op de binnenlandse overheden. De Europese Commissie is vaak in de rol van beleidsinitiator en geldschieter dan ook zeer sterk geïnteresseerd in de goede besteding van de middelen en de mate van doelbereiking. Dit mechanisme heeft ervoor gezorgd dat de lidstaten moeten meestappen in het evaluatiegebeuren. In Vlaanderen hebben heel wat overheden en administraties praktijkervaring met evaluaties voor Europa, onder meer in de sectoren milieu en natuur, plattelandontwikkeling, landbouw en visserij, werkgelegenheid, regionale ontwikkeling.

Een derde impuls komt er via de talrijke hervormingsprocessen die momenteel bestaan in de verschillende bestuurslagen in België. Varone & Jacob (2003) wezen op het belang van bijkomende inspanningen op het vlak van organisatorische inbedding en ontwikkeling van evaluatie. Het zijn niet alleen voorwaarden om op een kwaliteitsvolle en verantwoorde wijze evaluaties te kunnen uitvoeren. Ook de politieke en publieke zichtbaarheid van evaluatieresultaten en eraan gekoppelde beleidsverbeteringen vergroot.

De verdere professionalisering en modernisering van de overheden wordt mee gestuurd door een overtuiging in de rol van de overheid ten behoeve van de samenleving. Overheden moeten echter met de schaarse beschikbare middelen meer en meer de confrontatie aangaan met steeds ingewikkeldere problemen. De keuze voor een zogenaamd "evidence-based policy" is daarvoor het gepaste antwoord: beleid moet geïnformeerd en onderbouwd worden. Evaluaties leveren een belangrijke bijdrage aan het informeren van beleid, het formuleren van adequate antwoorden die efficiënt en effectief zijn.

3.2 Enkele trends en uitdagingen

- Wettelijke verankering in de beleidscyclus

De versterkte aandacht voor beleidsevaluatie keert terug binnen alle bestuurslagen in België en Vlaanderen. Een aandachtspunt daarbij is de koppeling van evaluatie met de recurrente beleidscyclus. Vlaanderen heeft een zekere traditie om beleidsinitiatieven en beleidsinstrumenten te verankeren in wetgeving. In recente kaderwetgeving vinden we dan ook verschillende aanwijzingen terug om evaluatie te verankeren in de beleidscyclus.

Voor Vlaanderen is de hervorming *Beter Bestuurlijk Beleid* daartoe een aanleiding geweest. Het kaderdecreet belast de departementen met beleidsondersteunende taken. De Memorie van Toelichting (art.4) vermeldt daarbij de evaluatie van de beleidsuitvoering, in het bijzonder het nagaan van de effectiviteit van de ingezette instrumenten, en de relaties tussen output en effecten. De agentschappen dienen als input voor de beleidsevaluatie relevante beleids- en beheersinformatie te genereren.

Het uitvoeringsbesluit bij het kaderdecreet omschrijft de taakstelling van de departementen nader op dit vlak. Ze dienen de minister te ondersteunen:

“1° bij het uitwerken van diens beleid: daartoe heeft het departement een beleidsvoorbereidende en beleidsevaluerende opdracht

2° bij de aansturing en opvolging van de beleidsuitvoering, die in de regel toevertrouwd wordt aan agentschappen.”

Op het vlak van beleidsvoorbereiding en -evaluatie hebben departementen de volgende activiteiten als taak ten behoeve van de minister:

- “a) de permanente monitoring en omgevingsanalyse van het beleidsdomein (met inbegrip van het beheer van databanken en de verzameling en analyse van statistische informatie);
- b) de evaluatie op macroniveau van de beleidsuitvoering (ingezette instrumenten, effecten, enz.) met het oog op eventuele bijsturing van het beleid of bijsturing van de aansturing van agentschappen;
- c) het volgen van de ontwikkelingen op het internationale niveau;
- d) de aansturing van het beleidsgericht wetenschappelijk onderzoek en de aanwending van de resultaten van dat onderzoek;
- e) het ontwikkelen van een beleidsvoorbereidend instrumentarium (modellen, scenarioanalyse, benchmarking, enz.);
- f) het voorbereiden en het opstellen van ontwerpen van regelgeving;
- g) het opbouwen van netwerken (intern met de andere actoren van het beleidsdomein of met andere beleidsdomeinen, extern met andere actoren);
- h) het uitwerken van voorstellen met het oog op de beleidsbepaling (welke instrumenten, benodigde middelen, financieringsmechanismen, periodiek benodigde beleids- en beheersinformatie, rapporteringen, verantwoordings- en toezichtsmechanismen) en de toetsing van beleidsvoornemens;

- i) het opstellen van ontwerpen van beleidsnota's en beleidsbrieven, met inbegrip van de begrotingsvoorstellen;
- j) de coördinatie en integratie van de begroting en het informatiemanagement voor het beleidsdomein;
- k) het verlenen van advies met het oog op de beleidscoördinatie en -afstemming;
- l) de beleidsmatige informatie en communicatie en het woordvoerderschap;
- m) het secretariaat van de beleidsraad.”

Het comptabiliteitsdecreet (art.6 §3) vermeldt de beleidseffectenrapportage: een rapportering waarin de realisaties van de beleidsopties gedurende het vorige begrotingsjaar geëvalueerd worden aan de hand van indicatoren en kengetallen die gekoppeld zijn aan de strategische doelstellingen (maatschappelijke effecten) en operationele doelstellingen (beleidsprestaties of output) en die vermeld staan in de beleidsnota's.

Recent deed het Instituut voor de Overheid (KU Leuven) een screening van de Vlaamse beleidsnota's voor de bestuursperiode 2004-2009 op de mate waarin en de wijze waarop de basisprincipes van *Beter Bestuurlijk Beleid* werden doorvertaald naar deze beleidsdocumenten. Uit de conclusies kwam onder meer naar voor dat er een algemene behoefte en streven bestond naar een zogenaamd “evidence-based” beleid. De meeste beleidsnota's bevatten ook verschillende verwijzingen naar geplande beleidsevaluaties en/of de uitbouw van evaluatie-instrumenten.

Ook voor de lokale besturen vinden we, gelijktijdig met een versterkte aandacht voor strategische planning, ook referenties naar beleidsevaluatie terug in het nieuwe Gemeentedecreet en Provinciedecreet. Zo hebben de gemeentesecretaris en provinciegriffier onder meer verantwoordelijkheden op het vlak van de evaluatie van het beleid, en moet het managementteam de coördinatie van de diensten ondersteunen bij de beleidsevaluatie. Op het moment van het schrijven van deze handleiding kon nog niet beoordeeld worden in welke mate en op welke manier deze decretale aanzet voor beleidsevaluatie zich doorvertaalt naar de evaluatiepraktijk binnen het lokale en provinciale bestuursniveau.

- Institutionalisering

Versterkte aandacht voor beleidsevaluatie doet overheden ook nadenken over de organisatie van een evaluatiefunctie. In verschillende administraties en afdelingen binnen de federale en Vlaamse overheid, maar ook binnen een aantal steden en gemeenten, werd de laatste jaren werk gemaakt van de institutionalisering van monitoring en beleidsevaluatie door de oprichting van een gespecialiseerde cel of

task force. Gegeven de evolutie naar een sterkere wettelijke verankering van beleidsevaluatie kan men aannemen dat deze trend van institutionalisering zich in de toekomst nog verder zal doorzetten in het kader van een capaciteitsopbouw om evaluaties te plannen, uit te voeren of op te volgen, te beoordelen en te verwerken in de beleidsprocessen.

- Afstemming van capaciteit

Een uitdaging die met het voorgaande punt samenhangt, ligt in de afstemming van evaluatiecapaciteit tussen verschillende overheidsorganisaties enerzijds en tussen de interne capaciteit binnen de overheid en externe evaluatieteams anderzijds. Het onderzoek van Varone & Jacob (2003) toonde bijvoorbeeld aan dat op het federale niveau in al heel wat evaluaties plaatsvinden, maar op een onduidelijk gestructureerde en niet gecoördineerde manier. Er bestaat met andere woorden nog een progressiemarge naar een betere kwaliteit van evaluaties en een grotere impact van beleidsaanbevelingen.

- Aandacht voor de bijzondere vraagstukken

In het bijzonder op het federaal en regionaal vlak is er een groeiende aandacht in het politieke en ambtelijke discours voor beleidsdomeinoverschrijdende effecten en impact van beleid. Het merendeel van de beleidsnota's van Vlaamse regering verwijst naar effectenrapportering op het niveau van het beleidsdomein. In verschillende beleidsdomeinen bestaan dergelijke rapporteringen al, zoals mobiliteitseffectenrapportage, milieueffectenrapportage, de gezinsindex, en dergelijke meer. De aandacht voor de evaluatie van domeinoverschrijdend beleid is echter niet meteen terug te vinden in de hoger aangehaalde decreten.

Vershillende uitdagingen manifesteren zich op het vlak van monitoring en evaluatie: van begripsuitklaring over de ontwikkeling van aangepaste indicatoren, evaluatiemechanismen en - technieken tot de aggregatie van de verzamelde informatie en de verwerking hiervan tot duidelijke syntheseconclusies en gekoppelde beleidsaanbevelingen.

Ook de evaluatie van interbestuurlijk beleid krijgt meer aandacht. Zo stelt de het agentschap Binnenlands Bestuur een jaarlijks rapport (jaarbeeld) op waarin wordt teruggekoppeld naar de lokale besturen met beleidsrelevante informatie die de centrale overheid haalt uit rapporteringen vanuit de lokale besturen in het kader van het administratieve toezicht. Op termijn kan gedacht worden aan een breder monitoringsysteem voor de opvolging van lokale besturen, dat zowel functioneel is voor de centrale en lokale overheid in de context van interbestuurlijke relaties binnen diverse beleidssectoren (De Peuter en Bouckaert, 2004).

Literatuur

- Conings V., Sterck M., Van Dooren W. en Bouckaert G. (2005). *Beleidsnota's versus Beter Bestuurlijk Beleid. Een toetsing van de beleidsnota's aan het Comptabiliteitsdecreet en het Kaderdecreet Bestuurlijk Beleid*. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen, 102 p.
- De Peuter B. en Bouckaert G. (2004). *Een monitoringsysteem voor lokale bestuurskracht: naar de modellering van een Vlaams design*. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen, 111 p.
- Jacob S. et Varone F. (2003). *Evaluer l'action publique: état des lieux et perspectives en Belgique, Série Modernisation de l'Administration*, Gent: Academia Press, 244 p.
- Vlaams Parlement (07-05-2004). *Decreet houdende regeling van de begrotingen, de boekhouding, de controle inzake subsidies en de controle door het Rekenhof ('Comptabiliteitsdecreet')*.
- Vlaams Parlement (09-12-2005). *Decreet Provinciedecreet*.
- Vlaams Parlement (15-07-2005). *Decreet Gemeentedecreet*.
- Vlaams Parlement (18-07-2003). *Decreet Kaderdecreet Beter Bestuurlijk Beleid*.
- Vlaamse Regering (03/06/2005). *Besluit van de Vlaamse Regering met betrekking tot de organisatie van de Vlaamse administratie*.

4. Ontwerpen en plannen van een evaluatie

4.0 Leeswijzer

Dit hoofdstuk is gewijd aan de eerste standaardstap van elk evaluatieproces: het structureren en plannen van een evaluatie. Dat we in deze handleiding sterk de nadruk op deze standaardstap leggen, heeft een aantal gegronde redenen. Het is de eerste, noodzakelijke stap, zowel wanneer een evaluatie intern zal worden uitgevoerd, als wanneer ze zal worden uitbesteed. Ten tweede is het een essentiële stap doordat de uitkomst ervan een belangrijke impact heeft op alle verdere standaardstappen in het evaluatieproces: dataverzameling, data-analyse en de beoordeling ofwel de formulering van conclusies. Deze stap, die we ook kunnen omschrijven met de term *evaluatiedesign*, omvat twee kerntaken:

- het bepalen van de inhoudelijke scope van de evaluatie
- het plannen van de volgende standaardstappen in het evaluatieproces, meer bepaald de uitvoering van de evaluatie

Paragraaf 1 en 2 overlopen de voornaamste deelstappen en aandachtspunten met betrekking tot deze twee kerntaken. In paragraaf 3 leggen we de link met de context waarin men beslist om de evaluatie uit te besteden.

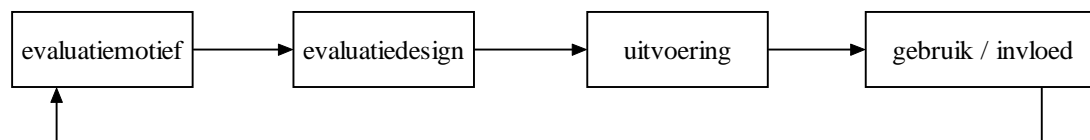
De eigenlijke uitvoering van de andere standaardstappen in het evaluatieproces wordt niet behandeld in dit hoofdstuk. Deel 2 van de handleiding gaat echter nader in op monitoring als ondersteunende functie voor evaluaties. Monitoring vormt namelijk een belangrijke bron voor dataverzameling. We verwijzen de lezer naar dat boekdeel van de handleiding voor informatie over de relatie tussen monitoring en evaluatie, een stappenplan voor de opzet van monitoringsystemen, en het gebruik van monitoring. Deel 3 van de handleiding bespreekt een hele reeks van concrete technieken die ingezet kunnen worden tijdens de standaardstappen van dataverzameling, analyse en beoordeling. We verwijzen de lezer naar dat boekdeel van de handleiding voor informatie over de gebruikscontext, de praktische toepassing en de mogelijkheden en beperkingen van deze technieken.

4.1 Het bepalen van de scope van een evaluatie

Het bepalen van de scope van een evaluatie vormt de eerste kerntaak bij het evaluatiedesign. De kwaliteit van het resultaat en dus van evaluatiedesign is om diverse redenen belangrijk.

Vanuit praktisch oogpunt zijn er de beperkingen van budget en tijd. Net als voor de beleidsvoering zijn ook de financiële middelen voor evaluatie niet onbeperkt. Men zal dus een selectie in mogelijke evaluatieonderwerpen moeten maken. Het budget is doorgaans vooraf bepaald aan de hand van een maximum enveloppe. Binnen deze enveloppe kan de interne uitvoering ofwel de uitbesteding gefinancierd worden. De concrete kostprijs zal afhangen van meerdere factoren (cf. infra). De tijdsbeperking is van belang in relatie tot het beslissingsproces. Hoe breder de scope van de evaluatie, hoe meer tijd nodig is om de evaluatie uit te voeren en hoe langer het duurt vooraleer de resultaten en conclusies gekend zijn. Beleidsmakers staan echter vaak onder tijdsdruk om beslissingen te nemen. Wanneer men evaluaties wil gebruiken ter ondersteuning en onderbouwing van deze beslissingen, dient men de timing en tijdsduur af te stemmen op het beleidsproces. Het afbakenen van de evaluatiescope is ook vereist om een evaluatie beheersbaar te houden: zowel voor de evaluator die de evaluatie uitvoert als voor de gebruiker. Een overdaad aan informatie verlaagt de kans om te komen tot heldere conclusies. Hoe vager de conclusies van een evaluatie, hoe minder waarschijnlijk het wordt dat de evaluatie gebruikt zal worden of invloed zal kunnen hebben in het kader van beleidsbeslissingen.

Vanuit beleidsmatig oogpunt is de (al dan niet opgelegde) beslissing om te evalueren steeds af te leiden van een bepaald evaluatiemotief en een daarmee samenhangende vooropgestelde gebruikswijze (cf. supra). Onderstaand schema toont nogmaals de logische sequens tussen het evaluatiemotief, het design, de uitvoering en het gebruik van een evaluatie.



Figuur 8: Logische sequens in een evaluatieproces

Een duidelijke definiëring van de evaluatiescope speelt een belangrijke rol voor de legitimiteit van de evaluatiebevindingen. Wanneer bij aanvang de scope duidelijk én in consensus wordt afgelijnd, weten alle betrokken partijen wat ze van de evaluatie kunnen en mogen verwachten.

Het afbakenen van de scope van een evaluatie behelst twee aspecten:

- (1) Een accurate definiëring van het te evalueren onderwerp of evaluandum
- (2) Een gedetailleerd antwoord op de vraag: “Wat wil men weten over het evaluandum?”

4.1.1 Definiëren van het evaluandum

Het eerste aspect is een accuraat definiëren van het evaluandum. Daarbij moet men aandacht hebben voor volgende elementen of kenmerken:

- Het institutionele kader

Wat is het institutionele kader van het beleid? Dit kenmerk verwijst naar de structurele plaats en aard van de betrokken beleidsactoren.

Welke bestuurlaag (lokaal, provinciaal, regionaal, federaal, supranationaal) is betrokken bij het te evalueren (deelaspect van het) beleid? Welke type overheidsorganisatie is verantwoordelijk bij de planning en/of uitvoering? Gaat het om kernadministraties (bv. een gemeentelijke dienst, departement, federale overheidsdienst) of verzelfstandigde satellieten (bv. een autonoom gemeentebedrijf, intern of extern verzelfstandigd agentschap, parastatale instelling), of een samenwerkingsverband (bv. een publiekprivaat samenwerkingsverband (PPS), projectvereniging, interlokale vereniging, dienstverlenende vereniging, opdrachthoudende vereniging)?

- Beleidsinhoud en context

Waaruit bestaat het te evalueren beleid of deelaspect in concreto? Dit kenmerk verwijst enerzijds naar de specifieke inhoud van het beleid en het operationele proces van de dienstverlening of uitvoering van de beleidstaak. De beleidsdoelstellingen vormen een essentieel onderdeel van deze informatie. Anderzijds verwijst het naar de beleidssetting: welke actoren zijn er betrokken? Deze informatie slaat hier niet enkel op de betrokken overheidsdiensten, maar ook over de doelgroep van het beleid, een beschrijving van het ruimere beleidsveld. Een beschrijving van de beleidsinhoud en context wordt best in één samenhangend en structureel geheel weergegeven.

- Tijdsdimensie

Het te evalueren beleid dient ook in de tijd afgebakend te worden. Welke voorbije, lopende of toekomstige tijdsperiode is relevant om in aanmerking te nemen voor de evaluatie?

- Geografische dimensie

De ruimtelijke afbakening is het vierde element. Voor de evaluatie valt ruimtelijke scope overigens niet noodzakelijk samen met het ganse toepassingsgebied van het beleid. Men kan ook een selectie maken van regio's of locaties die aan bepaalde voorwaarden of selectiecriteria beantwoorden. Dergelijke selectie kan bijvoorbeeld verband houden met de ernst of omvang van de beleidsproblematiek, met het proportionele aandeel van de specifieke doelgroep in de ruimere bevolking, of met een pilootproject waarbij een alternatieve aanpak ten aanzien van het bestaande beleid wordt getest. De ruimtelijke afbakening zal met andere woorden vaak mee afhangen van het antwoord op de vraag wat men wil weten over het evaluandum.

4.1.2 Het antwoord op de vraag: “Wat willen we weten over het evaluandum?”

Beleidsmakers dienen niet enkel keuzes te maken in wat ze willen evalueren. Voor een gegeven evaluandum (een beleidsplan, programma, project, een aparte maatregel, een beleidsinstrument of deelproces van het beleid enz.) moet men vervolgens kiezen welke concrete aandachtspunten onderzocht moeten worden.

Vandaar het tweede aspect bij het bepalen de scope van een evaluatie: het antwoord op de vraag wat men wil weten over het evaluandum. Daartoe zijn twee deelstappen en elementen van belang:

- het identificeren van stakeholders
- het selecteren van evaluatievragen en -criteria

In de volgende secties gaan we verder in op deze deelstappen.

4.1.3 identificeren van stakeholders

Belangrijk is het lokaliseren van de vraag naar en interesse voor een beleidsevaluatie. Daarvoor moet men de zogenaamde stakeholders identificeren. Stakeholders zijn de actoren die een individueel of collectief belang hebben bij het beleid en bijgevolg bij de evaluatie. Zij vormen dan ook de actoren die het antwoord kunnen geven op de vraag waarop de evaluatie moet focussen.

Wanneer men de stakeholders oplijst, kunnen hun verwachtingen, opmerkingen en gevoeligheden in rekening worden gebracht bij het bepalen van de scope van een evaluatie. Er zijn echter diverse categorieën van stakeholders denkbaar:

- de beleidsmakers (politiek verantwoordelijken en topmanagement)
- de uitvoerders van het beleid (lijnmanagement)
- de doelgroep van het beleid
- derden die rechtstreeks of onrechtstreeks voor- of nadeel ondervinden van het beleid, zonder dat ze tot de beoogde doelgroep behoren
- derden die noch tot de doelgroep behoren noch voor- of nadeel van het beleid ondervinden, maar een algemene interesse in het beleid (en de evaluatie daarvan) betonen

Een voorbeeld kan dit illustreren. Stel dat een gemeente een sluitingsuur heeft vastgesteld voor horecazaken op haar grondgebied en beslist om een tussentijdse evaluatie uit te voeren. Volgende stakeholders kunnen geïdentificeerd worden:

- de gemeenteraad en het schepencollege, i.c. de burgemeester
- de lokale politie
- de uitbaters van de horecazaken
- de buurtbewoners en de klanten
- het bestuur van een omliggende gemeente, de lokale media

Voor het bepalen van de scope van een evaluatie is het identificeren van de verschillende types van stakeholders slechts een tussenstap. In welke mate betreft men stakeholders bij het evaluatieproces?

Doorheen het brede spectrum van evaluatiebenaderingen werden sterk uiteenlopende visies ontwikkeld over de rol van stakeholders in het evaluatieproces, van minimalistisch tot maximalistisch. Dergelijke visies hangen meestal samen een bredere kijk op wat de meerwaarde van evaluatie omvat en op de daaraan beantwoordende evaluatiepraktijk.

Michael Scriven is de bedenker van de “goal-free evaluation”. Hij stelt dat evaluatie een wetenschappelijk proces is waarbij de vraag centraal staat in welke mate de vastgestelde effecten van het beleid adequate antwoorden vormen op de beleidsproblematiek en de noden van de doelgroep van het beleid. Het radicale aspect van deze benadering zit in de stelling dat beleidsdoelstellingen en ambitieniveaus van beleidsmakers daarbij van secundair belang zijn.

David Fetterman is de ontwerper van de “empowerment evaluation” methode. Hij vertrekt van de grondgedachte dat het evaluatieproces een vehikel is om de positie van bepaalde groepen in de maatschappij te versterken. Binnen deze benadering heeft de evaluator eerder een rol van begeleider van de betrokken stakeholders die de evaluatie grotendeels zelf uitvoeren.

Verschillende scenario's zijn denkbaar. Bovenstaande benaderingen zijn eerder polen van een continuüm. De initiële vraag kan verder verfijnd worden: *In welke mate betreft men de verschillende categorieën van stakeholders doorheen het evaluatieproces?*

De “utilization-focused evaluation” methode van Michael Quinn Patton vormt een mainstream benadering van stakeholders. Vertrekpunt bij de aanpak is de erkenning van het belang om de

verschillende types van stakeholders te identificeren en het besef dat zij verschillende en niet zelden tegenstrijdige belangen hebben. Omwille van de eerder vermelde redenen hecht de auteur sterk belang aan het afbakenen van de scope van een evaluatie. Daarom dient de evaluator bij de eerste stap in het evaluatieproces zich te beperken tot wat Patton noemt de primaire gebruikers van de evaluatie. Primaire gebruikers zijn stakeholders die iets willen weten over het evaluandum én die gebruik kunnen maken van de evaluatiebevindingen. In dat perspectief komen vooral de beleidsmakers en uitvoerders van het beleid in beeld. Met het oog op het gebruik en de invloed van evaluatie in het kader van beleidsbeslissingen wijst Patton verder op het belang van interactie met de primaire gebruikers bij de structurering van de evaluatie.

Ondanks zijn pleidooi voor de selectie van stakeholders, benadrukt Patton terecht dat men verderop in het evaluatieproces de andere categorieën van stakeholders niet uit het oog mag verliezen. In het bijzonder de doelgroep kan heel vaak belangrijke informatie aanreiken voor de uitvoering van de evaluatie. Daarnaast kan de communicatie van de evaluatiebevindingen gericht worden aan de diverse categorieën van stakeholders. Daarbij is een aangepaste werkwijze en selectie van kanalen belangrijk. Op het communicatieaspect komen we verderop in dit deel van de handleiding nog terug.

In de praktijk zal de selectie van stakeholders, de mate waarin en de wijze waarop ze betrokken worden in de verschillende fasen van het evaluatieproces afhangen van meerdere factoren:

- het evaluatiemotief
- de kenmerken van het evaluandum
- de concrete evaluatievragen (cf. infra)
- de politieke gevoeligheid van de informatie
- het belang dat aan inspraak en participatie van stakeholders wordt gehecht door de opdrachtgever van de evaluatie

4.1.4 Opstellen en selecteren van de evaluatievragen en –criteria

Het kiezen en opstellen van evaluatievragen en het selecteren van één of meerdere evaluatiecriteria behoort tot de kern van het evaluatiedesign. Aan de hand van de evaluatievragen en -criteria krijgt de evaluatie een concrete inhoudelijke focus aangemeten. Daarop wordt de uitvoering van heel het verdere evaluatieproces gebaseerd. Een gebrek aan evaluatievragen of het onvoldoende expliciteren ervan is de perfecte manier om een evaluatie totaal de verkeerde richting uit te sturen.

Beslissen op welke **vragen** over het beleid de evaluatie antwoorden moet aandragen is een cruciale maar geen eenvoudige opdracht. Uit de veelheid van mogelijke vragen dient men een keuze te maken. Bovendien is de concrete formulering van de vraag van belang. Een vraag lichtjes anders verwoorden kan aanleiding geven tot een andere focus van de evaluatie en bijgevolg de inzet van andere instrumenten of technieken in de verdere stappen van het evaluatieproces.

Evaluatievragen zijn in vier generieke categorieën in te delen.

- (1) *Beschrijvende* vragen hebben als doel het verzamelen van informatie over “wat is...?”. Hun functie is het schetsen en beschrijven van een toestand, verandering, kenmerken, processen enz. Ze verschaffen basisinformatie rond het beleid, de werking van het beleid, de resultaten enz.
- (2) *Causale* vragen hebben betrekking op oorzaak en gevolg relaties. Ze hebben als doel het verzamelen van informatie over het bestaan, de sterkte, richting en randvoorwaarden van causale verbanden. Ze zijn functioneel voor het verwerven van inzichten in en begrijpen van bijvoorbeeld de oorzaken van de beleidsproblematiek en mate waarin het beleid (output) bijdraagt aan de geobserveerde verandering of effecten, d.i. de effectiviteit van het onderzochte beleid. Dergelijke vragen zijn meestal retrospectief: ze kijken naar *bestaande* gebeurtenissen, processen en veranderingen met een oorsprong in het verleden, en die op het moment van de evaluatie al dan niet beëindigd zijn.
- (3) *Normatieve* of *prescriptieve* vragen verwijzen naar een gewenste situatie of verandering. Zij hebben als doel het verzamelen van informatie over “wat moet er gebeuren om...?”. Ze zijn functioneel voor het vergelijken tussen een bestaande en gewenste toestand. De gewenste situatie kan vertaald zijn in een vooropgestelde beleidsdoelstelling of norm. Ze kan ook verwijzen naar een ideale of optimale situatie. Dit soort vragen stelt men doorgaans met het oog op het uitwerken van verbeteringsacties en –trajecten. Ze passen met andere woorden binnen een instrumenteel gebruik van een evaluatie.
- (4) *Voorspellende* vragen hebben betrekking op mogelijkheid van situaties of veranderingen in de toekomst. Ze hebben als doel het verzamelen van informatie over de mogelijkheid, kans en risico waarmee deze situaties of veranderingen kunnen optreden. Ze zijn functioneel voor het verwerven van inzichten bijvoorbeeld in de effecten en impact van gepland beleid en/of de vergelijking beleidsalternatieven, toekomstige evoluties van externe of omgevingsfactoren en de relatie tussen deze factoren en het onderzochte beleid. Vaak hebben deze vragen dus ook betrekking op causale verbanden, maar in tegenstelling tot de causale vragen zijn ze altijd prospectief. Ze verwijzen naar *eventuele* gebeurtenissen, processen en veranderingen in de toekomst.

Meestal zal de evaluator of opdrachtgever niet in deze termen spreken over evaluatievragen. Toch biedt deze typologie een kapstok om vragen in een concrete evaluatie te herkennen en bijgevolg te weten wat de focus van de evaluatie is en om welke soort evaluatie het gaat. Beschrijvende vragen keren in de meeste evaluaties terug. Causale vragen staan meestal centraal in een ex post evaluatie, terwijl voorspellende vragen duiden op een ex ante evaluatie. Normatieve vragen kunnen zowel in een ex ante of ex post evaluatie opgenomen worden. Bij normatieve forecasting (ex ante) stelt men de vraag welk beleid nodig is om naar een gewenste toekomstige situatie of scenario te evolueren. Een ex post evaluatie kan naast en gebaseerd op een beoordeling van een voorbij of bestaand beleid de vraag bevatten hoe het beleid bijgestuurd moet worden. Tussentijdse evaluaties zijn hybride doordat ze naar het verleden en/of de toekomst kunnen kijken. Elk van de bovenstaande soorten evaluatievragen kan dus in principe aan bod komen in een tussentijdse evaluatie.

De volgorde in bovenstaande typologie van evaluatievragen werd niet zomaar gekozen. Er zit een zekere hiërarchie in met betrekking tot de moeilijkheidsgraad waarmee men ze kan beantwoorden. Bij beschrijvende vragen ligt deze een stuk lager in vergelijking met de andere soorten. Causale en normatieve vragen hebben een veel grotere moeilijkheidsgraad. Voor het accuraat beantwoorden van normatieve vragen is vaak ook kennis over causale verbanden vereist. Bij voorspellende vragen speelt de problematiek van onzekerheid bij het formuleren de antwoorden. Dergelijke vragen hebben bijvoorbeeld betrekking op verwachte, geschatte, berekende effecten van nog niet uitgevoerd beleid.

Welke werkwijze is raadzaam voor het opstellen en selecteren van evaluatievragen?

Meestal zal men twee fasen doorlopen: een eerste fase gekenmerkt door divergentie, gevolgd door een fase van convergentie. In de eerste fase vindt een consultatieronde onder de geselecteerde stakeholders plaats. Het evaluatieteam verzamelt daarbij een breed gamma van potentiële vragen. Er wordt nog geen schifting gemaakt omdat een vraag in eerste instantie onhaalbaar of onverdedigbaar lijkt. Doordat verschillende stakeholders verschillende vragen en prioriteiten stellen zal het geheel aan geformuleerde vragen onsamenhangend en zeer gevarieerd zijn. Vervolgens zal men uit deze longlist een selectie van vragen moeten maken waarover de stakeholders bovendien een consensus bereiken. In deze twee fase is de beargumentering van het al dan niet weerhouden van de initieel geformuleerde evaluatievragen belangrijk. Deze werkwijze levert voordelen op voor de evaluator en de stakeholders. De stakeholders worden betrokken en bepalen gemeenschappelijk de scope van de evaluatie waardoor hun verwachtingen gelijkgesteld worden. De (interne of externe) evaluator kan vertrekken van een duidelijke, onderbouwde en compacte set van evaluatievragen om de evaluatie uit te voeren.

Het proces impliceert vanzelfsprekend een onderhandelingsmoment tussen de stakeholders. Stakeholders verschillen niet enkel op het vlak van belangen en prioriteiten, maar vaak ook van (machts)positie. Een belangrijk aandachtspunt is dat elke stakeholder die betrokken wordt bij de structurering van de evaluatie ook daadwerkelijk inspraak heeft en kan participeren in deze stap van het evaluatieproces. Er bestaan technieken om precies te bewaken dat de verschillende standpunten en perspectieven van stakeholders aan bod komen. Enkele van deze technieken, waaronder colour voting, worden besproken in deel 3 van deze handleiding.

Bovenstaand proces om vragen te selecteren garandeert nog geen kwalitatief goed geformuleerde vragen. Daarom is een kwaliteitscheck voor elke evaluatievraag belangrijk. Wat is een goede evaluatievraag? Volgende checklist reikt de 7 voornaamste aandachtspunten aan.

1. *Voor wie is de informatie bestemd? Wie zal de evaluatieresultaten gebruiken?*
2. *Zal het antwoord op de vraag nieuwe informatie opleveren?*
3. *Bevat het antwoord belangrijke informatie om beleidskeuzes te maken?*
4. *Brengt de vraag de afbakening en/of de begrijpbaarheid van de evaluatie niet in het gedrang?*
5. *Zal het antwoord op de vraag daadwerkelijk een impact hebben op het beleid?*
6. *Is het mogelijk de vraag te beantwoorden, gegeven de beperking op het vlak van capaciteit (budget, personeel, tijd, expertise)?*
7. *Is het mogelijk de vraag te beantwoorden, gegeven de beleidscontext (stand van zaken beleid, beschikbaarheid en/of verzamelbaarheid van data, politieke gevoeligheid,...)?*

Tabel 5: Checklist voor kwaliteitsvolle evaluatievragen

Bij het selecteren van vragen kan men een matrix opstellen waarbij elke initieel geformuleerde evaluatievraag wordt beoordeeld en gescoord aan de hand de bovenstaande elementen. Men kan per vraag ook de bedenkingen en argumenten vanwege de stakeholders ophoofden.

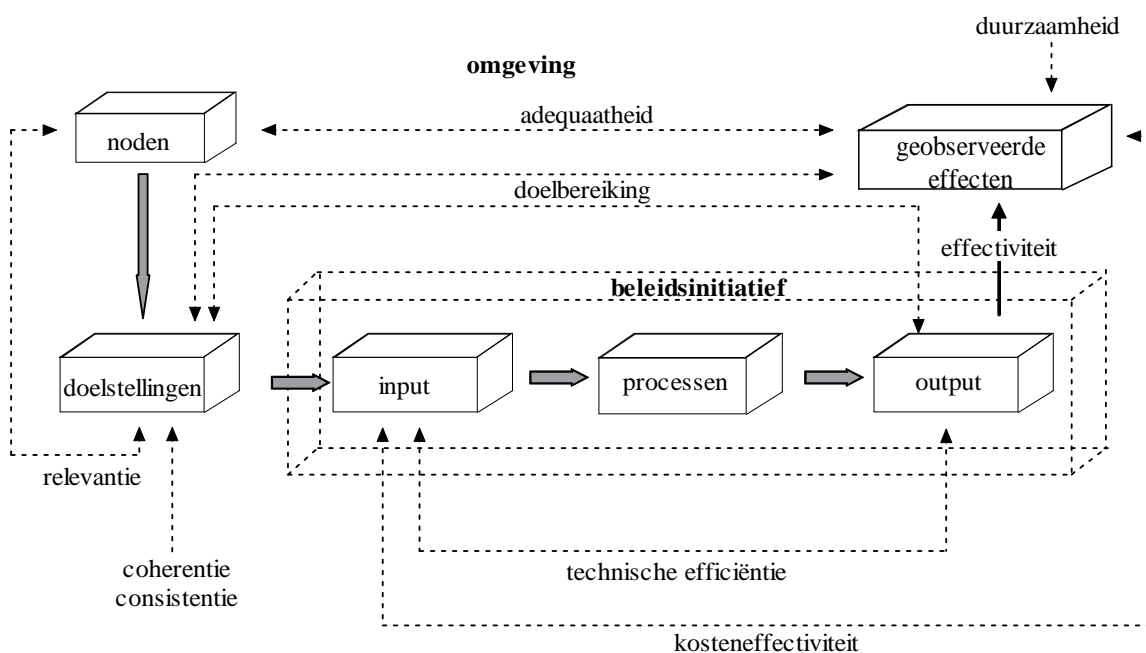
De twee laatste elementen uit de checklist verwijzen respectievelijk naar de organisatorische en inhoudelijke *evalueerbaarheid* van de evaluatievraag. Een inschatting van de evalueerbaarheid is een essentieel onderdeel van de kwaliteitscontrole op de evaluatievragen. Een vraag die omwille van praktische of inhoudelijke redenen niet of onvoldoende onderzocht kan worden, zal bij voorbaat geen accuraat en bruikbaar antwoord kunnen opleveren. Het is van belang om vragen die (nog) niet evalueerbaar zijn te filteren en uit te sluiten van de evaluatie. Het aandragen van niet evalueerbare

vragen kan immers een strategie zijn van tegenstanders van de evaluatie die het evaluatieproces misschien niet kunnen tegenhouden maar wel aanwenden om de evaluatie en diens mogelijke invloed op beleidsbeslissingen te ondermijnen.

Het inschatten van de evalueerbaarheid is een techniek die overigens niet alleen op individuele evaluatievragen kan toegepast worden. Een 'evaluability assessment' is ook hanteerbaar als een analyse-instrument met betrekking tot een beleidsinitiatief. De facto gaat het dan om een speciale vorm van ex ante evaluatie. Het is een meta-evaluatie die de vraag stelt of aan de randvoorwaarden is voldaan om het beleid ex ante, tussentijds of ex post te kunnen evalueren.

Een goede evaluatievraag ten slotte bevat ook een evaluatiecriterium. Zowel de koppeling van de vraag en het criterium, als de regel dat één vraag slechts één criterium mag bevatten zijn daarbij belangrijk.

Evaluatiecriteria vormen eveneens een kapstok om het soort evaluatie te typeren. Eerder in dit boekdeel hebben we de verschillende generieke evaluatiecriteria gerelateerd aan de beleidscyclus als een open systeemmodel. De volgende figuur herneemt de schematische voorstelling van deze metafoor.



Evaluatiecriteria hebben betrekking op enerzijds kenmerken van de componenten uit het open systeemmodel en anderzijds op relaties tussen deze componenten. In het bijzonder zijn in het bovenstaande schema volgende evaluatiecriteria opgenomen.

- *Beleidsrelevantie*

De mate waarin de doelstellingen tegemoet komen aan de maatschappelijke noden en behoeften.

- *Interne coherentie*

De mate waarin de strategische en operationele doelstellingen een samenhangend geheel vormen.

- *Externe consistentie*

De mate waarin de doelstellingen compatibel zijn met doelstellingen binnen andere beleidsvelden.

- *Technische efficiëntie*

De verhouding tussen input en output.

- *Doelbereiking*

De mate waarin de doelstellingen en meer bepaald de ambitieniveaus of normen worden gehaald.

- *Effectiviteit*

De mate waarin de geobserveerde effecten of veranderingen toe te schrijven zijn aan de beleidsoutput.

- *Kosteneffectiviteit*

De verhouding tussen input en effect.

- *Duurzaamheid van effecten*

De mate waarin effecten blijven bestaan na de beëindiging van het beleid.

- *Adequaatheid van effecten*

De mate waarin de effecten of veranderingen tegemoet komen aan de maatschappelijke noden en behoeften.

Bovenstaand schema is niet exhaustief. Bijkomende belangrijke evaluatiecriteria zijn:

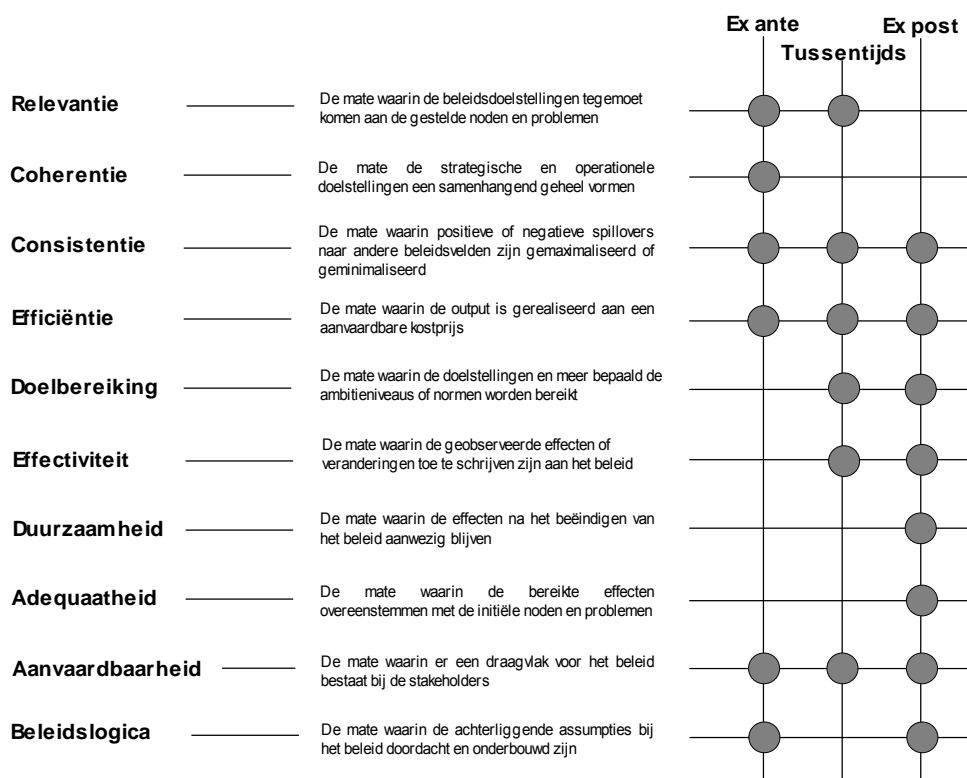
- *Draagvlak voor het beleid*

De mate waarin de doelstellingen, maatregelen en instrumenten van het beleid aanvaard worden door de stakeholders, en in het bijzonder door de doelgroep.

- *Sterkte van de achterliggende beleidslogica*

De mate waarin de assumpties achter een beleidsinitiatief doordacht en onderbouwd zijn.

Stakeholders zullen vaak geïnteresseerd zijn in mogelijke effecten die geen deel uitmaken van de beleidsopzet. De aandacht voor zogenaamde *neveneffecten* en meer bepaald ongewenste neveneffecten dient men situeren in het kader van evaluaties waarbij het criterium van effectiviteit of consistentie centraal staat.



Figuur 9: Evaluatiecriteria voor ex ante, tussentijdse en ex post evaluatie

De bovenstaande criteria zijn te beschouwen als generieke of basiscriteria om beleid te evalueren. In eenzelfde evaluatie kunnen meerdere van deze evaluatiecriteria centraal staan. Echter, hoe meer criteria, hoe uitgebreider en complexer de evaluatie zal zijn.

Er zijn daarnaast nog tal van andere maatstaven om een evaluandum te beoordelen en/of onderling te vergelijken met alternatieven of varianten. Dergelijke maatstaven hebben zijn meestal af te leiden uit de beleidsdoelstellingen of hebben betrekking op specifieke kenmerken. Zo kan men bijvoorbeeld alternatieve maatregelen of varianten van eenzelfde maatregel gaan vergelijken op hun effecten en impact. Bijvoorbeeld kan men de beleidsalternatieven spoorlijnverbinding vs. extra autoweg beoordelen en vergelijken op hun impact op de verkeersintensiteiten op het bestaande verkeersnet, op het milieu enz. Binnen de beleids optie van de spoorlijnverbinding kan men verschillende spoorwegtracés gaan vergelijken op een aantal kenmerken en effecten (reistijdwinst, lawaaihinder, etc.).

Het laatste aandachtspunt dat we hier willen aanreiken betreft de **koppeling tussen evaluatievragen en -criteria**. Vertrekkende vanuit het evaluatiemotief zal men meestal eerst keuzes maken tussen de mogelijk evaluatiecriteria en vervolgens in functie daarvan de vragen opstellen. Bij de structurering van de evaluatie en meer bepaald bij de vraagformulering dient men ervoor te zorgen dat elke evaluatievraag maximaal één evaluatiecriterium bevat. Zo vermijdt men dat bepaalde aspecten van een vraag niet of onvoldoende zullen worden afgedekt en doorvertaald in het totale evaluatiedesign en in de eigenlijke uitvoering van de evaluatie.

Met het opstellen en selecteren van evaluatievragen en -criteria wordt de inhoudelijke scope van de evaluatie afgerond. De tweede kerntaak om te komen tot een totaal evaluatiedesign is het plannen van de volgende fasen van het evaluatieproces. De volgende secties gaan in op het plannen van de gegevensverzameling, de analyse en de beoordeling.

4.2 Het plannen van de uitvoering van de evaluatie

De tweede kerntaak in het evaluatiedesign, na het bepalen van de inhoudelijk scope, is het plannen van de volgende standaardstappen in het evaluatieproces. Deze stappen zijn achtereenvolgens de gegevensverzameling, de data-analyse en de beoordeling. Het plannen van deze stappen komt in hoofdzaak neer op het maken van keuzes met betrekking tot het instrumentarium om de evaluatie uit te voeren.

De eerstvolgende sectie heeft een dubbel doel:

- het uitklaren van een aantal begrippen
- het aanreiken van een aantal kernprincipes om de volgende stappen van het evaluatieproces te plannen.

De daarop volgende secties gaan dieper in op de volgende stappen in het evaluatieproces en de bijhorende aandachtspunten. Boekdeel 3 van deze handleiding is volledig gewijd aan evaluatietechnieken per standaardstap van het evaluatieproces. Naast de omschrijving van elke techniek, wordt in dat deel aandacht besteed aan de gebruikscontext, de toepassing van de techniek, en de mogelijkheden en beperkingen.

4.2.1 Inleiding: enkele begrippen en kernprincipes

Wanneer men spreekt over de uitvoering van een evaluatie is het nuttig om een aantal vaak terugkerende begrippen uit te klaren.

Zo maken we een onderscheid tussen de termen methode en techniek. Met een *methode* verwijzen we naar de algemene aanpak bij een evaluatie. Het gaat om een werkwijze die kenmerkend is voor elke evaluatie apart. Een methode heeft dus steeds een ad hoc karakter, net zoals een evaluatiedesign, terwijl deze laatste term verwijst naar een breder geheel van zowel inhoudelijke elementen als methodologische aspecten.

De term methode komt bijgevolg overeen met wat we eerder hebben geduid als een evaluatiebenadering. Ten grondslag van een methode of benadering liggen een aantal fundamentele uitgangspunten of waarden waaraan de evaluator sterk belang hecht. Een benadering kan bijvoorbeeld gekenmerkt zijn door een sterke nadruk op de participatie van verschillende stakeholders doorheen het evaluatieproces, of het belang dat men hecht aan de beleidstheorie om een evaluatie vorm te geven.

De term *methodologie* reserveren we hier voor de ‘wetenschap’ of kunde met betrekking tot het ontwerpen van methoden. Deze handleiding is echter eerder gericht op het bijdragen tot praktische expertise dan puur wetenschappelijke kennis over methoden voor evaluatie.

Een *techniek* definiëren we als een gestandaardiseerde werkwijze. De term *instrument* leunt hier zo dicht bij aan dat we beide termen in deze handleiding als synoniemen gebruiken. Echter, een instrument kan ook verwijzen naar een object. De techniek is in dat geval de manier waarop het object gehanteerd wordt. Het Engelse leenwoord ‘tool’ is eveneens een synoniem voor techniek.

Het plannen van de volgende stappen in het evaluatieproces komt in hoofdzaak neer op het maken van keuzes tussen technieken om de evaluatie uit te voeren. Voor het maken van deze keuzes kan men zich laten leiden door enkele kernprincipes en criteria.

- In functie van evaluatievragen, niet omgekeerd

Ook voor de keuze van instrumenten geldt het evaluatiemotief en daarmee samenhangend het type van de evaluatie het vertrekpunt. Hieruit worden de evaluatievragen en -criteria afgeleid (cf. supra). Een eerste kernprincipe is dan ook dat de keuze van evaluatietechnieken moet gebeuren in functie van de evaluatievragen die men wil beantwoorden en niet omgekeerd. Met de soort evaluatievragen hangen diverse kenmerken samen. Ten eerste het tijdsperspectief van de evaluatie; niet elke techniek is even geschikt voor toepassing in een ex ante, tussentijdse of ex

post evaluatie. Ten tweede het soort informatie: heeft men eerder nood aan kwantitatieve of kwalitatieve informatie? Ten derde de hoeveelheid informatie; sommige technieken zijn meer geschikt dan andere om een bepaald datavolume te verzamelen of te verwerken.

Het belang van dit kernprincipe kan niet genoeg onderstreept worden. In de praktijk zijn opdrachtgevers en evaluatoren niet in dezelfde mate vertrouwd met het brede spectrum aan evaluatietechnieken. Het risico bestaat dat men voortgaat om wat men kent, zonder na te gaan of de techniek inderdaad de meest aangewezen optie is. Het daarom van fundamenteel belang dat opdrachtgevers en evaluatoren kennis hebben van de diversiteit aan technieken en de mogelijkheden en beperkingen van een concrete techniek. Deel 3 van de handleiding tracht tegemoet te komen aan deze nood.

- Aangepast aan het evaluandum

Specifieke kenmerken van het evaluatieobject kunnen ervoor zorgen dat bepaalde technieken meer geschikt zijn dan andere. De aard van het beleidsveld of een beleidsprogramma bijvoorbeeld kan mee bepalen dat kwalitatieve technieken (bv. interviews, focus groepen) meer bruikbare informatie zullen opleveren dan kwantitatieve technieken (bv. survey). De evaluatie van domeinoverschrijdend beleid vereist mogelijk een combinatie technieken om verschillende aspecten te evalueren. Zo stelt men bijvoorbeeld dat duurzame ontwikkeling steunt op drie pijlers en de combinatie van economische, sociale en ecologische maatregelen. Voor het evalueren van dergelijk beleid kan een combinatie van evaluatietechnieken een optie zijn om elk van deze pijlers aan bod te laten komen. Anderzijds kan men belang hechten aan het formuleren van een geaggregeerde of synthesebeoordeling. Niet gelijk welke evaluatietechniek kan daarvoor oplossingen aanreiken.

- Technische kwaliteit

De betrouwbaarheid van een techniek is vanzelfsprekend ook een selectie criterium. Technieken die al veelvuldig werden toegepast hebben meestal een proces van verfijning en verbetering doorgemaakt. Bij dergelijke technieken zijn de mogelijke knelpunten dan genoegzaam bekend waardoor men erop kan anticiperen.

Hoe hoger de mate van toepassing, des te stabielere de techniek kan worden beschouwd. Wanneer het echter gaat om een pas ontwikkelde techniek, zal het gebruik ervan innovatief zijn. Nieuwe technieken kunnen innoverend zijn in de mate dat ze een alternatief vormen voor bestaande, stabiele technieken met dezelfde functie bv. gegevensverzameling of analyse. Een bestaande techniek kan ook uitgetest worden in een meer ongewone functie of context.

- **Kostprijs**

Een laatste criterium dat niet zonder belang is bij de selectie van instrumenten is de *kostprijs* van een bepaalde techniek. Deze kostprijs hangt vaak samen met de mate van technische complexiteit. De kostprijs is een belangrijke randvoorwaarde maar mag niet het primaire selectiecriterium zijn. Anderzijds kan men beter opteren voor een technisch minder complexe evaluatie die goed is afgelijnd, dan een evaluatie waarbij zeer geavanceerde en daardoor mogelijk duurere technieken worden ingezet maar waarvan de meerwaarde voor de concrete evaluatiecontext overschat wordt. Tevens dient vermeden dat een welbepaalde techniek het uithangbord wordt van een evaluatie en ermee vereenzelvigd wordt. Een evaluatie gelijkstellen aan een symbolisch instrument is steeds een vertekening. Een techniek vervult immers een bepaalde functie binnen het ganse evaluatieproces. De volledige uitvoering van een evaluatie is niet mogelijk met één en dezelfde techniek. Het negeren van de andere noodzakelijke functies ondermijnt de kwaliteit van de evaluatie als geheel.

4.2.2 Planning van de gegevensverzameling

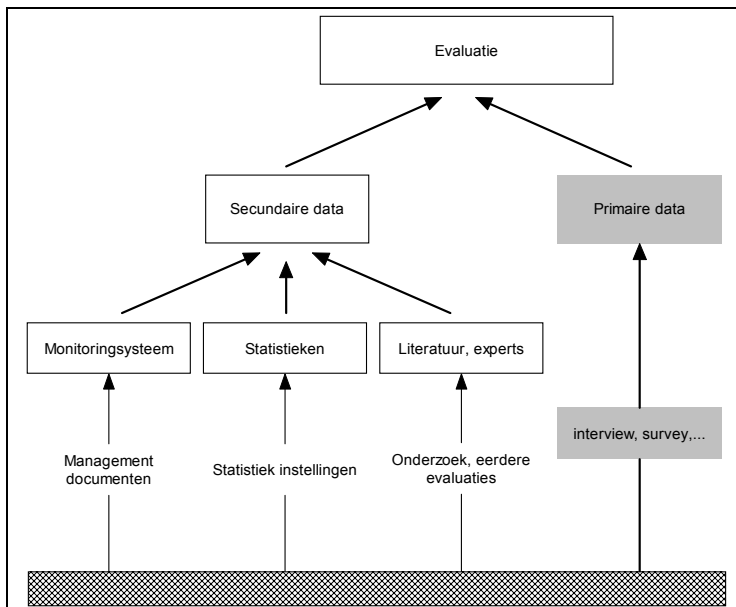
Beleidsbeoordelingen zijn functioneel voor het voeren van een 'evidence-based' ofwel een met kennis ondersteund en onderbouwd beleid. Kennis wordt stelselmatig opgebouwd (Russell Ackoff, 1989; Weggeman, 1998). Het vertrekpunt vormen ruwe gegevens. Informatie ontstaat door het toekennen van een betekenis aan deze gegevens vanuit een referentie- of interpretatiekader. De stap van informatie naar kennis wordt gezet via een leermoment; wanneer men verbanden en verklaringen ontdekt. Een evaluatieproces bevat dergelijk cumulatief leerproces.

Een evaluatie kan dus niet uitgevoerd worden zonder eerst gegevens te verzamelen. Elke evaluatiesetting is uniek, waardoor (de planning van) de gegevensverzameling context- en tijdsgebonden is. Bij het plannen van de gegevensverzameling keren de volgende vragen terug:

- Welke gegevens en informatie zijn noodzakelijk?
- Welke bronnen zijn beschikbaar?
- Welke methoden en technieken zijn mogelijk?
- Welke tijdsdimensie en frequentie van gegevensverzameling zijn relevant?

Bij een goed gestructureerd evaluatiedesign kan men het antwoord op de eerste vraag afleiden uit de scope van de evaluatie: de evaluatievragen en -criteria. In functie daarvan dienen de andere vragen beantwoord te worden.

Evaluaties kunnen gevoed worden met twee soorten gegevens: primaire en secundaire data. Gegevens die nog niet verzameld zijn bij aanvang van de evaluatie zijn primaire data. Al bestaande en verzamelde gegevens zijn secundaire data. Onderstaand schema geeft de verschillende bronnen van secundaire en primaire data weer.



Figuur 10: Bronnen voor gegevensverzameling

Bij de **bestaande of secundaire bronnen** kan men volgende opdeling maken:

- monitoringsystemen van de organisaties betrokken bij het beleid
 - gegevens van statistiekinstellingen
 - eerder uitgevoerd beleidsgericht onderzoek en evaluaties
- Monitoring

De organisaties betrokken bij het beleid hebben vaak eigen, interne prestatie-meetsystemen of monitoringsystemen waarmee het management beheers- en beleidsinformatie verzamelt op een systematische en permanente basis. Het management kan dergelijke informatie aanwenden ter ondersteuning van beslissingen met betrekking tot de werking van de organisatie. Indicatoren vormen de basisbestanddelen van een meetsysteem. Afhankelijk van de focus van het meetsysteem kunnen de indicatoren gericht zijn op een of meerdere componenten van de beheers- en beleidscyclus:

- input
- processen
- output
- effecten
- omgevingsfactoren

Inputindicatoren leveren informatie over de ingezette en geplande financiële middelen, het personeel, materiële middelen etc. Procesindicatoren leveren informatie over de activiteiten die de organisatie uitvoert. Outputindicatoren leveren informatie over de prestaties van de organisatie. Dat kunnen fysieke producten (bv. identiteitskaart, vergunning, riolering) zijn, bestemd voor individuele burgers of collectieve goederen, of een bepaalde dienstverlening (bv. advies, informatie, begeleiding). Effectindicatoren leveren informatie over factoren en fenomenen buiten de organisatiegrens (bv. verkeersongevallen, klantentevredenheid, bouwaanvragen etc.). Beleid is doorgaans gericht op het beïnvloeden van deze factoren. Effecten zijn dan in se de veranderingen bij deze factoren. Deze veranderingen kan men plannen en observeren. Omgevingsindicatoren leveren eveneens informatie over factoren buiten de organisatiegrens. In tegenstelling tot effectindicatoren zijn ze niet rechtstreeks gekoppeld aan beleidsdoelstellingen. De invloed van omgevings- of externe factoren kan naast de beleidsoutput, een alternatieve of aanvullende verklaring vormen voor de vastgestelde effecten of veranderingen.

Doordat indicatoren gerelateerd kunnen worden aan meerdere componenten en fasen in de beheers- en beleidscyclus, kan monitoring een zeer belangrijke informatiebron vormen in het kader van een evaluatie. De evaluator dient oog te hebben voor de focus van het monitoringsysteem. In welke mate is het relevant voor het verzamelen van de informatie die noodzakelijk is om de evaluatie uit te voeren? Om als secundaire bron relevant te zijn voor een evaluatie, dient de focus van het monitoringsysteem overeen te stemmen met de focus van de stakeholders bij de evaluatie.

De toegankelijkheid van dergelijke bron voor de evaluator hangt af van de betrokken organisatie. De mate waarin dit aspect een rol speelt, zal verschillen naargelang het om een interne of externe evaluatie gaat. Informatie gegenereerd door monitoringsystemen wordt vaak verwerkt in managementrapportages en beleidsdocumenten. Dergelijke documenten bevatten dan ook vaak verwijzingen naar bestaande meetsystemen die voor de evaluator relevant kunnen zijn.

Boekdeel 2 van deze handleiding is volledig gewijd aan meten en monitoring en de relatie met evaluatie. Het permanente karakter van monitoring en het vaak ad hoc karakter van evaluatie vormt een spanningsveld. Een belangrijke strategische uitdaging ligt dan ook in de onderlinge afstemming van beide beleidsondersteunende instrumenten. Tevens bevat boekdeel 2 een stappenplan om een meetsysteem te ontwerpen met het oog op het kunnen evalueren van beleid. Daarbij wordt ook stilgestaan bij kritieke slaagfactoren en kwaliteitsdimensies op het vlak van het ontwerpen van indicatoren en de meetprocedures. Deze laatste vormen dan ook belangrijke aandachtspunten wanneer men overweegt om gegevens uit monitoring te gaan gebruiken in het kader van een evaluatie.

- Statistieken

Openbare statistieken en databanken vormen ook een potentieel belangrijke bron voor een evaluatie. Instituties zoals EUROSTAT, het NIS en de Studiedienst van de Vlaamse Regering (voorheen APS) besteden veel aandacht aan de betrouwbaarheid van gegevens door het standaardiseren van de meetprocedures en bijkomende kwaliteitscontroles. Anderzijds heeft dergelijk publiek toegankelijke informatie ook een aantal beperkingen. Los van de vraag of de informatie relevant is in het kader van een evaluatie, is de informatie niet steeds op het gewenste detailniveau beschikbaar. Gegevens zijn vaak op een geaggregeerd niveau consulteerbaar. Zogenaamde metagegevens in verband met het niveau van registratie, de gehanteerde procedures en technieken voor verzameling en de grootte van de samples zijn vaak moeilijker te achterhalen. De grootte van de dataset is nochtans van belang met het oog op het veralgemenen of generaliseren van conclusies. De actualiteit van de gegevens vormt soms ook een probleem. De evaluator zal de voorkeur geven aan data die zoveel als mogelijk up-to-date zijn. Vaak zijn gepubliceerde openbare statistieken echter gebaseerd op data die enkele maanden of zelfs jaren voorheen verzameld werden. Een zeker tijdsinterval is bijna steeds in te calculeren; real-time verwerking en rapportering is in veel gevallen niet kosteneffectief.

- Eerder beleidsgericht onderzoek en uitgevoerde evaluaties

Relevante secundaire gegevens kunnen ook door middel van een literatuurstudie gevonden worden, meer bepaald in de rapporteringen van eerder beleidsgericht onderzoek of al uitgevoerde evaluaties over hetzelfde of aanverwant thema. In menig opzicht kunnen deze bronnen de evaluator helpen. Ze bevatten mogelijk interessante modellen, technieken en bronnen met betrekking tot dataverzameling over het onderwerp. Wanneer een evaluatievraag al onderzocht is in andere evaluaties, kan de evaluator soms de tijd besparen. Belangrijk is dan de kwaliteit van de analyse in de eerdere studie en de mate waarin de conclusies veralgemeend kunnen worden naar bijvoorbeeld andere beleidssectoren of settings, of regio's. Aanvullend kan de evaluator ook informatie inwinnen door rechtstreeks de experts te consulteren die de eerdere studie of evaluatie hebben uitgevoerd.

In veel gevallen zal de evaluator echter aangewezen zijn op **primaire data**: wanneer secundaire data ontbreken, niet alles afdekken of omwille van technisch beperkt of helemaal niet bruikbaar zijn.

Het komt er dan op aan om gegevens te verzamelen die op maat van de evaluatievragen gesneden zijn. De evaluator kan daarvoor zich richten op verschillende categorieën van stakeholders:

- de ontvangers van het beleid of de doelgroep
- de beleidsmakers en beleidsverstrekkers: politici en ambtenaren
- derden die kennis hebben over de beleidsontvangers en de bredere beleidscontext: experts, wetenschappers, ouders van leerlingen, etc.

De evaluator heeft voor de verzameling van informatie vaak de toestemming van beleidsmakers en ontvangers nodig. Daarbij is een aantal ethische beginselen na te leven, bijvoorbeeld het waarborgen van de privacy van de deelnemers en/of het vertrouwelijk en discrete behandelen van de verzamelde gegevens.

Eenmaal bepaald is welke informatie verzameld moet worden en bij welke stakeholders en actoren, dient een selectie gemaakt van methoden en technieken. In boekdeel 3 van de handleiding worden enkele technieken voor dataverzameling uitvoeriger besproken, waaronder het interview, focusgroepen en de schriftelijke vragenlijst of survey. Er wordt eveneens aandacht besteed aan het samenstellen van steekproeven. Dit boekdeel vergelijkt ook de mogelijkheden en beperkingen van kwantitatieve en kwalitatieve methoden voor dataverzameling.

4.2.3 Het instrumentarium voor analyse en beoordeling

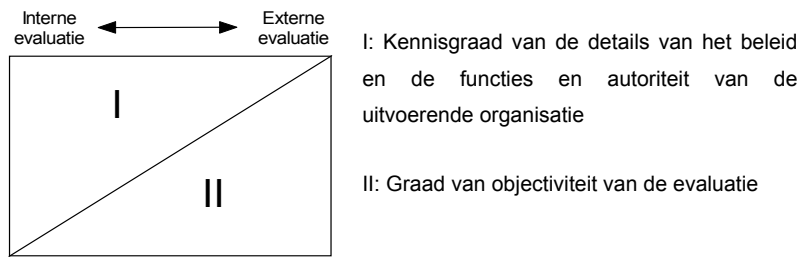
De planning van de standaardstappen van data-analyse en beoordeling verschilt naargelang de evaluatie intern of extern uitgevoerd wordt. Voor een interne evaluatie kan men op dit vlak sneller gedetailleerde keuzes maken dan wanneer de evaluatie uitbesteed zal worden. In het laatste geval is het raadzaam om de methodologische keuzes nog grotendeels open te laten en de kandidaat-evaluatoren hierover een voorstel te laten formuleren. Het voordeel voor de opdrachtgever is dat men op die wijze alternatieve voorstellen voor de aanpak van de evaluatie kan vergelijken vooraleer men de evaluatieopdracht gunt.

Dat neemt niet weg dat men als opdrachtgever al wel vaak een idee of voorkeur heeft op vlak van technieken. In het bestek van de evaluatieopdracht werkt men in dat geval aan de hand van richtlijnen een kader uit waarbinnen de kandidaat-evaluatoren nog een marge hebben om eigen accenten te leggen.

In de inleidende sectie werden al de kernprincipes en criteria toegelicht die als leidraad gelden voor het selecteren van concrete technieken voor analyse en beoordeling.

4.3 Uitbesteding van een evaluatie

In hoofdstuk 2 werden de voor- en nadelen van interne en externe evaluaties al opgelijst en vergeleken. De conclusie was dat er een trade-off bestaat tussen de twee vormen, waarbij het kennisgehalte over het beleid en de (gepercipieerde) objectiviteit van de evaluatie twee voorname parameters zijn.



Figuur 11: Kenniswaad en objectiviteit bij interne en externe evaluatie

Bovendien is een combinatie van beide vormen niet uitgesloten en een oplossing om de voor- en nadelen van de aparte vormen respectievelijk te benutten en te vermijden. Wanneer men beslist om de evaluatie uit te besteden, creëert men de facto dergelijke hybride vorm: de planning van de evaluatie gebeurt intern, de uitvoering extern. In voorgaande paragrafen van dit hoofdstuk kwamen de verschillende aspecten van het evaluatiedesign aan bod. Het belang van goede interne voorbereiding en planning van de evaluatie vormt één van de vertrekpunten van deze handleiding. Het bepalen van de inhoudelijke scope is een kerntaak voor de opdrachtgever. Hij kan daarin wel bijgestaan worden door de evaluator om deze taak af te ronden. In geen geval mag deze taak volledig overgelaten worden aan een externe evaluator wanneer de vraag tot evaluatie vanuit de opdrachtgever komt. Anders worden alle kansen tot gebruik van de evaluatie bij voorbaat gehypothekeerd.

Ingeval van aanbesteding vindt deze voorbereiding idealiter zijn neerslag in het evaluatiebestek. De volgende sectie gaat in op de elementen die in een bestek voor een evaluatieopdracht moeten worden opgenomen. De daarop volgende sectie staat stil bij de fase van selectie van de externe evaluator.

4.3.1 Opmaak van een bestek voor een evaluatieopdracht

Het bestek is het document waarin de modaliteiten en randvoorwaarden van een evaluatieopdracht worden omschreven. De opmaak van een accuraat bestek is dan ook een essentiële stap wanneer een evaluatie wordt uitbesteed, aangezien het de grondslag vormt voor een contractuele verbintenis tussen de opdrachtgever en de uitvoerder van de evaluatie. Nochtans is deze stap evenzeer van belang wanneer de evaluatie gedeeltelijk in eigen beheer zal worden gedaan, ook al wordt het nut ervan in dergelijke gevallen vaak over het hoofd gezien.

In deze sectie gaan we dieper in de op verschillende ingrediënten van een goed bestek en wordt aldus een sjabloon van bestek aangereikt. Onderstaande tabel geeft de noodzakelijke bestanddelen van een bestek weer.

1.	Wettelijk kader en opdrachtgever
2.	Beleidsmatige scope van de evaluatie
3.	Opzet en stakeholders
4.	Evaluatievragen en -criteria
5.	Beschikbare gegevens en informatie
6.	Methode en technieken voor dataverzameling, analyse en beoordeling (richtlijnen)
7.	Tijdslijn
8.	Indicatief budget
9.	Vereiste kwalificaties voor de evaluator
10.	Structurering van offertes
11.	Bepalingen ter indiening en toewijzingscriteria

Tabel 6: Bestanddelen van een evaluatiebestek

- Wettelijk kader en opdrachtgever

Het wettelijke en institutionele kader voor de evaluatie is een eerste element. Meer bepaald wordt hier gespecificeerd welke de regelgevende bepalingen zijn waarop de evaluatie wordt gebaseerd of, bij het ontbreken daarvan, welke de aanleiding is voor de evaluatieopdracht. Daarbij kan al een eerste algemene omschrijving worden gegeven van wat de opzet is van de geplande evaluatie.

Tevens dient hier duidelijk vermeld wie of welke instanties de opdrachtgever en initiator zijn en welke actoren betrokken werden bij de opmaak van de opdracht.

- Beleidsmatige scope van de evaluatie

Het belang van het adequaat afbakenen van de inhoudelijke scope van een evaluatie kwam reeds in voorgaande paragrafen aan bod. Het bestek dient te verduidelijken welk(e) maatregel, project of programma, integraal of transversaal beleid geëvalueerd moet worden, alsook het relevante tijdssegment, het tijdsperspectief ten aanzien van de beleidscyclus (ex ante / tussentijds / ex post) en de relevante geografische grenzen.

- Opzet en stakeholders

Het belangrijk te vermelden wat de opzet van de evaluatie is. Zoals eerder beschreven zijn er verschillende perspectieven en motieven die ten grondslag kunnen liggen aan een evaluatie. Het toelichten van de opzet betekent dan ook dat men aangeeft op welke wijze en in welke context de resultaten van de evaluatie zullen gebruikt worden. Tevens dienen daarbij de voornaamste belanghebbenden of stakeholders opgegeven te worden. De identificatie van de stakeholders is een belangrijke stap in het design van een evaluatie (cf. supra). In het bestek dient verder aangegeven te worden in welke mate bepaalde stakeholders tijdens de uitvoering en/of opvolging van de evaluatie betrokken moeten worden. Dergelijke betrokkenheid kan verschillende vormen

aannemen, bijvoorbeeld via een technische opvolgingscommissie, als doelgroep bij de datavergaring, een klankbordgroep, e.d.

- Evaluatievragen en -criteria

Welke evaluatievragen staan centraal? Op welk evaluatiecriterium wordt gefocused? (cf. supra). Belangrijk is om bij de evaluatieopdracht de klemtoon te leggen op een beperkte reeks van vragen waarop de evaluatie een antwoord moet geven. Dit zorgt niet enkel voor een welomlijnde focus van de evaluatie, maar laat ook een betere kwaliteitscontrole toe. Het selecteren van evaluatievragen kan gebeuren aan de hand van een toetsing op evalueerbaarheid die eerder werd besproken.

- Beschikbare gegevens en kennis

Kandidaat-evaluatoren moeten hun voorstel van methode kunnen afstemmen op de omvang en aard van de gegevensverzameling en -analyse die tijdens de evaluatie dient te gebeuren. Daarom is het raadzaam om in het bestek een stand van zaken op te nemen over de bestaande kennis over het evaluatieobject en over de beschikbare en consulteerbare gegevens. Elementen in dit overzicht kunnen zijn: uittreksels van of verwijzingen naar planningsdocumenten, een lijst van uitgevoerde analyses en evaluaties, een beschrijving van de bestaande meetsystemen, toegankelijke databanken en raadpleegbare documenten.

- Methode en technieken (richtlijnen)

Elke evaluatie hanteert een ad hoc aanpak, opgesteld met akkoord van de opdrachtgever en gebaseerd op het voorstel van de evaluator waarin specifieke evaluatietechnieken worden vermeld. In de fase van het bestek is het aangewezen om nog geen gedetailleerde omschrijving van de aanpak en evaluatietechnieken te maken, maar om nog voldoende vrijheidsgraden te laten voor de kandidaat-evaluatoren om op dat vlak aan te geven hoe zij te werk zullen gaan. In bestek kan de opdrachtgever in hoofdlijnen vermelden wat deze als behoeften ziet met het oog op een resultaatgerichte evaluatie, namelijk het verschaffen van antwoorden op de gestelde evaluatievragen. Eventueel kan verwezen worden naar bepaalde technieken die in de lijn liggen van de intenties, bijvoorbeeld een survey onder de doelgroep van het betrokken beleid, een comparatieve analyse, e.d. Op deze wijze hebben kandidaat-evaluatoren de mogelijkheid om zich te onderscheiden op het vlak van de relevantie en duidelijkheid van hun voorgestelde methode. Dat laatste is van belang in het licht van de kwaliteitstoetsing van de voorstellen en de uiteindelijke selectie van de externe evaluator.

- Tijdslijn

Bij het uittekenen van de tijdslijn voor de evaluatie dienen verschillende beperkingen in acht genomen, meer bepaald deze in verband met het beslissingsproces gekoppeld aan het beleid dat wordt geëvalueerd. Ook bepaalde deadlines die samenhangen met de procedure tot uitbesteding moeten ingecalculeerd worden. Enkele fasen van de evaluatie, bv. dataverzameling, zullen steeds een minimum tijdsduur vergen, terwijl bepaalde onderdelen niet met elkaar kunnen overlappen en deel uitmaken van het zogenaamde kritisch pad van activiteiten: wanneer hier ergens een vertraging optreedt, vertraagt het hele proces.

Met bovenstaande elementen in acht genomen, kan men in het bestek de mogelijke totale tijdsduur van de evaluatie aangeven. Het is aan te bevelen om in een eerste fase te reserveren voor het verder verfijnen van de planning en structurering van de evaluatie door de evaluator in samenspraak met de opdrachtgever. Deze eerste fase kan uitmonden in een startrapport waarin de meer gedetailleerde methode en tijdsindeling worden beschreven. In de fase van de resultaten is het raadzaam om voldoende tijd te voorzien voor de bespreking en eventuele amendementen aan de draft van eindrapportage.

- Indicatief budget

Het is raadzaam om een indicatief budget op te nemen in het bestek en de kandidaat-evaluatoren te laten aangeven wat zij binnen deze financiële marge kunnen aanbieden. Op deze wijze kan de kostprijs van de evaluatie beter beheerst worden en kunnen de inkomende voorstellen op efficiëntie worden beoordeeld. Een alternatieve werkwijze bestaat erin dat men de kostprijs laat schatten door de kandidaat-evaluatoren op basis van wat zij als noodzakelijke activiteiten zien. In beide gevallen wordt best in het bestek een richtlijn opgenomen omtrent een transparante kostenstructuur waarbij een opdeling wordt gemaakt naar verschillende kostenplaatsen.

- Vereiste kwalificaties voor de evaluator

Het bestek dient tevens een aantal voorwaarden te stipuleren over het beoogde profiel van de externe evaluator. Daarbij kan verwezen worden naar de noodzakelijke methodologische kwaliteiten, ervaring met gelijkaardige opdrachten, kennis over de institutionele context en de regio (indien van toepassing), professionele achtergronden en specifieke expertise en de capaciteit om een evaluatie uit te voeren en op te leveren binnen bepaalde tijdsgrenzen.

Met het oog op de zorg voor onafhankelijkheid van een externe evaluator kan de potentiële evaluator ook worden gevraagd om in de offerte op te bevestigen dat er geen belangenconflicten bestaan binnen het evaluatieteam. Anderzijds wordt best ook beschreven in het bestek hoe de evaluator toegang zal kunnen hebben tot de relevante beleids- en managementverantwoordelijken en tot de achtergrondinformatie die noodzakelijk is voor de uitvoering van de evaluatie.

- Structurering van offertes

Om de verwerking van de voorstellen en de toewijzing te vergemakkelijken is het raadzaam om tevens enkele richtlijnen in het bestek op te nemen over de structuur van de offertes. Bepaalde elementen die zeker aan bod moeten komen in de offerte, kunnen gesuggereerd worden. Eventueel kan men ook een maximum lengte per onderdeel vermelden.

- Bepalingen ter indiening en toewijzingscriteria

Ten slotte dient duidelijk in het bestek vermeld wat de deadline voor het indienen van voorstellen is, de verschillende mogelijk manieren, en de geldigheidsduur van voorstellen.

Daarnaast is het aan te bevelen om ook de criteria te expliciteren op basis waarvan de voorstellen zullen beoordeeld worden en een selectie zal gemaakt worden. Tevens kan men daarbij het relatieve gewicht van de verschillende criteria aangeven. Bijvoorbeeld:

x % op de kwaliteit van de methode of aanpak

y % op de kwalificaties en ervaring van de evaluator

z % op de kostprijs

Het spreekt voor zich dat als dergelijke criteria vermeld worden in het bestek, deze consequent dienen te worden toegepast tijdens het beoordelen van de verschillende offertes.

4.3.2 Selectie van de uitvoerder van de evaluatieopdracht

De opmaak van een goed bestek zal bijdragen aan een transparante en gerichte keuze van de externe evaluator. Bij het selecteren van een externe evaluator spelen het profiel van het evaluatieteam en de methodologische kwaliteit van het voorstel een belangrijke rol.

Het profiel wordt enerzijds bepaald door de positie of het type instelling waartoe de evaluator behoort. Dit kan een universitaire onderzoeksinstelling of een private actor zijn. Terwijl de eerste categorie het voordeel heeft van een hoge onafhankelijkheid en geloofwaardigheid omwille van academische vereisten, kan de tweede categorie soms makkelijker capaciteit vrijmaken op korte termijn.

Een andere factor die het profiel van de evaluator beïnvloedt, is op welke deelterreinen de ervaring en expertise zich concentreert. Afhankelijk van een bepaalde evaluatieopdracht zal eerder gespecialiseerde kennis van het evaluatieobject of de ruimere context van belang zijn, hetzij meer algemene vaardigheden en kennis met betrekking tot specifieke evaluatietechnieken. Dit hangt deels ook af van de aanpak die de kandidaat-evaluator voorstelt. Een aanpak die een mix van externe en interne bijdragen aan de evaluatie voorziet, kan even goede of zelfs betere resultaten opleveren, dan

wanneer de evaluatie volledig wordt uitbesteed. Een interdisciplinair team kan vaak het best inspelen op de combinatie van bepaalde expertise en ervaring.

Belangrijk is dat de opdrachtgever of het door hem aangestelde selectiecomité een eigen oordeel kan vellen over de voorstellen op basis van de criteria die geëxpliciteerd werden in het bestek. Deze handleiding wil bijdragen aan dergelijke kenniscapaciteit. Centraal in de kwaliteitstoetsing is functionaliteit van de voorgestelde methode in relatie tot de gestelde evaluatievragen. Onderstaande tabel bevat een (niet exhaustieve) checklist van elementen die afgetoetst kunnen worden. Sommige items hebben betrekking op elke evaluatievraag afzonderlijk, andere vergen een inschatting voor het geheel van de gestelde evaluatievragen.

Wat is de kwaliteit van de voorgestelde methode op vlak van...	...met betrekking tot:									
	Evaluatievraag 1					Evaluatievraag 2				
- Dataverzameling: voldoende en relevant?	--	-	+/-	+	++	--	-	+/-	+	++
- Analyse: is deze gebaseerd op zorgvuldige technieken?	--	-	+/-	+	++	--	-	+/-	+	++
- Evaluatiecriteria: ondubbelzinnig uitgeklaard?	--	-	+/-	+	++	--	-	+/-	+	++
- Resultaatgerichtheid: kans op geloofwaardige bevindingen?	--	-	+/-	+	++	--	-	+/-	+	++
	Geheel van evaluatievragen									
- Volledigheid: komen alle evaluatievragen voldoende aan bod?	--	-	+/-	+	++					
- Interactie met stakeholders: voldoet de voorziene aanpak om stakeholders te betrekken aan de verwachtingen?	--	-	+/-	+	++					

Tabel 7: Checklist voor de beoordeling van evaluatieoffertes

Daarnaast dient de voorgestelde methode te worden getoetst aan het profiel van de evaluator. De werkwijze waarbij het bestek enkel de richtlijnen voor de methode en technieken aangeeft, kan immers ertoe leiden dat kandidaat-evaluatoren teveel aandacht besteden aan de originaliteit van de aanpak en onvoldoende aan hun capaciteit om het voorstel daadwerkelijk toe te passen.

Belangrijk is de beoordeling van de verschillende voorstellen een kwalitatieve en geen kwantitatieve inschatting is. De items in de checklist zijn ook kwalitatief van aard. In het geval van een survey-techniek bijvoorbeeld, kan de omvang van de steekproef of het aantal casestudies minder belangrijk zijn dan de kwaliteit van de procedure van de steekproeftrekking gebeurt of de selectie van casestudies.

Literatuur

Ackoff, R. L. (1989). "From Data to Wisdom", *Journal of Applied Systems Analysis*, Vol. 16, pp. 3-9.

European Commission (1999). *The Means Collection – Evaluating socio-economic programmes*. Luxembourg: Office for Official Publications of the European Communities.

Fetterman D.M. (1997). *Empowerment Evaluation : knowledge and tools for self-assessment and accountability*. Beverly Hills, CA: Sage.

Fetterman D.M. Wandersman A. and Millet R.A. (2005). *Empowerment Evaluation Principles in Practice*. New York: Guilford.

Patton M.Q. (1997). *Utilization-focused Evaluation – The New Century Text*. Thousand Oaks, CA: Sage.

Scriven M. (1991) *Evaluation Thesaurus*. Thousand Oaks: Sage.

Weggeman M. (1998). *Kennismanagement: inrichting en besturing van kennisintensieve organisaties*. Schiedam: Lannoo, 221p.

5. Evaluatietypes

In dit hoofdstuk gaan dieper in op de mogelijk vraagstellingen en criteria om te evalueren. We doen dit aan de hand van een aantal basistypologieën van evaluatie die eerder al werden voorgesteld in hoofdstuk 2. De bedoeling is om de lezer in de eerste plaats een overzichtelijk kader aan te reiken om verschillende types van evaluatie te plaatsen. Ten tweede wil dit hoofdstuk een handreiking bieden de lezer vanuit een bepaald evaluatiemotief en bepaalde evaluatievragen en -criteria te gidsen naar een type van evaluatie en de bijhorende aanpak. De specifieke evaluatietechnieken komen uitgebreider aan bod in deel 3 van de handleiding.

Paragraaf één herneemt kort de basistypologieën van evaluaties die ruggengraat vormen voor het overzicht. We beklemtonen in het bijzonder de dwarsverbanden die bestaan tussen de typologieën.

Paragraaf twee gaat dieper in op het aspect beleidstheorie en de rol van beleidstheorie voor beleidsevaluatie.

In paragraaf drie stellen we een conceptueel raamwerk van de beleidstheorie voor. Dit raamwerk vormt een bijkomend ankerpunt om de verschillende mogelijke vraagstellingen in een evaluatie te plaatsen.

Paragraaf vier en vijf gaan respectievelijk dieper in op vormen van ex ante en ex post evaluatie. Telkens wordt per type van evaluatie volgende vragen beantwoord. Wat is de opzet en de centrale vraagstelling in deze evaluatie? Waarin ligt de meerwaarde voor het beleid? Welke aanpak en evaluatietechnieken zijn aangewezen? Wat zijn de kritieke slaagfactoren? Korte voorbeeldjes zullen diverse aspecten illustreren. Voor uitgebreide cases verwijzen we de lezer naar boekdeel 4 van de handleiding.

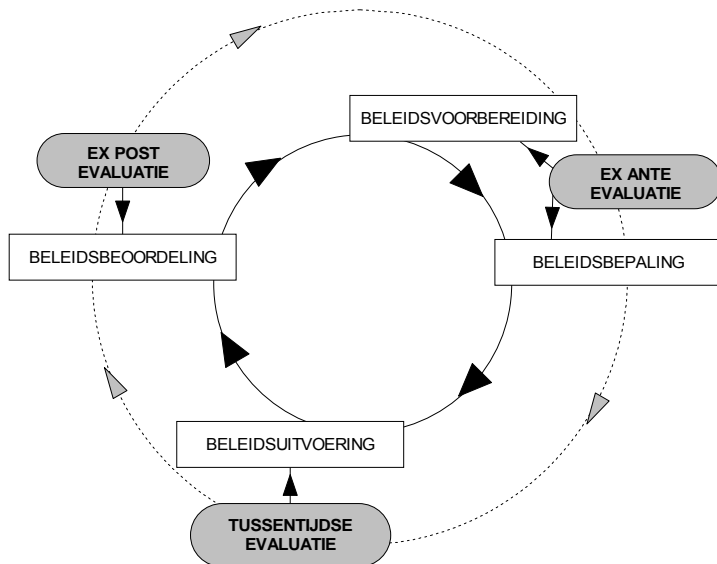
5.1 Typologieën en dwarsverbanden

In hoofdstuk 2 werden een aantal basistypologieën van beleidsevaluaties uitgewerkt. We hernemen ze kort. Vervolgens duiden we de dwarsverbanden. Deze dwarsverbanden vormen de ruggengraat voor het overzicht van evaluatietypes in paragrafen 4 en 5.

- Positie in de beleidscyclus: het tijdsperspectief

Eén van basisideeën rond beleidsevaluatie waarop deze handleiding steunt, heeft betrekking op de situering van evaluatie in de beleidscirkel of -cyclus. We breken in het bijzonder met de klassieke bestuurskundige opvatting dat evaluatie de allerlaatste fase is in de beleidscyclus, na de fasen van

beleidsplanning en beleidsuitvoering. Onze alternatieve opvatting stelt daartegenover dat beleidsevaluatie moet gedacht worden als een mogelijk, nuttig en noodzakelijk onderdeel of proces doorheen de ganse beleidskring. Met andere woorden, evaluatie is niet beperkt tot de laatste fase van beleidscyclus.



Figuur 12: Evaluatie gepositioneerd in de beleidscyclus

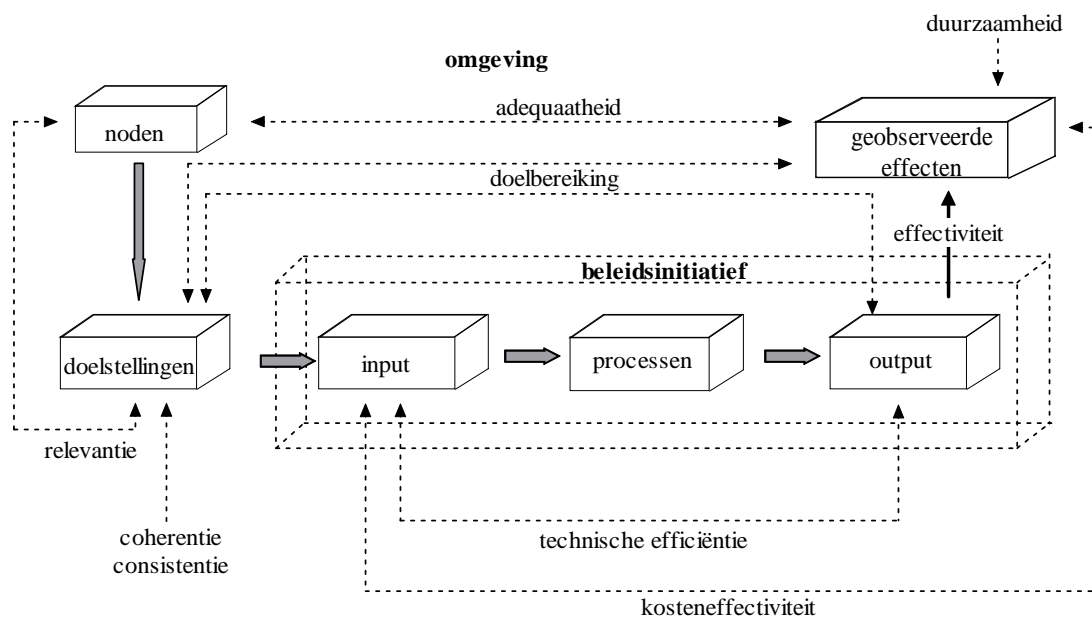
Afhankelijk van wanneer de evaluatie plaats heeft en op welke fase van de beleidscyclus ze betrekking heeft, spraken we van ex ante evaluatie, tussentijdse evaluatie of ex post evaluatie.

- Focus

Beleidsbeoordelingen kunnen ook ingedeeld worden volgens de algemene focus die ze hebben. We onderscheiden daarbij twee categorieën: evaluatie van de beleidsinhoud en evaluatie van de beleidsimpact. Een evaluatie van de beleidsinhoud gaat dieper in op aspecten zoals het concept, de logica, de instrumenten en de processen van het beleid. Een evaluatie van de beleidsimpact gaat dieper in op de veranderingen die het beleid al dan niet teweeg brengt in de maatschappelijke omgeving: de effecten op korte en lange termijn.

- Evaluatiecriterium

Een derde opdeling van evaluaties kan gebeuren aan de hand van het evaluatiecriterium dat centraal staat voor de afvoering van het beleid. Een groot aantal van deze generieke evaluatiecriteria kunnen afgeleid worden uit de schematische voorstelling van de beleidscyclus van de overheid als een open systeem model.



Figuur 13: Evaluatiecriteria doorheen de beleids- en beheerscyclus

Tussen de types uit de verschillende indelingen bestaat geen 1-op-1 relatie. Wel kunnen evaluatietypes geclusterd worden. Er bestaan met andere woorden dwarsverbanden tussen de verschillende typologieën. Deze dwarsverbanden maken het mogelijk om een soort van stroomdiagram op te stellen. Dit stroomdiagram biedt een leidraad in het vinden van het relevante type van evaluatie in functie van de evaluatienoden. Op de volgende pagina wordt dit stroomdiagram voorgesteld. De grondstructuur is gebaseerd op volgend verband:

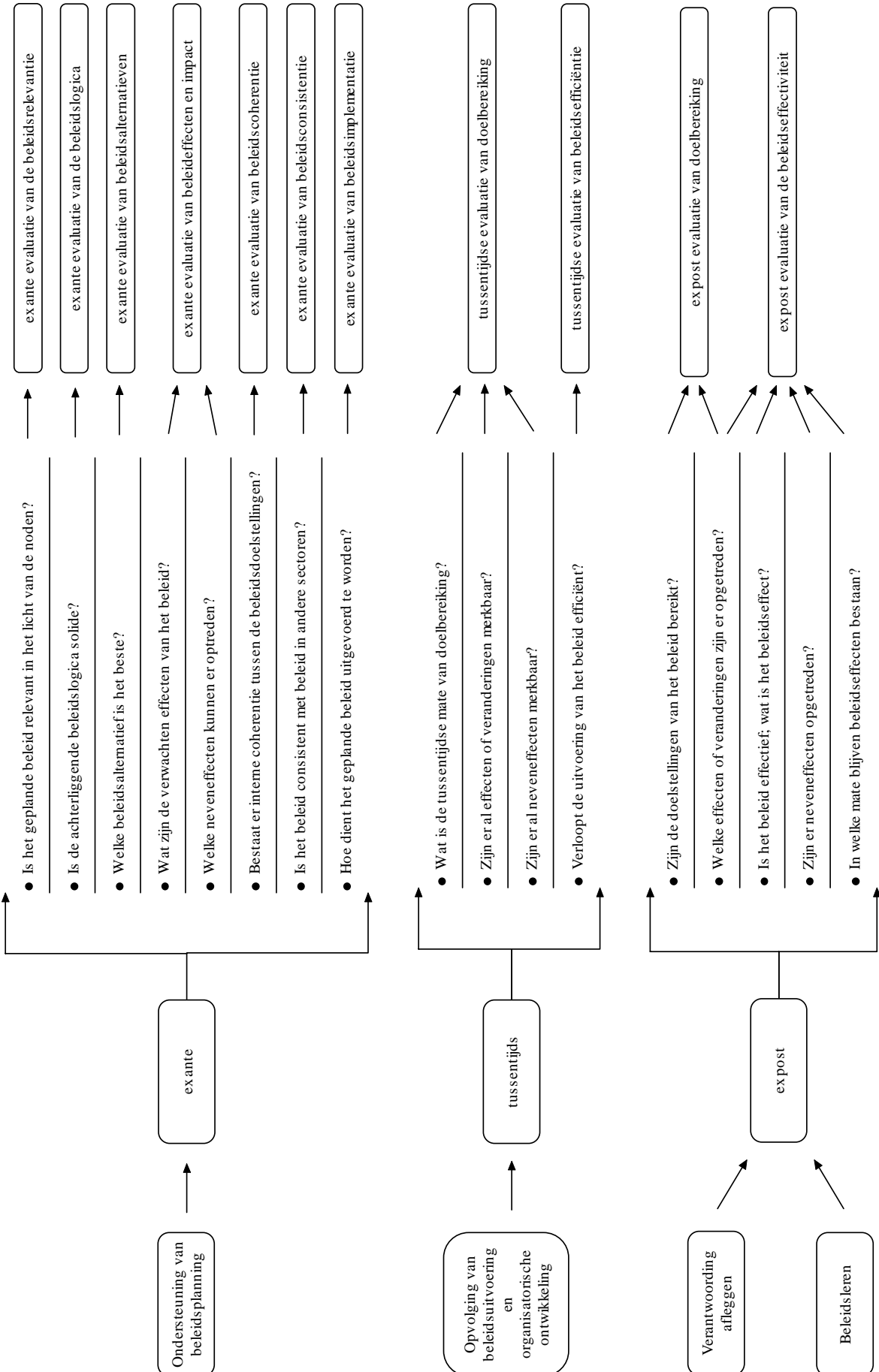


Figuur 14: Sequens voor de selectie van een evaluatietype

Het vertrekpunt vormt een van de vier evaluatiemotieven. Het evaluatiemotief hangt samen met een bepaalde fase in de beleidscyclus. Op basis daarvan krijgt de evaluatie een tijdsperspectief opgekleefd. Een volgende belangrijke stap is dan de selectie van evaluatievragen en -criteria. Naargelang de fase in de beleidscyclus en dus het tijdsperspectief van de evaluatie, zullen bepaalde vragen en criteria centraal staan. Belangrijk is dat bepaalde vragen en criteria uit meerdere tijdsperspectieven behandeld kunnen worden. Toch zullen er belangrijke verschillen zijn: een ex ante evaluatie van de verwachte beleidseffecten vereist een andere aanpak dan een ex post evaluatie van de beleidseffecten. Uit de combinatie van evaluatiemotief, het tijdsperspectief, de evaluatievraag en het bijhorende evaluatiecriterium, zal een bepaald evaluatietype kunnen worden afgeleid.

Elk evaluatietype veronderstelt een specifieke aanpak en technieken. Het verband tussen het evaluatietype en de aangepaste aanpak en technieken komt aan bod in de paragrafen vier en vijf.

Figuur 15: Stroomdiagram voor het afleiden van een evaluatietype



5.2 De rol van beleidstheorie voor evaluatie

Ongeacht het type van evaluatie, vinden we in de brede waaier van benaderingen van de evaluatiepraktijk een vaak terugkerend element terug: de aandacht voor beleidstheorie. Bepaalde auteurs zoals Chen en Donaldson hebben het concept zelfs centraal geplaatst in hun zogenaamde beleidstheorie-gedreven evaluatiebenadering. Beleidstheorie vormt daarin het vertrekpunt voor het design van een evaluatie: de afbakening van de inhoudelijke scope en keuzes omtrent de volgende stappen in het evaluatieproces. Gelet op de wijdverspreide en toenemende aandacht hiervoor, staan we in deze paragraaf stil bij de definitie en het profiel van beleidstheorie en vervolgens bij de rol ervan in evaluaties.

Wat is een beleidstheorie? In de context van beleidsevaluatie verwijst deze term naar de achterliggende redenering van beleidsmakers en/of andere stakeholders met betrekking tot een concreet beleidsinitiatief. Chen heeft baanbrekend werk geleverd in het duiden van het belang van beleidstheorie en de betekenis ervan voor het ontwerpen en uitvoeren van evaluaties. Volgens zijn definitie omvat een beleidstheorie twee aspecten:

- een beschrijving van een concreet beleidsinitiatief en van de wijze waarop het bepaalde doelstellingen zal bereiken of effecten zal genereren
- een standpunt over wat er gedaan moet worden om deze doelstellingen te bereiken.

Een beleidstheorie is met andere woorden tegelijk beschrijvend en prescriptief van aard. De beschrijvende assumpties slaan op de causale mechanismen die ten grondslag liggen van het maatschappelijke probleem en op de causale processen waardoor het beleid werkt en daarmee de beoogde verandering optreedt. Met andere woorden, een beleidstheorie geeft relevante factoren aan waarop het beleid kan inspelen om de gewenste effecten te bereiken. De prescriptieve assumpties slaan op de uitvoering van het beleid: ze geven aan welke middelen en processen noodzakelijk zijn om het beoogde veranderingsproces op te starten.

Kenmerkend voor veel beleidstheorieën bovendien is dat ze impliciet zijn. De redenering en de veronderstellingen achter een bepaald beleidsinitiatief zijn niet steeds helemaal uitgesproken of neergeschreven in beleidsdocumenten.

De term beleidstheorie in de context van evaluatie heeft dus niet de betekenis van een bestuurskundige theorie over beleid.

Wat is het belang van een beleidstheorie? Een niet doordachte of onvoldoende gefundeerde beleidstheorie wordt als een belangrijke faalfactor beschouwd. Een accurate achterliggende beleidstheorie is van essentieel belang met het oog op een efficiënt en effectief beleid. De mate waarin de beschrijvende veronderstellingen geldig zijn, bepaalt immers in belangrijke mate de effectiviteit van het beleid. De kwaliteit van de prescriptieve veronderstellingen bepaalt dan weer mee de efficiënte organisatie van de uitvoering van het beleid.

Het expliciteren van de beleidstheorie is dan ook een belangrijk aandachtspunt. In een ex ante evaluatie ter ondersteuning van de beleidsplanning kan de verfijning van de beleidslogica zelf het voorwerp of evaluandum zijn. Bij een ex post evaluatie dient het expliciteren van de beleidstheorie om het te evalueren beleidsinitiatief scherp in beeld te krijgen. Pas dan kan men beslissen over de focus en uitvoering van de evaluatie via de selectie van evaluatievragen, -criteria, indicatoren en evaluatietechnieken. Bij de verdere bespreking van de types van evaluaties komen we hier nog uitgebreider op terug.

Wat is het fundament van een beleidstheorie? Een beleidstheorie zal in veel gevallen gebaseerd zijn op inzichten en overtuigingen van beleidsmakers, praktijkdeskundigen en belangengroepen. Zij hebben bepaalde percepties en preferenties bij het ontwerpen van beleid, waarbij deze gebaseerd kunnen zijn op ervaring, conventionele kennis, discussie met collega's uit het veld, advies van experts, vertrouwde met een bepaalde aanpak en zelfs intuïtieve ideeën. Daarin ligt een reden waarom deze beleidstheorieën vaak impliciete theorieën zijn. Men noemt ze daarom soms ook stakeholdertheorieën. Deze laatste term verwijst in deze context niet naar wetenschappelijke theorieën over de invloed en het gedrag van stakeholders, maar wel naar de ideeën en assumpties die stakeholders hebben met betrekking tot een beleidsinitiatief.

Een beleidstheorie kan ook gebaseerd zijn op bredere wetenschappelijke theorieën. Dergelijke theorieën zijn doorgaans empirisch getest. Het hanteren van deze wetenschappelijk onderbouwde kaders of modellen voor het ontwerpen van beleid kan een moeilijke zoektocht naar relevante factoren dus vermijden. Beleidsmakers baseren zich dan op wetenschappelijke inzichten in de redenen waarom een gegeven factor als hefboom voor het bereiken van de gewenste effecten kan dienen. Waarom bijvoorbeeld, een sensibilisatiecampagne bijdraagt tot een attitude- of gedragsverandering.

Wanneer men een bepaalde aanpak van het beleid wetenschappelijk kan staven, zal het voorgestelde beleid sneller legitimiteit verwerven. Een mogelijk nadeel van wetenschappelijke theorie als beleidsgrondslag kan liggen in het feit dat ze onvoldoende de praktijk reflecteert. De wetenschappelijk

theorie is misschien gebaseerd op testen in een andere beleidscontext, of op resultaten met minder complexe beleid, bijvoorbeeld zonder te kijken naar de wisselwerking met beleidsinitiatieven binnen andere beleidssectoren.

In welke mate een beleidstheorie achter een concreet beleidsinitiatief vooral steunt op wetenschappelijke inzichten, dan wel op ervaringen en expertise van beleidsmakers, zal van geval tot geval verschillen. In een aantal sectoren zal er bijvoorbeeld nog niet voldoende wetenschappelijke kennis opgebouwd zijn om een innoverende beleidsaanpak te onderbouwen. Of beleidsmakers hebben positieve ervaringen met een bepaalde aanpak; ze zullen dan geneigd zijn om deze verder aan te houden in toekomstige initiatieven.

Auteurs, zoals Donaldson, die hun evaluatiebenadering volledig ontwikkelen rond het concept beleidstheorie, beklemtonen het potentieel dat zit in een kruisbestuiving tussen verschillende 'niveaus' van theorieën. Concrete beleidstheorieën, en evaluaties die daarop worden gebaseerd kunnen verrijkt worden met inzichten uit bredere sociaalwetenschappelijke theorieën. Maar ook vice versa kunnen concrete beleidstheorieën bijdragen aan de verfijning van de bredere wetenschappelijke theorieën, door kennis aan te dragen over hoe overheidsoptreden maatschappelijke verandering met zich mee brengt.

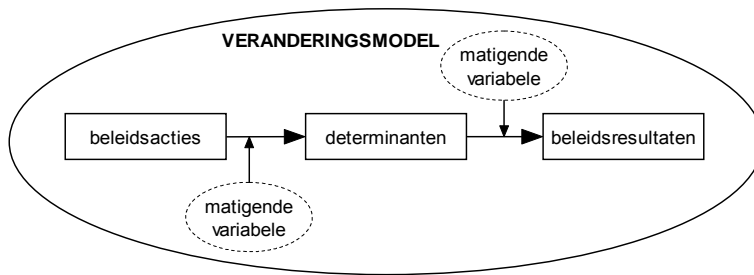
In een volgende paragraaf wordt het concept beleidstheorie uitgewerkt naar een conceptueel raamwerk. Dit kader biedt een bijkomende kapstok voor de lezer om de focus van een beleidsevaluatie te herkennen en te plaatsen.

5.3 Naar een conceptueel raamwerk van beleidstheorie

Volgens de eerder vermelde definitie bevat een beleidstheorie een beschrijving van de beoogde verandering die het beleid wil bewerkstelligen en richtlijnen voor de uitvoering van het beleid in functie daarvan. Dat betekent dat een beleidstheorie schematisch voorgesteld kan worden als een veranderingsmodel en een implementatiemodel. Het veranderingsmodel bevat de meer theoretische redenering achter de beleidsaanpak en de beoogde verandering, terwijl het implementatiemodel deze ideeën doorvertaalt naar praktijkgerichte procedures voor de uitvoering van het beleid. In deze paragraaf construeren we deze twee modellen in een algemeen raamwerk van de beleidstheorie en we staan we stil bij de afzonderlijke componenten.

5.3.1 Het veranderingsmodel

Het veranderingsmodel bevat drie componenten die in een causale keten worden geplaatst: beleidsacties, determinanten en beleidsresultaten. Daarnaast is er een vierde soort component: de modererende of matigende variabele. Onderstaande figuur geeft het veranderingsmodel weer.



Figuur 16: Het veranderingsmodel

De *beleidsacties* stemmen overeen met de output van het beleid: de concrete dienstverlening of producten die de uitvoering van het beleid genereert. De operationele beleidsdoelstellingen geven de output van het beleid aan. De beleidsacties zijn de aanzet of katalysator van het beoogde veranderingsproces.

Determinanten nemen in het veranderingsmodel een centrale plaats in. Ze zijn de weergave van de elementen waar men van veronderstelt dat het de hefboomen zijn om de beoogde verandering teweeg te brengen. De hypothese is dat wanneer het beleid deze hefboomen in beweging zet, de verwachte effecten gegenereerd zullen worden. Daarmee zijn de determinanten die men opneemt in het veranderingsmodel medebepalend voor de focus van het beleidsinitiatief. Dit is van belang omdat maatschappelijke problemen vaak verschillende aanleidingen en oorzaken hebben. Beleidsmakers dienen bijgevolg keuzes te maken en determinanten te selecteren waarvan men veronderstelt dat ze de belangrijkste oorzaken reflecteren en bijgevolg het meest bijdragen tot de verwachte effecten. Dergelijke keuze zal echter ook gebaseerd moeten zijn op een beoordeling van de haalbaarheid om deze determinanten beleidsmatig te beheersen of te beïnvloeden. Ze vormen een belangrijke schakels in de logica van het beleid; de brug tussen output en verwachte effecten.

De *beleidsresultaten* stemmen overeen met de verwachte (lees: gewenste) effecten van het beleid en reflecteren dus de strategische beleidsdoelstellingen. Ze zijn het eindpunt van het beoogde veranderingsproces. Men kan ze nog verder opdelen tussen resultaten of effecten op korte termijn en op lange termijn (impact).

In het veranderingsmodel zit dus een causale keten vervat: de fundamentele assumptie is dat wanneer de beleidsacties gerealiseerd worden door het uitvoeren van het beleid, de determinanten veranderd worden, waardoor op hun beurt de verwachte effecten gegenereerd worden. Toch kan deze causale keten in een gegeven beleidscontext meer of minder opgaan. Het veranderingsproces kan met andere woorden onderhevig zijn aan randvoorwaarden.

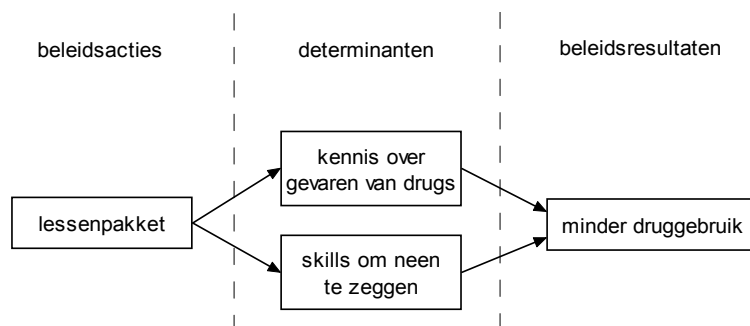
Deze randvoorwaarden worden aangegeven door de *modererende of matigende variabelen*. Voor een goed begrip: variabelen zijn factoren die een bepaalde waarde kunnen aannemen. Die waarde kan

kwantitatief (0, 99, enz.) zijn of kwalitatief (hoog, laag, enz.) zijn. De waarde is niet vast, maar kan variëren in de tijd, naar plaats, per persoon enz.

Modererende variabelen werken in op de causale relaties tussen beleidsacties, determinanten en beleidsresultaten. Ze beïnvloeden de sterkte van causale relaties tussen deze componenten; ze bevorderen of remmen af. Ze kunnen helpen verklaren waarom bepaalde oorzaakgevolg relaties wel of niet opgaan, of in mindere of meerdere mate dan verondersteld. Matigende variabelen zijn dan ook belangrijk met het oog op het ontwerpen van zogenaamde flankerende beleidsmaatregelen. Dit zijn maatregelen die als doel hebben om gunstige randvoorwaarden te scheppen waardoor de effectiviteit verhoogt van de maatregelen die de kern van het beleidsinitiatief vormen.

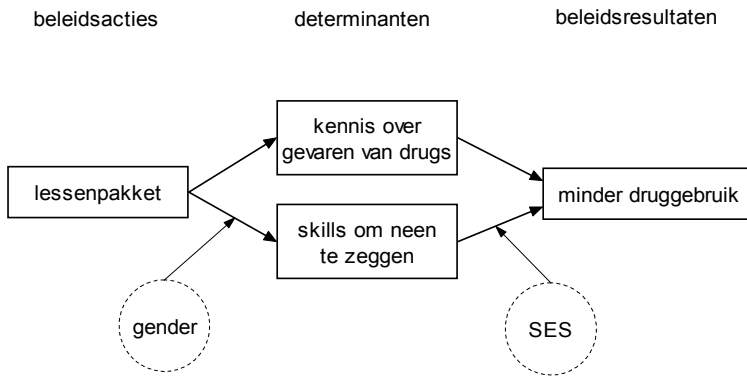
Bovenstaande elementen kunnen verduidelijkt worden aan de hand van enkele voorbeelden.

Stel dat een beleidsinitiatief wordt opgestart om kinderen te leren nee zeggen tegen drugs om zo het druggebruik in de scholen te laten dalen. In dit voorbeeld is een verminderd druggebruik het gewenste beleidsresultaat of effect en het lessenpakket is de beleidsactie of output. De beleidsmakers mikken op twee determinanten of tussenliggende variabele in het veranderingsmodel: een betere kennis over de gevaren van drugs voor de gezondheid en het versterken van de skills van de kinderen om nee te zeggen, ook onder 'peer pressure'.



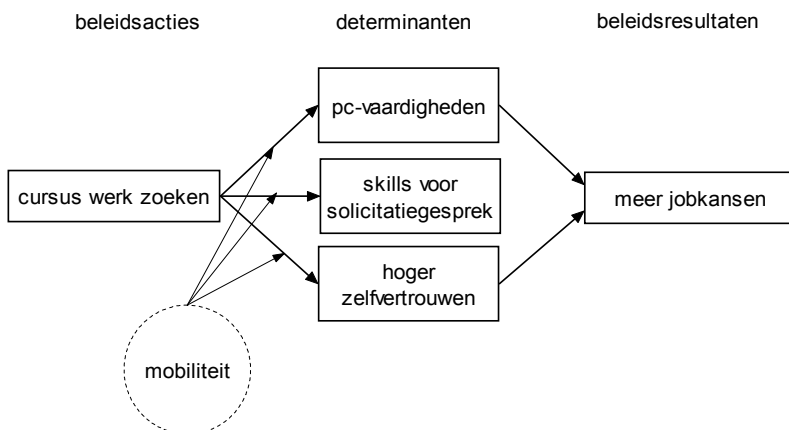
Figuur 17: Eenvoudig veranderingsmodel achter antidrugsbeleid

Echter, stel dat de mate waarin kinderen via het lessenpakket deze skills aanleren, verschilt. Er speelt een genderverschil jongens minder dan meisjes oor hebben naar de raad om drugs te weigeren. Stel ook men heeft vastgesteld dat kinderen met een lagere sociaaleconomische status relatief meer vatbaar blijven voor drugs. Het zijn twee modererende of matigende variabelen die men in het schema moet toevoegen.



Figuur 18: Uitgebreid veranderingsmodel achter antidrugsbeleid

Een tweede voorbeeld uit het werkgelegenheidsbeleid. Een begeleidingstraject om werklozen te herscholen en daardoor sneller en gericht aan een job te helpen, wordt opgestart. De cursus mikt op het verhogen van het zelfvertrouwen en het aanleren van pc-vaardigheden. Tevens wordt aan de hand van een rollenspel sollicitatiegesprekken gesimuleerd. Een matigende variabele in deze context kan de toegankelijkheid zijn van de opleidingslocatie. Bepaalde personen uit deze categorie kunnen beperkte vervoersmogelijkheden hebben, hetzij omdat ze kansarm zijn hetzij door een gebrek aan een openbaar vervoersaanbod. Wanneer iemand zeer moeilijk op de plaats van de opleiding kan geraken, is de kans groter dat de persoon in de loop van het begeleidingstraject afhaakt, waardoor het gewenste effect niet gehaald wordt. De mobiliteit van deelnemers is dus van invloed op alle relaties tussen de aangeboden cursus en de intermediaire effecten of determinanten.

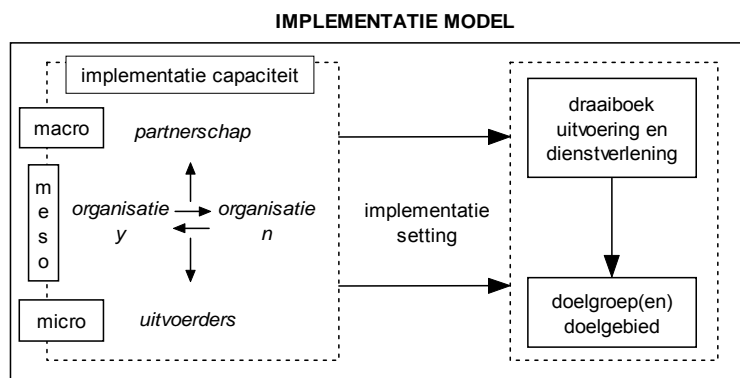


Figuur 19: Veranderingsmodel achter beleid in de sector werkgelegenheid

Beide voorbeelden tonen het belang aan van het identificeren van alle componenten van het veranderingsmodel; op deze wijze kan de logica van het beleid getoetst en verfijnd worden (cf. infra).

5.3.2 Het implementatiemodel

Het implementatiemodel geeft aan welke elementen van belang zijn voor het plannen van de uitvoering van het beleid. Deze elementen zijn de implementatiecapaciteit op micro-, meso- en macroniveau, de implementatiesetting, het draaiboek voor de dienstverlening en de beoogde doelgroep. Vaak krijgt de implementatie onterecht weinig aandacht in zowel de bestuurskundige theorie als bestuurlijke praktijk. Een succesvol beleid vergt evenzeer een accuraat implementatiemodel als een geldig veranderingsmodel.



Figuur 20: Het implementatiemodel

De *implementatiecapaciteit* verwijst naar een structuur en de actoren die instaan voor de diverse taken en functies in het kader van de beleidsuitvoering. Men kan deze capaciteit indelen naar drie niveaus. Ten eerste is er het individuele of microniveau. De kwaliteit van de uitvoering en daarmee de efficiëntie en effectiviteit van het beleid hangt in sterke mate af van de skills en expertise van individuele personen die taken vervullen in het kader van de beleidsuitvoering. Deze personen werken in één of meerdere organisaties die belast zijn met de implementatie. Dit resulteert in een capaciteit op organisatie- of mesoniveau. De betrokken organisatie als geheel moet over adequate capaciteit beschikken om haar uitvoerende functie waar te maken. Verschillende organisaties kunnen ook partnerschappen aangaan. Op dit macroniveau zijn taakverdeling en coördinatie belangrijke aandachtspunten.

Naast de structurele capaciteit is het uitvoeringsproces een belangrijke component in het implementatiemodel. De neerslag van het uitvoeringsproces kan men aangeven met de term *draaiboek van de dienstverlening*. Het is in wezen de beschrijving van de specifieke stappen met het oog op de dienstverlening of ontplooiing van maatregelen in het veld. Dergelijk vademecum bevat bijvoorbeeld de procedures voor de concrete toepassing van beleidsinstrumenten, richtlijnen over de omgang en communicatie met de doelgroep, de beschrijving van taken en verantwoordelijkheden, enz. Het kan ook bestaan uit een verzameling van documenten die de verschillende deelaspecten behandelen. Voor het

weslagen van de uitvoering van grootschalige of complexe beleidsinitiatieven kan een draaiboek of vademecum beschouwd worden als een kritieke succesfactor.

De *doelgroep* verwijst naar de verzameling van personen, administratieve entiteiten (openbare besturen, bedrijven, scholen) of natuurlijke elementen (dieren, planten, bosbestand, enz.) waarop het beleidsinitiatief zich richt. Vaak is deze doelgroep ook territoriaal afgebakend. We kunnen dan spreken van een doelgebied (bebouwde kom, provincie, land, natuurgebied enz.).

In het kader van een evaluatie zijn doorgaans drie assumpties van belang met betrekking tot de doelgroep: het bestaan van geldig vastgestelde selectiecriteria voor de afbakening van de doelgroep, de haalbaarheid om de doelgroep te bereiken. In de context van bepaalde beleidssectoren, bv. de gezondheidszorg, is ook de bereidheid van doelgroepleden om zich te engageren of zich minstens open te stellen ten aanzien van het beleidsinitiatief een belangrijke factor voor een doeltreffend beleid. Zo is bij behandelingen van patiënten met het oog op persoonlijke gedragsveranderingen het screenen van mogelijke patiënten (op fysieke en mentale gereedheid) en het kennen van de persoonlijke noden van groot belang.

Met de *implementatiesetting* verwijzen we naar de verzameling van een aantal contextuele parameters die belangrijke aandachtspunten zijn met het oog op een succesvolle uitvoering van het beleid. De implementatiesetting varieert naargelang het soort beleid. Ripley en Franklin maken bijvoorbeeld het onderscheid tussen verdelingsbeleid (publieke diensten en goederen), herverdelingsbeleid (bv. fiscaal beleid) en reguleringsbeleid (marktwerking en consumentenzaken). Anderen stellen deze indeling wel in vraag. Enerzijds omdat de categorie van organiek beleid (ontwerpen van instituties, bv. via een gemeentedecreet) ontbreekt. Anderzijds omdat in de praktijk voorbeelden bestaan van beleid dat in meerdere van de aangehaalde categorieën kan ingedeeld worden. Bovenstaande classificatie werd ontworpen om slaagfactoren in de context van beleidsuitvoering te identificeren, daarmee indirect verwijzend naar het belang van de kans op conflict en externe interferentie.

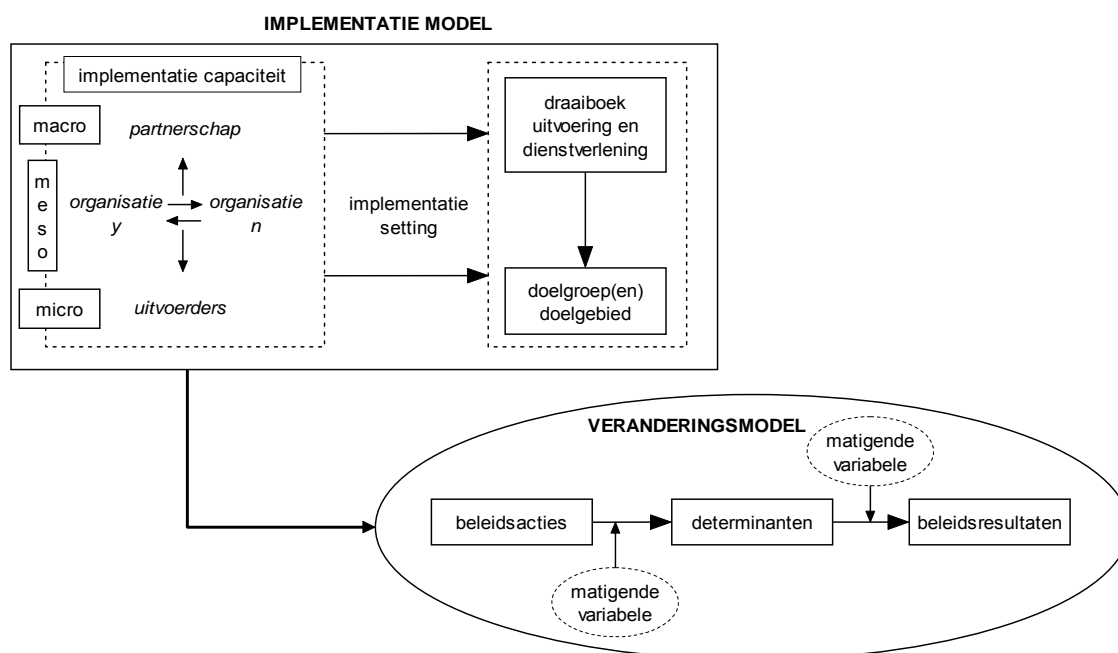
Een belangrijke parameter bij de implementatiesetting is het bestaan van aangepaste sturingsrelaties tussen de actoren verantwoordelijk voor de uitvoering onderling en de relatie tussen deze actoren en doelgroep van het beleid. Er zijn een globaal drietal types van sturingsrelaties mogelijk: hiërarchie, markt en netwerk. Afhankelijk van de concrete beleidscontext zal een bepaald type meer aangewezen zijn dan andere. Voor handhaving bijvoorbeeld is een hiërarchische sturing van de doelgroep vereist. De snelheidscontroles in het verkeer en bijhorende sanctionering zijn hiervan een voorbeeld. In de context van energiebeleid is het markttype belangrijk: consumenten moeten kunnen kiezen bij welke leverancier

ze klant zijn. Anderzijds zal de overheid in het kader van haar reguleringsbeleid een hiërarchische relatie hebben ten opzichte van de marktspelers aan de aanbodzijde. Zo kan ze bijvoorbeeld fusies verhinderen die de goede marktwerking in gevaar brengen. Op Europees vlak is de rol van de Commissie hierin welbekend.

In de context van welzijnsbeleid zal een netwerktype dan weer adequater zijn waarbij dienstverlener en doelgroep op een gelijke voet ten opzichte van elkaar staan. Werken met individuele personen en beleid gericht op gedragsveranderingen is een heel andere context dan een beleid gericht op natuurbeheer. Inwerken op persoonlijke gedragsveranderingen vergt vaak een andere aanpak naargelang het beleidsveld: in het welzijnsbeleid is de band met het individu veel directer bijvoorbeeld in vergelijking met afvalpreventiebeleid.

5.3.3 Het conceptueel raamwerk

Samen vormen het veranderingsmodel en het implementatiemodel het conceptueel raamwerk van de beleidstheorie met betrekking tot een standaard beleidsinitiatief. Onderstaande figuur geeft de link tussen beide modellen aan: het implementatiemodel zet het veranderingsmodel in beweging. Door de uitvoering van het beleid komt de causale keten van beleidsoutput over determinanten tot en met effecten op gang, waarbij matigende variabelen de sterkte van de causale relaties beïnvloeden.



Figuur 21: Conceptueel raamwerk van beleidstheorie

Het conceptueel raamwerk is een generiek schema is. Het heeft als voordeel dat het een holistisch perspectief biedt op het beleid, door zowel de implementatie als de beoogde werking van het beleid in

beeld te brengen. Natuurlijk kan een specifieke evaluatie inzoomen op ofwel het implementatiemodel ofwel het veranderingsmodel, en verder op bepaalde componenten van het model en de onderlinge relaties. Aan de hand van het conceptueel raamwerk kan men de focus van een evaluatie herkennen en plaatsen.

In het kader van de zogenaamde beleidstheorie-gedreven evaluatiebenadering vormt het raamwerk tevens een sjabloon voor een specifieke evaluatietechniek, meer bepaald de logische modellering. Met dit instrument kan men een beleidsinitiatief inhoudelijk uitklaren en/of de achterliggende logica (beleidstheorie) van gepland of uitgevoerd beleid verifiëren door deze visueel te (re)construeren. Met hetzelfde instrument kan men ook aanknopingspunten identificeren voor het ontwerpen van flankerend beleid (cf. supra: modererende of matigende variabelen).

De ex ante en ex post evaluatie van de beleidslogica komt verderop in dit hoofdstuk aan bod. De bijpassende evaluatietechniek, de logische modellering, wordt besproken in deel 3 van de handleiding.

In de volgende paragrafen gaan we dieper in op de verschillende types van evaluatie. Zoals eerder gesteld hanteren we een indeling gebaseerd op enerzijds het tijdsperspectief en anderzijds op de focus van de evaluatie (beleidsinhoud - beleidsimpact). Binnen deze categorieën zullen de verschillende evaluatiecriteria aan bod komen.

5.4 Ex ante evaluatie

5.4.1 Inleiding

Wanneer we de beleidscyclus als vertrekpunt nemen, kan beleidsevaluatie in de eerste plaats ondersteuning bieden in de fase van beleidsvoorbereiding en -bepaling. Centraal daarbij staat het evalueren van de meerwaarde van een beleidsinitiatief (maatregel, programma of beleid) dat nog in een planningsfase zit. We spreken in dit geval van een ex ante evaluatie omdat deze evaluatie de (eventuele) uitvoering van het beleid vooraf gaat in de tijd.

Kenmerkend voor dergelijke ex ante evaluaties is dat de finaliteit ligt in het bijdragen tot de ontwikkeling van een beleid en het beoordelen van een gepland beleid of van beleidsalternatieven op de meerwaarde. Dit is een wezenlijk verschil met een beoordeling van de meerwaarde van een beleid waarvan de uitvoering is voltooid (cf. infra).

Afhankelijk van de evaluatievraag en het bijhorende criterium zijn er verschillende subtypes van ex ante evaluatie te onderscheiden. Volgende klemtonen werden afgeleid:

- de *relevantie* of *geldigheid* van het beleid;
- de *interne logica* van het beleid;
- de *afweging van alternatieven* voor beleid;
- de *coherentie en consistentie* van het beleid;
- de *verwachte effecten en impact* van het beleid
- de *implementatie* van het beleid;

Merk op dat, naargelang het subtype, een ex ante evaluatie de link kan leggen naar de aanleiding, de conceptie, de uitvoering of de verwachte effecten van het geplande beleid. Hoewel het te evalueren beleid nog niet uitgevoerd wordt of werd, kan men prospectief al naar de uitvoering of naar de resultaten kijken. In dat laatste geval is de focus gericht op de beleidsinhoud, terwijl bij de andere subtypes de focus ligt bij de beleidsimpact.

De aangehaalde subtypes zijn ook complementair: een ex ante evaluatie kan meerdere evaluatiecriteria combineren. Dit verandert de facto niets aan de werkwijze: voor het beantwoorden van elke vraagstelling is een bepaalde aanpak aangewezen. Het aantal vragen en criteria heeft natuurlijk wel gevolgen voor de duurtijd en de kostprijs van een evaluatie.

In de volgende secties gaan we nader in op elk van subtypes van ex ante evaluatie. Telkens komen volgende aspecten aan bod, verrijkt met korte voorbeelden:

- opzet en vraagstellingen
- meerwaarde
- bijhorende benaderingen en technieken
- kritieke slaagfactoren voor de uitvoering

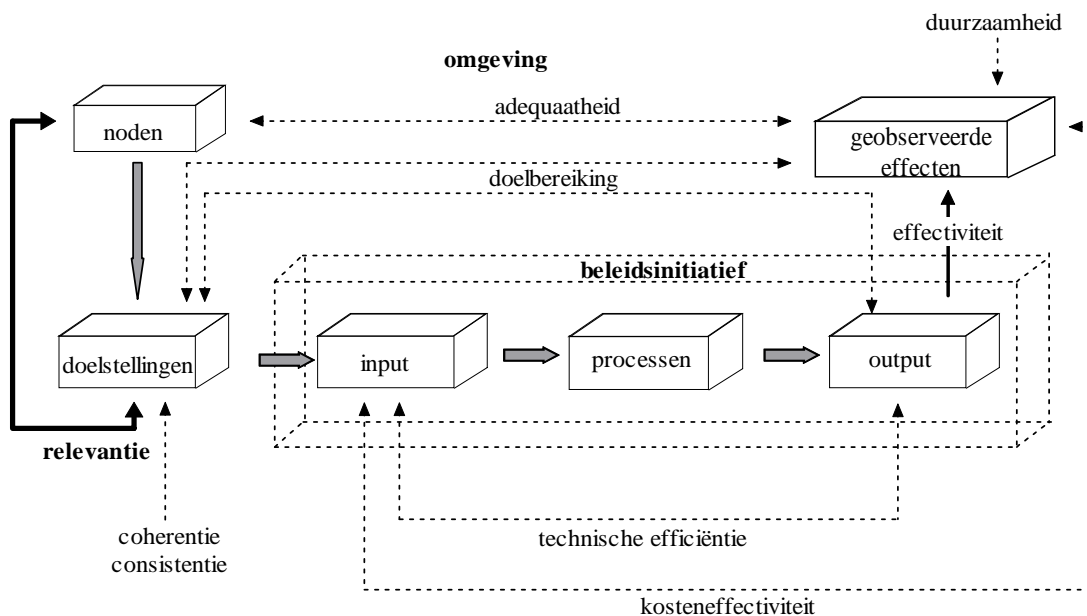
5.4.2 Relevantie van het beleid

5.4.2.1 Opzet en vraagstelling

Vanuit normatief oogpunt ligt de bestaansreden van een beleidsinitiatief van de overheid in het probleemoplossende vermogen, een probleemvermijdend vermogen en/of een pro-actief benutten van kansen met het oog op het genereren van een maatschappelijke meerwaarde.

Een ex ante evaluatie kan focussen op een maatschappelijke probleemstelling, het antwoord van het beleid en de relatie tussen beide. De evaluatie zal de gepercipieerde problemen en behoeften ‘an sich’ niet in vraag stellen. Datgene wat erkend wordt als een maatschappelijk probleem of sociale nood behoort immers tot het primaat van de politiek en wordt afgeleid uit een visie over de rol van de overheid. Een vraagstelling die wel centraal kan staan in een ex ante evaluatie is: “Is het geplande beleid of zijn de beleidsdoelstellingen relevant in het licht van de gegeven en geformuleerde probleemstelling?”.

Met relevantie als het voornaamste evaluatiecriterium waarop getoetst zal worden, staat het nut van beleid centraal. Hierbij focust het evalueren zich op de link tussen probleemstelling en de formulering van doelstellingen. De hier bedoelde evaluatie richt zich dus met andere woorden op het aangrijpingspunt van de beleidscyclus: de vraag waarom een gegeven beleidsinitiatief wordt gepland en of de doelstellingen van het beleid een accuraat antwoord vormen op de maatschappelijke vraag of nood. Onderstaande figuur geeft de klemtoon van dergelijke ex ante evaluatie weer.



Figuur 22: Evaluatie van de beleidsrelevantie

5.4.2.2 Meerwaarde

Bepaalde van de voorwaarden waaraan voldaan moet zijn om te kunnen spreken van een geslaagd beleid kunnen slechts na de uitvoering van het beleid worden getoetst. Ze liggen vervat in de criteria van doelbereiking, effectiviteit en adequaatheid. In een ex post evaluatie kan men inderdaad nagaan of de vooropgestelde doelstellingen werden gehaald op operationeel niveau (output) en/of strategisch niveau (effect, impact), de mate waarin de *geobserveerde* effecten toe te schrijven zijn aan het beleid en in hoeverre de geobserveerde effecten tegemoet komen aan de maatschappelijke noden.

Maar in een ex ante evaluatie kan men wel al nagaan of de geformuleerde doelstellingen waarin normalerwijze de *gewenste* effecten vervat liggen, een accuraat antwoord zijn op de maatschappelijke vraag. Dit is niet zonder belang. Immers, wat is het nut van een beleid waarvan het vertrekpunt verkeerd is?

Een beleid dat niet relevant is in het licht van de maatschappelijke nood, zal niet leiden tot effecten die vanuit maatschappelijk oogpunt adequaat zijn, ook al kan achteraf een grote mate van doelbereiking en een sterke effectiviteit aangetoond worden.

De relevantie van het beleid is met andere woorden een *conditio sine qua non*. Op zichzelf is het nog geen voldoende voorwaarde, maar het is wel een noodzakelijke voorwaarde die vervuld moet zijn om te kunnen spreken van een geslaagd beleid. Zekerheid willen hebben over de relevantie van het beleid in het licht van de probleemstelling is dus een pertinente aspiratie.

Een ex ante evaluatie van de relevantie van het beleid past dus onder één van de twee fundamentele vraagstellingen van de overheid, namelijk: "Doen we de goede dingen?".

De meerwaarde van een evaluatie op het relevantie criterium kan zowel zinvol zijn bij het opstarten van een nieuw beleid als bij het voortzetten van een bestaand beleid. Vooraleer het (voorgezette) beleidsinitiatief daadwerkelijk wordt uitgevoerd, kan men het nog bijsturen in functie van het (veranderd) probleem of (veranderde) maatschappelijke nood.

De meerwaarde ligt zowel op politiek als op administratief vlak. Vanuit politiek oogpunt is de relevantie van het beleid is een belangrijke legitimerende factor. Het kan onzekerheid wegnemen bij beleidsmakers en andere belangrijke stakeholders, zowel over de aard van de problematiek, de geschiktheid van de beoogde maatregelen als over de aanvaardingskans bij de doelgroep van het beleid. Zekerheid over het nut van het beleid verhoogt het draagvlak voor het beleid. Vanuit administratief oogpunt kan men via een ex ante evaluatie van de beleidsrelevantie een verspilling van middelen en tijd vermijden.

Een toetsing op het relevantie criterium past dus onder verschillende evaluatiemotieven. Het kan niet enkel beleidsplanning gericht ondersteunen en gepland beleid verbeteren. Tevens kan men het ook beschouwen als een voorschot op een (ex post) verantwoording.

5.4.2.3 Aanpak en technieken

De algemene aanpak om een grotere zekerheid over de relevantie van het beleid te verkrijgen, is het verzamelen van beleidsvoorbereidende informatie. Empirische gegevens over de maatschappelijke problematiek of de noden op het terrein kunnen een diagnose onderbouwen als vertrekpunt voor beleidsbepaling. Dergelijke informatie kan dus richtinggevend zijn voor de selectie van doelstellingen en doelgroepen.

Wat zijn de bronnen voor beleidsvoorbereidende informatie? Ten eerste kan men deze halen uit bestaande en reed beschikbare gegevens (secundaire bronnen): bijvoorbeeld demografische gegevens uit volkstellingen, of meer specifieke gegevensbestanden bijgehouden door de administratie.

Ten tweede kan men ook nieuwe gegevens verzamelen in het kader van een behoefteonderzoek of een nadere verkenning van de kenmerken van de potentiële doelgroep. Dergelijke onderzoeken bestuderen bijvoorbeeld de culturele gewoonten en kenmerken van de doelgroepleden of peilen naar de opinie en verwachtingen van toekomstige klanten van dienstverlening over geplande doelstellingen en maatregelen.

Op het strategische niveau kan ook beleidsvoorbereidende informatie verzameld en verwerkt worden aan de hand van een analyse waarbij zowel de sterkten en zwakten als de kansen en bedreigingen met betrekking tot de context van een gepland beleidsinitiatief in kaart worden gebracht. Vervolgens worden deze gegevens gestructureerd en gekoppeld met het oog op de ontwikkeling en toepassing van een algemene strategie. Deze algemene strategie bestaat erin om de sterktes te valoriseren, de zwaktes te verminderen of weg te werken, met het oog op het maximaal benutten van de kansen en het afwenden van de bedreigingen.

Aan de hand van dit analysekader kan men de focus van het beleid mee bepalen, maar ook gepland beleid evalueren op de relevantie. Men kan nagaan in hoeverre de strategische doelstellingen een antwoord vormen dat strookt met de hoger beschreven generieke strategie. De hiervoor aangewezen techniek is een SWOT-analyse.

- SWOT-analyse

De meeste technieken die voor een ex ante evaluatie op beleidsrelevantie aangewezen zijn, hebben als voornaamste functie het verzamelen van gegevens. Informatie over de (potentiële) doelgroep kan men verkrijgen via volgende technieken:

- interview

- focusgroep
- documentanalyse
- schriftelijke vragenlijst
- etnografische en participerende observatie

De meeste bovenstaande technieken worden verder toegelicht in deel 3 van de handleiding.

5.4.2.4 Voorbeelden

Het gebruik van beleidsvoorbereidende achtergrondinformatie in een ex ante evaluatie kan worden geïllustreerd aan de hand van volgend voorbeeld. In een zuiders land werd beslist een programma op te starten om leerlingen in het basisonderwijs aan te leren hoe ze huidkanker kunnen voorkomen. Alvorens het geplande beleidsinitiatief uit te voeren werd een ex ante evaluatie gedaan met het oog om de achtergrond van de doelgroep beter te leren kennen, om eventueel de beleidstheorie en het uitgetekende veranderingsmodel bij te sturen. De beleidsmakers wilden meer weten over de ideeën en activiteiten van de leerlingen met betrekking tot bescherming tegen de zon en huidkanker. Meer bepaald wilden ze de respons van de doelgroep op verschillende pedagogische hulpmiddelen en beschermingsvoorschriften doorgronden. De gegevensverzameling gebeurde aan de hand van groepsdiscussies, focusgroepen en interviews met de kinderen, hun ouders en personen met een staffunctie in organisaties actief in de recreatiesector. Uit de informatie bleek dat de praktijk van het zich beschermen tegen de zon inconsistenties vertoonde, hoewel er een wijdverspreid bewustzijn was in verband met huidkankerpreventie. Echter, uit het onderzoek bleek dat kinderen niet begrepen wat huidkanker was en dat ze het niet als een bedreiging beschouwden. Deze informatie bevestigde de relevantie van de vooropgestelde doelstellingen van het geplande beleidsprogramma.

Een ex ante evaluatie van de beleidsrelevantie kan ook een ruimere scope krijgen, wanneer men naast de doelstellingen ook de implementatiestrategie toetst aan de beleidsvoorbereidende informatie. In het hoger beschreven voorbeeld hielp de evaluatie de beleidsmakers bij de keuze van determinanten en het vinden van een accurate implementatiestrategie. Uit de informatie bleek immers dat de leerlingen afkerig stonden tegenover het dragen van lange broeken, kledij met lange mouwen en breedgerande hoeden om zichzelf tegen de zon te beschermen. De boodschappen tijdens een sensibilisatiecampagne die oproepen om deze soort kledij te dragen zouden met andere woorden als te extreem overkomen bij de doelgroep, te extreem om ze te accepteren. Een betere boodschap, volgens de ex ante evaluatie, richtte zich op een graduele verandering en aanpassing in de klederdracht. Uit gesprekken met de ouders en het personeel uit de recreatiesector bleek ook hun zorg dat veranderingen die het beleid nastreefde acceptabel zouden zijn binnen de lokale cultuur met een voorkeur voor lichte klederdracht die over

generaties heen tot stand was gekomen. De volwassenen bleken echter het geplande beleid te steunen dat erop gericht was om hun kinderen vaardigheden aan te leren om zich te beschermen tegen de zon. Daaruit bleek voor de beleidsmakers het belang van creatieve en wervende boodschappen die deels in recreatiefaciliteiten zouden worden verspreid en deels moesten overgedragen worden in familiekring.

Dit praktijkvoorbeeld toont aan dat het verzamelen van beleidsvoorbereidende achtergrondinformatie over de doelgroep de beleidsverantwoordelijken niet alleen helpt om de relevantie van het beleid te toetsen, maar ook nuttig kan zijn om de beleidslogica bij te sturen. In een volgende sectie wordt verder ingegaan op de beleidslogica als centraal aandachtspunt van een ex ante evaluatie.

5.4.2.5 Kritieke slaagfactoren

- timing en afstemming met het beslissingsproces
- politieke openheid om de focus van de beleidsdoelstellingen in vraag te stellen met het oog op het vergroten van de relevantie van het beleid en daarmee de kansen op daadwerkelijke impact
- beschikbaarheid van beleidsvoorbereidende informatie

5.4.3 Interne logica van het beleid of de beleidstheorie

5.4.3.1 Opzet en vraagstelling

Een tweede mogelijke vraagstelling bij een ex ante evaluatie op de beleidsinhoud richt zich op de robuustheid van de logica in de achterliggende beleidstheorie. Niet enkel de strategische en operationele doelstellingen zijn in beeld, maar ook het verband tussen beide. Daarmee staat dus het conceptueel raamwerk van de beleidstheorie in het brandpunt van de evaluatie, en in het bijzonder het veranderingsmodel. De strategische doelstellingen verwijzen naar de beleidsresultaten, de operationele doelstellingen verwijzen naar de beleidsacties.

Zoals uiteengezet in paragraaf twee, vormen determinanten een belangrijke tussenliggende component in het veranderingsmodel dat in de beleidstheorie vervat zit. De beleidstheorie draagt de fundamenten aan waarop het beleid wordt gebaseerd; assumpties over de causale processen waardoor de geplande beleidsacties leiden naar de gewenste resultaten. De beleidstheorie draagt dan ook argumenten aan voor het selecteren van determinanten die men met het beleid beoogt te beïnvloeden. Deze determinanten kunnen we zoals eerder vermeld beschouwen als hefboomen voor bepaalde causale processen. Ze vertegenwoordigen belangrijke schakels in het ontstaan van maatschappelijke problemen. Daardoor vormen ze de aangrijpingspunten voor een pro-actief of reactief beleid.

Een ex ante evaluatie van de beleidslogica verifieert in het bijzonder de (impliciete of expliciete) kernredeneringen achter het geplande beleid. Het expliciteren van de sequens en causale keten van

beleidsacties over determinanten naar resultaten weerspiegelt dus het proces van beleidsplanning vanuit een evaluerende invalshoek.

5.4.3.2 Meerwaarde

Beleidsverantwoordelijken en belangrijke stakeholders kunnen 1001 ideeën hebben voor hun beleid. Echter, vaak is het niet zo evident om deze ideeën te verduidelijken en onderling op een systematische en coherente manier te verbinden. Bij dit type van ex ante evaluatie staat de kern van de beleidsbepaling zelf in het brandpunt.

Met dergelijke toetsing kan men het beleid in een vroeg stadium, d.w.z. nog vóór de uitvoeringsfase, nog bijsturen met het oog op een maximale beleidseffectiviteit. Deze bijsturing kan zowel gericht zijn op het veranderingsmodel achter het beleid als op het bijhorende implementatiemodel.

Een groot deel van de meerwaarde ontleent dergelijke ex ante evaluatie dan ook precies aan haar timing ten aanzien van de beleidscyclus. Paradoxaal is het dat net deze timing ingaat tegen de intuïtieve of traditionele opvattingen over de beleidscyclus. Ex ante evaluatie krijgt weinig of geen aandacht in het korte termijn perspectief dat vaak kenmerkend is voor de dynamiek van politieke besluitvorming.

Een ex ante evaluatie van de beleidslogica past eveneens onder de essentiële vraag: “Doen we de goede dingen?”. Net als de toetsing op relevantie is dit type evaluatie gericht op het ondersteunen van de planning van het beleid en het verbeteren van gepland beleid met het oog op een maximale effectiviteit ervan. Het hoeft geen betoog dat een weldoordachte beleidslogica meer slaagkansen heeft.

5.4.3.3 Aanpak en technieken

Dit subtype van evaluatie zal vaak intern worden uitgevoerd, hetzij door de beleidsmakers zelf, hetzij in samenspraak met een interne evaluator in een aparte functie. Deze ex ante evaluatie is dan ook sterk verbonden met het proces van beleidsplanning zelf. Sterker nog, ingeval ze intern wordt uitgevoerd, zal men meestal geen formele opdeling maken tussen planning en deze vorm van ex ante evaluatie. De redenering achter het beleid toetsen op zwakke assumpties of eventuele hiaten gebeurt dan als onderdeel van het planningsproces. Het feit dat er geen duidelijke scheidingslijn bestaat met de beleidsplanning zelf, betekent echter niet dat beleidsmakers automatisch hun beleidstheorie aan een systematische, kritische toetsing onderwerpen.

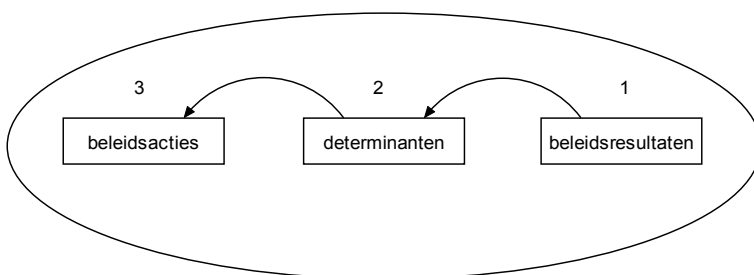
Anderzijds is het niet uitgesloten dat een externe evaluator de beleidslogica onder de loep neemt. Meestal zal dit echter kaderen in een ex post evaluatie waarbij een eerste stap bestaat uit het in kaart brengen van het uitgevoerde beleid. Dit vormt dan de opstap naar het bepalen van de scope van de evaluatie: men zal bepaalde causale relaties eruit pikken om nader te onderzoeken en te verifiëren in welke mate ze daadwerkelijk gerealiseerd zijn.

Bij het uitwerken van de evaluatiestrategie om de logica achter het beleid te expliciteren en te verfijnen, dient ook een bepaalde werkwijze gekozen te worden. In essentie komt het erop neer om het veranderingsmodel te (re-)construeren waarbij de causale relaties tussen de beleidsacties, determinanten en beleidsresultaten één voor één onder de loep genomen worden en concreet gemaakt worden. Hierbij heeft men de keuze tussen drie opties: het zogenaamde *voorwaarts redeneren*, *het terugredeneren* of *de combinatie van voorwaarts en terugredeneren*.

Het terugredeneren is een benadering die start bij het veranderingsmodel in het conceptueel raamwerk van beleidstheorie. Vervolgens gaat men stap voor stap terug binnen het model en, wanneer ook dit deel uitmaakt van de evaluatie, vervolgens naar het implementatiemodel. Het redeneren gebeurt in 'achterwaartse' richting doordat het proces in de tegengestelde richting verloopt als de sequens die getoond wordt in het raamwerk aan de hand van de pijlen.

De volgorde van de vragen is dan als volgt:

- 1) Welke resultaten willen we bereiken met het beleid en voor wie?
- 2) Op welke determinanten moet het beleid focussen?
- 3) Met welke beleidsacties zullen we deze determinanten op een accurate manier kunnen beïnvloeden?

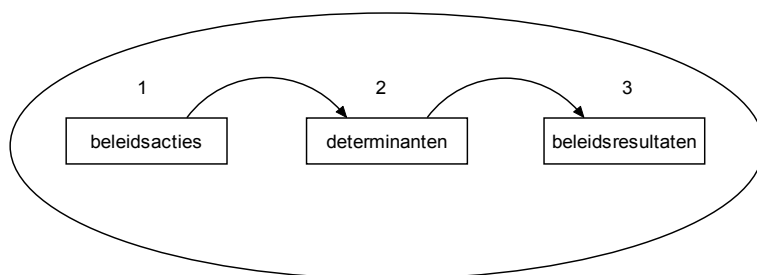


Figuur 23: Constructie van het veranderingsmodel via terugredeneren

Als ook het implementatiemodel wordt bekeken, dient men de vraag te stellen welke implementatiecapaciteit er vereist is op organisatie- en individueel niveau, of er nood is aan samenwerking en partnerschappen voor de uitvoering en hoe het uitvoeringsproces, c.q. de dienstverlening moet verlopen.

Het voorwaarts redeneren volgt wel het logische verloop binnen het veranderingsmodel. Wanneer ook het implementatiemodel wordt bekeken, ligt daar het vertrekpunt bij het voorwaarts redeneren en vervolgens schuift men op naar het veranderingsmodel. Het voorwaarts redeneren volgt met andere woorden de chronologie die ontstaat wanneer het theoretisch kader van de beleidstheorie in de praktijk wordt omgezet: het toepassen van het implementatiemodel zet immers het veranderingsmodel in beweging. Volgende vragen komen dan aan bod. In welke aspecten staan de uitvoerende organisaties sterk? Hoe kan de doelgroep bereikt worden? Zijn er daarbij bepaalde barrières of randvoorwaarden? Dit verwijst naar implementatiesetting: culturele gewoonten e.d. die een rol kunnen spelen. Binnen het veranderingsmodel volgt het voorwaarts redeneren de causale keten van beleidsacties over determinanten naar resultaten. De uitkomst is de formulering van doelstellingen die groeien uit initiële ideeën over het soort implementatiemodel dat vereist is. De volgorde van de vragen is dan:

- 1) Welke beleidsacties kunnen we voorzien?
- 2) Welke determinanten zullen veranderd worden door deze beleidsactie(s)?
- 3) Welke resultaten zullen we bekomen door deze determinanten te wijzigen?



Figuur 24: Constructie van het veranderingsmodel via voorwaarts redeneren

De twee benaderingen van voorwaarts en terugredeneren bij het ontwikkelen en expliciteren van de beleidstheorie, sluiten elkaar niet uit. De combinatie is tijdsintensiever, maar kan het beste van twee werelden samenbrengen. Men denkt dan afwisselend vooruit en achteruit doorheen de sequens in het beoogde veranderingsmodel (en het implementatiemodel). Generieke vragen die voortdurend terugkeren zijn dan:

- Waar willen we naartoe?
- Hoe kunnen we daar geraken?
- Wat gebeurt er als we dit doen?

Op basis van de wisselwerking tussen de antwoorden op deze vragen ontstaat er een ontwerp van beleidstheorie waarbij gelijktijdig de elementen van het veranderingsmodel en het implementatiemodel invulling krijgen. Zo kan men bijvoorbeeld starten met voorwaarts redeneren om het implementatiemodel op te stellen, vervolgens de techniek van terugredeneren gebruiken om het veranderingsmodel in te vullen, en ten slotte deze twee integreren om te komen tot een samenhangend geheel.

Deze duale aanpak is een goede keuze wanneer neveneffecten aan de orde zijn. Doordat de twee theorievormende procedures door elkaar worden gebruikt, zullen de betrokkenen alerter zijn voor eventuele gewenste of ongewenste neveneffecten. Dit is niet onbelangrijk omdat het om een ex ante evaluatie gaat en er al in fase van beleidsplanning aandacht ontstaat voor mogelijke neveneffecten van het beleid.

Bij de hoger beschreven aanpak kan ook een concrete techniek, logische modellering, worden gehanteerd. Werken met logische modellen kan de beleidstheorie en meer bepaald het beoogde veranderingsmodel achter het beleid schematisch en visueel voorstellen. In dat laatste schuilt het grote voordeel van deze techniek: het overleg tussen stakeholders en evaluator wordt vergemakkelijkt en de impliciete theorie achter een beleidsinitiatief wordt concreet gemaakt en letterlijk in beeld gebracht.

In paragraaf twee wezen we al op het belang van deze aanpak en techniek in het kader van de zogenaamde beleidstheorie-gedreven evaluatiebenadering. De aanpak is niet alleen in een ex ante evaluatie bruikbaar ter ondersteuning van de beleidsplanning en de verbetering van gepland beleid. Ook in een ex post evaluatie is het vaak aangewezen om de redenering achter het beleid te reconstrueren en op basis daarvan de evaluatievragen en -criteria te selecteren (cf. infra). In deel 3 komt de techniek van logische modellering uitgebreid aan bod.

➤ Logische modellering

Voor voorbeelden verwijzen we de lezer naar paragraaf twee.

5.4.3.4 Kritieke slaagfactoren

- De rol van de evaluator

We beklemtonen dat het doel van een ex ante evaluatie van de beleidslogica niet is om helemaal 'from scratch' het beleid uit te denken. De functie ervan ligt in het structureren van de ideeën van beleidsverantwoordelijken op een systematische en coherente manier. Zo kan men de achterliggende beleidstheorie, die vaak impliciet en fragmentair blijft, expliciteren en verfijnen. Desgevallend moet de rol van de interne of externe evaluator in dit proces dan ook duidelijk afgelijnd worden. Essentieel uitgangsprincipe is dat het eigenaarschap van de beleidstheorie steeds berust bij de beleidsverantwoordelijken en andere betrokken stakeholders. Wanneer een aparte evaluator wordt betrokken, heeft deze de rol van begeleider of 'facilitator'. Hij staat in voor

het efficiënte en doelgerichte verloop van de overlegmomenten tussen de beleidsverantwoordelijken en andere betrokken stakeholders waarbij zij trachten om hun ideeën over het beleidsinitiatief te verduidelijken en te verfijnen. Stakeholders kunnen naargelang de inhoudelijke scope en organisationele complexiteit van het beleidsinitiatief zeer diverse achtergronden, aandachtspunten en belangen hebben. Dergelijke diversiteit vergemakkelijkt het bereiken van consensus niet. De evaluator dient daarom aan te geven wat voor de groep de belangrijke discussiepunten zijn en waar welke stakeholder een bijdrage kan leveren op basis van de respectieve kennis, ervaring en ideeën. Op basis daarvan kan de evaluator vervolgens de discussies samenvatten en streven naar consensus.

Daarnaast kan de evaluator op basis van de eigen expertise in evalueren advies verlenen aan de stakeholders wanneer zij hierom vragen. De rol van de evaluator spitst zich toe op het structureren van de gesprekken, het aangeven van keuzemogelijkheden en opties zonder dat hij of zij eigen waarden naar voor schuift en oplegt. Anderzijds kan de evaluator, opnieuw vanuit eerdere evaluatie-ervaringen, wel ideeën naar voor brengen die voorwerp kunnen zijn van debat tussen de stakeholders.

- Aangepaste keuze van aanpak

Er zijn een drietal opties bij het (re-)construeren van de beleidstheorie: voorwaarts redeneren, terugredeneren en een combinatie van deze twee. Welke optie verkieslijk is, hangt af van een aantal randvoorwaarden. Volgende vuistregels kan men hierbij in acht nemen. Het terugredeneren werkt het best wanneer de beleidsmakers en andere belangrijke stakeholders vertrouwd zijn met sociaalwetenschappelijke methodologie en/of de inhoud van het beleidsveld of beleidsdomein. Deze elementen bevorderen het ontwerpen van het veranderingsmodel en het onderscheiden van de componenten van dit model. Bij de techniek van het terugredeneren vormt dit veranderingsmodel het vertrekpunt. Kennis over het beleidsterrein biedt sterktes op het vlak van het formuleren van doelstellingen. Dergelijke competenties vergemakkelijken ook het beantwoorden van opvolgingsvragen bij het terugredeneren, namelijk naar de oorzaken van een probleem en naar wat de meest doeltreffende beleidsacties zijn.

Het terugredeneren, volgt ook het gekende stamien van strategische beleidsplanning waarbij men eerste strategische doelstellingen (cf. beleidsresultaten of effecten) en daarna de meer operationele doelstellingen (cf. beleidsacties of output) formuleert.

Wanneer bovenstaande randvoorwaarden beperkt of niet vervuld zijn, is de aanpak van voorwaarts redeneren meer aangewezen. Het is immers van belang dat de conceptualisering van het beleid start met een aandachtspunt waarover stakeholders vlot kunnen meedenken en meespreken. Voorwaarts redeneren start vanaf de meer tastbare componenten van het

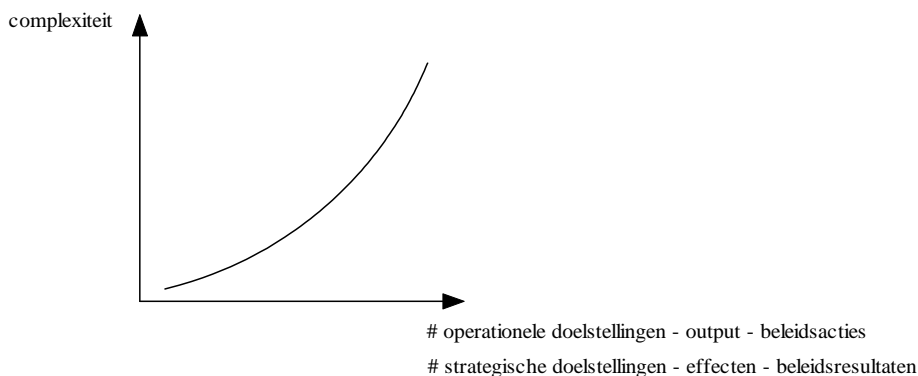
implementatiemodel. Op deze wijze wordt de beleidstheorie opgebouwd of verfijnd door te starten vanaf concrete knelpunten.

De combinatie van voorwaarts en terugredeneren, kan de voordelen van beide werkwijzen bundelen.

5.4.4 Coherentie en consistentie van beleid

5.4.4.1 Opzet en vraagstelling

De mate van complexiteit van een beleidsinitiatief vanwege de overheid kan sterk variëren. Dit hangt samen met de scope van het beleid. Onderstaande figuur schetst de (exponentieel) stijgende complexiteit naargelang het aantal doelstellingen. Of vertaald naar het veranderingsmodel: naargelang het aantal beleidsacties op operationeel niveau (output) en beleidsresultaten op strategisch niveau (effecten).



Figuur 25: Complexiteit van beleid

De meest eenvoudige situatie is het ontwerp van één maatregel in het kader van een bepaalde doelstelling. Vaak echter werkt men een pakket van maatregelen uit om eenzelfde doelstelling te bereiken. Een belangrijk aandachtspunt is dan het onderling afstemmen van de verschillende maatregelen in het pakket. Zij dienen complementair te zijn en/of elkaar te versterken. Tegenstrijdige maatregelen, dit wil zeggen met tegengestelde effecten, dienen vermeden.

Wanneer dat maatregelenpakket werd samengesteld in functie van meerdere doelstellingen in een beleidsveld te bereiken, is het risico op dergelijke interferenties groter. Verschillende maatregelen kunnen wel complementair zijn in functie van één van de doelstellingen, maar zijn daarom nog niet noodzakelijk allemaal functioneel of neutraal voor een andere vooropgestelde doelstelling.

De complexiteit is het sterkst wanneer het maatregelenpakket betrekking heeft op doelstellingen die de grenzen van verschillende beleidsvelden doorbreken of overstijgen. Er zijn twee algemene varianten:

- Integraal beleid: wanneer er sprake is van overkoepelende doelstellingen (bv. duurzame ontwikkeling) waartoe de doelstellingen binnen elk van de aparte beleidsvelden een bijdrage leveren.
- Transversaal beleid: wanneer men dezelfde doelstelling(en) nastreeft binnen verschillende beleidsvelden (bv. gelijke kansen). Deze doelstelling kan betrekking op een specifieke doelgroep (bv. jeugd, senioren, vrouwen, allochtonen, mindervaliden) of een afgebakende regio (bv. gebiedsgericht beleid voor stadsvernieuwing).

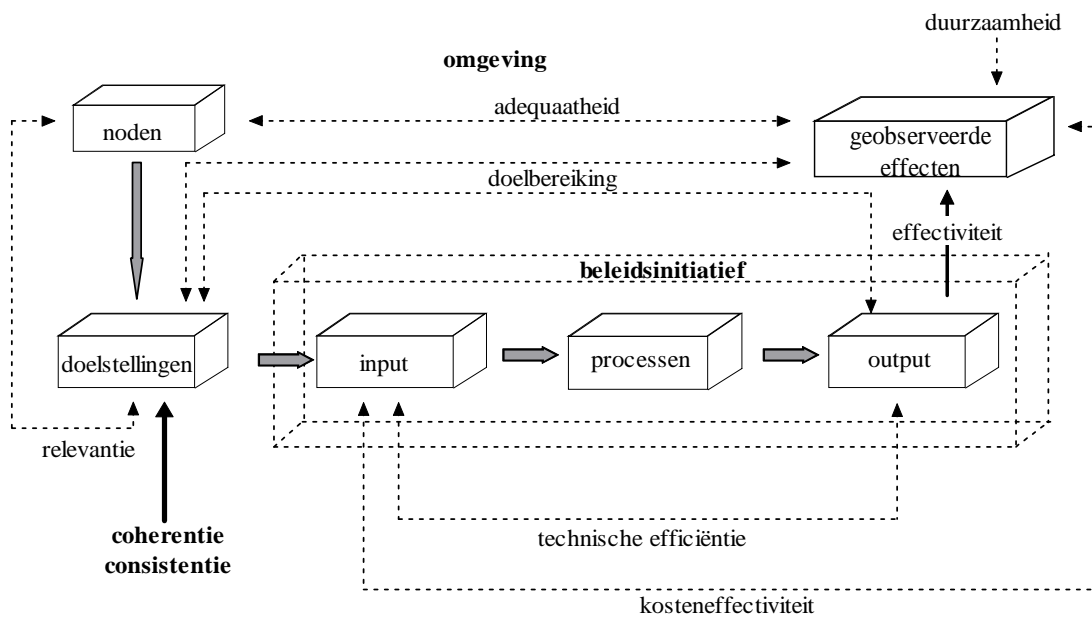
Naarmate de scope van het geplande beleid verbreedt en als gevolg daarvan de complexiteit verhoogt, neemt het belang van coherentie en consistentie toe als toetsingscriteria in een ex ante evaluatie.

Met coherentie verwijzen we naar de interne samenhang en cohesie van een beleidsinitiatief binnen een bepaald beleidsveld. De coherentie is af te lezen van de mate waarin de verschillende maatregelen bijdragen tot de doelstellingen die onder het beleidsinitiatief ressorteren en van de complementariteit tussen de doelstellingen.

Met consistentie verwijzen we naar de mate waarin positieve en negatieve spillovers naar andere beleidsvelden respectievelijk zijn gemaximaliseerd en geminimaliseerd. Het gaat met andere woorden hier om de samenhang tussen doelstellingen van verschillende beleidsvelden en tussen elk van deze doelstellingen en eventuele overkoepelende beleidsdoelstellingen. Consistentie is dus een belangrijk aandachtspunt wanneer de overheid een integrale beleidsvoering of een transversaal beleid nastreeft.

Wanneer we het beleidsveld als 'pivot' nemen, kan samenvattend gesteld worden dat coherentie een interne focus heeft, terwijl consistentie een externe, grensoverschrijdende focus heeft.

Vertaald naar het beleidstheoretische raamwerk wordt de vraag gesteld in welke mate de geplande beleidsacties in functie staan van de vooropgestelde beleidsresultaten binnen het veranderingsmodel.



Figuur 26: Coherentie en consistentie van beleid als evaluatiecriteria

5.4.4.2 Meerwaarde

De meerwaarde van een ex ante evaluatie op de coherentie en consistentie ligt in het uitklaren hoe verschillende maatregelen en doelstellingen op elkaar inhaken in meer complexe veranderingsmodellen. De toetsing op deze criteria kan aan het licht brengen dat bepaalde doelstellingen niet compatibel zijn. Of aantonen dat bepaalde maatregelen tegenstrijdig relaties met een doelstelling hebben. Dergelijke tegenstrijdige relaties kunnen immers nefast zijn voor de effectiviteit van het beleid als geheel of van bepaalde aspecten ervan. Ook ongewenste neveneffecten kunnen zo voor een deel vermeden of ingecalculereerd worden.

Net als de voorgaande types van ex ante evaluatie, kadert een evaluatie op (interne) coherentie en (externe) consistentie van het beleid in het ondersteunen van de beleidsplanning en de verbetering van gepland beleid.

5.4.4.3 Aanpak en technieken

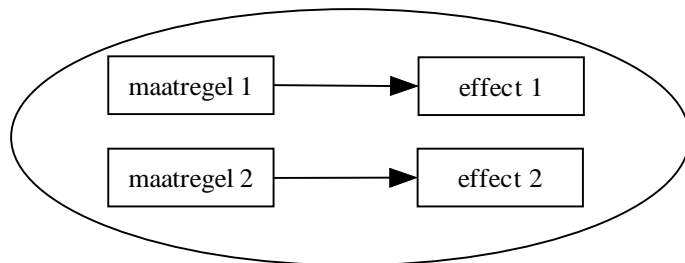
Een eerste wijze voor het toetsen van de (interne) coherentie van het beleid gebeurt door het verifiëren van de hiërarchie van de doelstellingen. In de eerste plaats dient men na te gaan of het onderscheid werd gemaakt tussen doelstellingen op het strategische niveau en deze op het operationele niveau. Verder kan men verifiëren of elke doelstelling op het juiste planniveau wordt ingedeeld. Ten derde kan men nagaan of elke strategische doelstelling werd doorvertaald naar het operationele niveau, en omgekeerd of elke geformuleerde operationele doelstelling ressorteert onder een van de strategische doelstellingen. Een

vierde stap bestaat uit een controle op tegenstrijdigheden of overlappingen tussen doelstellingen, zowel op strategisch als op operationeel niveau.

De doelstellingenstructuur kan visueel worden voorgesteld met behulp van een doelboom. In een diagram worden daarbij de meest algemene (strategische) doelstellingen bovenaan geplaatst. Vervolgens worden deze naar onder opgesplitst naar meer specifieke (operationele) doelstellingen. Vaak kunnen strategische doelstellingen uitgesplitst worden in meerdere dimensies, prioriteiten of accenten (bv. qua doelgroep). Ook operationele doelstellingen kan men vaak nog ontleden in meerdere subdoelstellingen die overeenkomen met aparte maatregelen.

Een tweede wijze om de (interne) coherentie van het beleid te toetsen is te kijken naar de causale verbanden tussen strategische en operationele (sub-)doelstellingen. Dit komt neer op een verificatie van de beleidslogica vanuit de vraag of een beleidsactie (maatregel) geformuleerd in functie van het bereiken van een bepaald beoogd beleidsresultaat (effect) geen negatieve gevolgen meebrengt ten aanzien van een ander beoogd beleidsresultaat (effect).

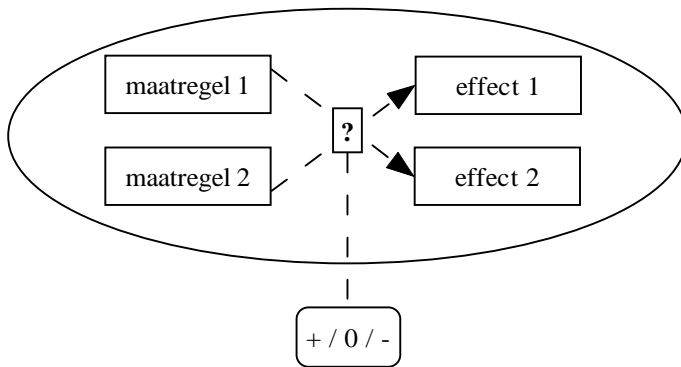
Stel dat een beleidsinitiatief 2 beoogde beleidsresultaten of effecten nastreeft en er voor elke effect één beleidsactie of maatregel wordt voorzien. Het (vereenvoudigde) veranderingsmodel, met abstractie van determinanten of tussenliggende variabelen, ziet er dan als volgt uit.



Figuur 27: Generiek veranderingsmodel bij complex beleid

Een ex ante check op coherentie bestaat er dan in om de relaties te onderzoeken tussen maatregel 1 en effect 2 enerzijds en tussen maatregel 2 en effect 1 anderzijds. Theoretisch zijn er immers telkens drie mogelijkheden:

- de maatregel of operationele doelstelling is functioneel voor het andere beoogde effect of strategische doelstelling (+);
- de maatregel is neutraal voor het andere beoogde effect (0);
- de maatregel is disfunctioneel voor het andere beoogde effect (-).



Figuur 28: Coherentie bij complex beleid

Een surplus aan functionaliteit vanuit een bijkomende maatregel voor eenzelfde doelstelling is in principe een gunstige situatie. Wel dient men na te gaan of er geen te grote (inhoudelijke) overlap bestaat tussen de twee geplande maatregels. In dat geval moet de vraag gesteld of beide maatregels noodzakelijk zijn. Wanneer andere maatregels neutraal zijn ten aanzien van een andere doelstelling, stelt er zich geen probleem.

Bij de derde mogelijke situatie is er echter sprake van een tegenstrijdigheid of incoherentie in het beleid: een maatregel is functioneel voor een van de beoogde doelstelling maar sorteert tevens een neveneffect dat contrasteert met een andere doelstelling. Er is dus sprake van een ongewenst neveneffect. Om dit te vermijden dient de maatregel bijgestuurd of vervangen te worden. Het nemen van een zogenaamde flankerende maatregel kan ook een oplossing zijn. Men scheidt dan een randvoorwaarde waardoor het neveneffect uitblijft. Anderzijds kan de disfunctionaliteit ook wijzen op een incoherentie op het niveau van de beoogde effecten, lees: de strategische doelstellingen. In dergelijk geval moet men deze doelstellingen van het beleid herbekijken.

Een beleidsinitiatief binnen één beleidsveld kan ook gevolgen hebben voor andere beleidsinitiatieven in andere beleidsvelden. Om de (externe) consistentie van een beleidsinitiatief binnen een beleidsveld met het beleid in andere beleidsvelden na te gaan, kan men analoog te werk gaan als hierboven beschreven. In het schema ressorteert effect 2 dan onder een ander beleidsveld als maatregel 1.

Een gepland beleidsinitiatief kan zich ook bewust uitstrekken tot meerdere beleidsvelden (cf. supra: integraal of transversaal beleid). Ook in deze gevallen is een ex ante evaluatie op de (externe) consistentie van het beleid aangewezen. Daarvoor kan men een logisch diagram van de verwachte effecten opstellen. In tegenstelling tot het instrument van de doelboom, worden hier de overkoepelende of globale effecten terug samengesteld vanaf de verwachte effecten binnen elk van de betrokken beleidsvelden.

Het mag immers wenselijk zijn om in een complex beleidsprogramma elke maatregel afzonderlijk af te wegen op diens bijdrage aan de uiteindelijke integrale of transversale doelstelling; vaak staan praktische hindernissen in de weg. De impact van een maatregel op bijvoorbeeld een globale doelstelling als economische en sociale cohesie van een regio omvat teveel complexe mechanismen om het voorwerp te zijn van een betrouwbare evaluatie. Veel van deze mechanismen vragen daarenboven een macro-economische benadering waarbinnen een robuuste analyse op maatregelniveau zelden mogelijk is. De analyse van verwachte effecten zal zich moeten beperken tot een tussenniveau. Voorbij dit niveau zal men weliswaar uitspraken kunnen doen op basis van extrapolatie en economische modellen, maar deze informatie laat niet toe om afzonderlijke maatregelen onderling af te wegen op hun relatieve effectiviteit.

Het logische diagram van verwachte effecten komt neer op de constructie van het veranderingsmodel van het beleidstheoretische raamwerk (cf. supra). Het veranderingsmodel wordt opgebouwd op programmaniveau waarbij beleidsacties vanuit meerdere beleidsvelden worden samengebracht.

De determinanten zijn hier te beschouwen als effecten op korte termijn en de beleidsresultaten als effecten op lange termijn of impact. Deze laatste komen overeen met de integrale of transversale beleidsdoelstellingen.

Het op te stellen veranderingsmodel om de consistentie van het beleid te toetsen heeft betrekking op sectoroverschrijdend en dus complex beleid. Daarom is het betrekken van verschillende actoren en instanties uit de respectieve beleidssectoren noodzakelijk. Vanuit hun expertise kunnen de verwachte effecten in kaart worden gebracht, en ook mogelijke dwarsverbanden tussen sectoren verkend worden. In deel 3 van de handleiding wordt een specifieke techniek voor deze context toegelicht:

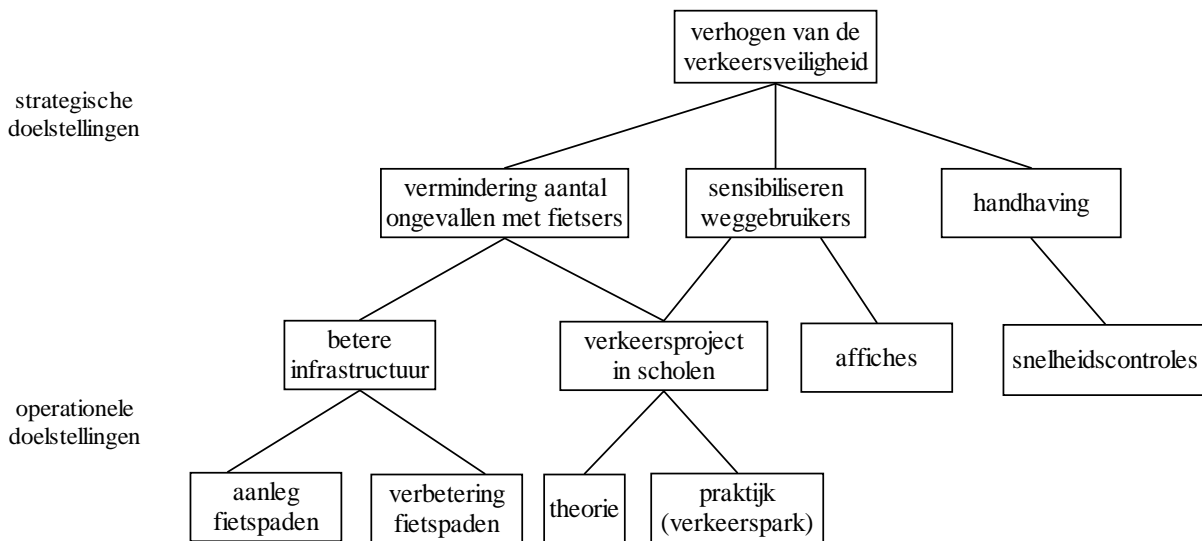
- Concept mapping van verwachte impact

Het instrument levert ook beschrijvende informatie op over verwachte impact of lange termijn effecten die verder gebruikt kan worden met het oog op monitoring (indicatorselectie) en ex post evaluatie (selectie van evaluatievragen en -criteria).

5.4.4.4 Voorbeeld

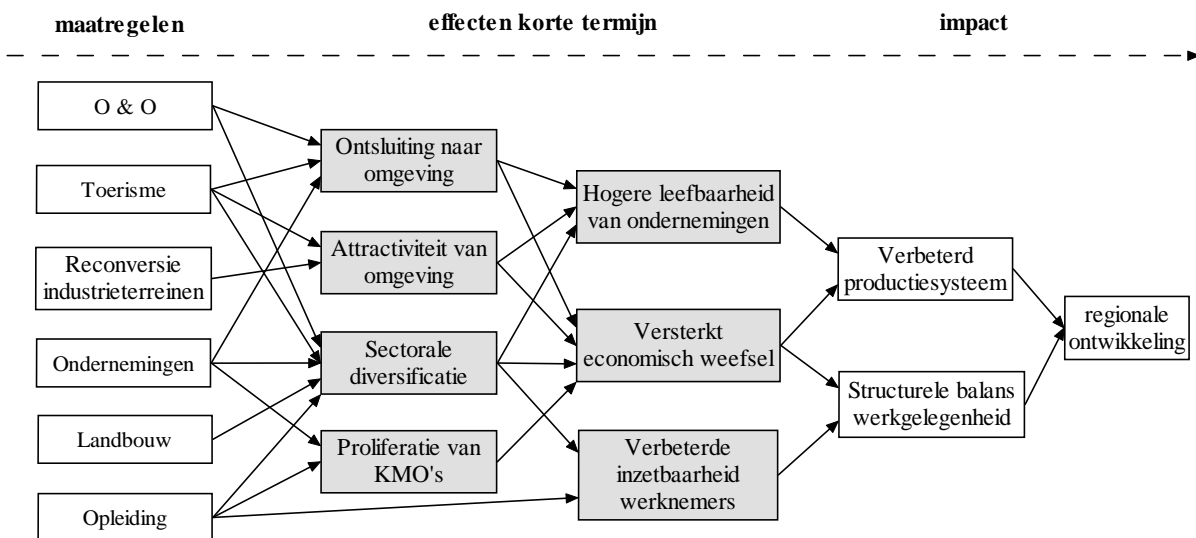
Het eerste onderstaande schema geeft een voorbeeld van een boomstructuur van doelstellingen of doelboom in het kader van een lokaal verkeersveiligheidsbeleid. Het tweede schema bevat een logisch diagram van verwachte impact in het kader van het integraal beleid gericht op regionale (sociaal-economische) ontwikkeling van de provincie Henegouwen, in het kader van Europese structuurfondsen.

Boomstructuur



Figuur 29: Boomstructuur van doelstellingen uit het verkeersveiligheidsbeleid

Logisch diagram van verwachte impact



Figuur 30: Logisch diagram van verwachte impact: het regionaal programma voor Hengouwen voor Europese steun (1994-1999)

5.4.4.5 Kritieke slaagfactoren

- Helder geformuleerde doelstellingen

Om gepland beleid accuraat te kunnen evalueren op interne coherentie en/of externe consistentie, is duidelijkheid over de wat de doelstellingen van het beleid zijn een noodzakelijke

voorwaarde. Ze vormen immers het evaluandum bij dit type van ex ante evaluatie. De mate waarin aan deze voorwaarde is voldaan hangt samen met de kwaliteit van de beleidsdocumenten. Wanneer hierin de doelstellingen onvoldoende helder omschreven staan, dient de evaluator daarover bijkomende informatie te verzamelen bij de beleidsmakers.

- Intersectorale samenwerking

In het bijzonder wanneer de consistentie van het beleid getoetst wordt, is een consensus tussen de verantwoordelijken en stakeholders van de verschillende betrokken beleidssectoren omtrent de opzet van de evaluatie belangrijk. De evaluatie tracht immers om een totaalbeeld op te bouwen over de grenzen van beleidsvelden heen. Daarvoor is een gemeenschappelijk beeld van de beleidsstructuur -het toegepaste veranderingsmodel - vereist.

5.4.5 Implementatie van het beleid

5.4.5.1 Opzet en vraagstelling

Een ex ante evaluatie kan niet enkel focussen op de inhoud van het beleid, maar kan ook vooruit kijken naar de uitvoering ervan. In dat geval staat het implementatiemodel van het beleidstheoretische raamwerk centraal. Meer bepaald zijn dan de implementatiecapaciteit, het uitvoeringsproces (cf. draaiboek) en/of de toegang tot en de omgang met de doelgroep het evaluandum.

De beleidsuitvoering is sterk gekoppeld aan een managementperspectief. Aspecten van intra- en inter-organisatorische sturing, coördinatie, verzelfstandiging, communicatie, human resources, voorraadbeheer, wachttijden, klantenbeleid, e.d. zijn daarbij vaak terugkerende aandachtspunten.

In het bestek van deze handleiding ontbreekt de plaats om uitgebreid in te gaan op de evaluatie van de uitvoeringsprocessen van beleid vanuit deze managementbenadering. Toch willen we het nut van een ex ante evaluatie op de uitvoering van het beleid benadrukken. We geven ook kort aan welke aanpak bij dit type van evaluatie behoort.

5.4.5.2 Meerwaarde

Een falend beleid is niet zelden terug te brengen tot een problematische uitvoering. Het is dan ook van belang om deze fase in de beleidscyclus voldoende aandacht te schenken, naast de planning van het beleid. Want het traditionele beeld dat beleidsplanning de uitvoering bepaalt, gaat niet meer op. Veeleer is er een wisselwerking: de uitvoering geeft in grote mate gestalte aan de inhoud van het beleid. Dit is ook af te leiden uit het conceptueel kader van de beleidstheorie. In de praktijk gaat het implementatiemodel vooraf aan het veranderingsmodel. Echter, voor het (re)construeren van de achterliggende beleidslogica

of -theorie kan men zowel voorwaarts als terugredeneren (cf. supra); zowel het implementatiemodel als het veranderingsmodel kan als vertrekpunt genomen worden.

Toch wordt vaak onvoldoende vooraf aandacht geschonken aan het implementatiemodel. Wanneer het beleid uitgetekend is en de doelstellingen bepaald zijn, hebben beleidsmakers vaak de neiging om meteen over te gaan naar de volgende fase van de beleidscyclus, namelijk de uitvoering. De implementatie krijgt in dat scenario pas meer aandacht wanneer achteraf blijkt dat het beleid niet de verwachte resultaten heeft opgeleverd.

Echter, door prospectief bepaalde aspecten van de beleidsuitvoering te evalueren op praktische uitvoerbaarheid of haalbaarheid, kan men bepaalde knelpunten tijdig ontdekken en daarmee problemen vermijden. Dergelijke ex ante evaluatie past dan ook onder het motief van het ondersteunen en verbeteren van de (operationele) beleidsplanning. Ze kan aangeven welke indicatoren belangrijk zijn om via monitoring op te volgen zodra het beleid in uitvoering gaat.

Ook dit type van evaluatie is zowel van politiek als administratief oogpunt relevant. Ze ondersteunt enerzijds een doelgerichte en efficiënte allocatie van middelen en vermijdt anderzijds de situatie dat politici oplossingen aankondigen die later onuitvoerbaar blijken te zijn.

5.4.5.3 Aanpak en technieken

De aanpak bij een ex ante evaluatie van de implementatie van het beleid is op zich vrij eenvoudig: de organisatie van een testfase of try-out van de uitvoering op compacte schaal. Een piloottest of proeftuinexperiment zijn aangewezen technieken om de praktische haalbaarheid van een beleidsinitiatief te bepalen en bijgevolg de grootschalige invoering van een mangelend beleid te vermijden.

Een testfase vervult een brugfunctie tussen beleidsbepaling en beleidsuitvoering. Daardoor echter zal deze test meestal onder een hoge tijdsdruk moeten gebeuren, in die mate zelfs dat ze beperkt kan worden tot bepaalde onderdelen van het geplande beleidsinitiatief.

De aanpak via een testfase kan overigens ook relevant zijn om de beleidseffectiviteit na te gaan. Daarmee is niet enkel het implementatiemodel maar ook het beoogde veranderingsmodel het evaluandum. Men kan bijvoorbeeld nagaan hoe het beleid de doelgroep bereikt en of de selectie van determinanten accuraat is. In dat opzicht vertoont deze aanpak overeenkomsten met het verzamelen van beleidsvoorbereidende informatie bij en ex ante evaluatie van de beleidsrelevantie (cf. supra). Voor dat evaluatietype is het scherp stellen van de maatschappelijke problematiek essentieel, maar kan ook het verzamelen van kenmerken van de beoogde doelgroep relevante informatie opleveren om keuzes te maken tussen de determinanten waarop men met het beleid wil inwerken. Het voornaamste verschilpunt met de aandacht voor de doelgroep in een uitvoeringstest zit in de timing ten aanzien van het ontwerpen van de beleidstheorie. Een ex ante evaluatie van de beleidsrelevantie gebruikt deze informatie om de

beleidstheorie vorm te geven, terwijl in een ex ante evaluatie van de beleidsimplementatie het implementatiemodel als bestaand onderdeel van de beleidstheorie getest kan worden.

Uit dit laatste zou men kunnen besluiten dat een testfase eigenlijk een ex post evaluatie is. Vanwege de beperking in tijd en omvang kan een piloottest echter eerder beschouwd worden als een ex ante evaluatie-instrument dan wel een beoordeling ex post.

5.4.5.4 Voorbeeld

De case van de spaarlampenbedeling demonstreert in negatieve zin de meerwaarde van een ex ante evaluatie van de uitvoering van het beleid; een terugblik op een gemiste vooruitblik.

In september 2003 wordt het beleidsinitiatief aangekondigd dat de Vlaamse overheid spaarlampen zal uitdelen aan elk Vlaams gezin. Dat zal gebeuren in 2 rondes, de eerste ronde in juni 2004, de tweede in de loop van 2006. De maatregel, begroot op 4,8 miljoen euro, kadert in het beleid gericht op duurzame ontwikkeling en meer bepaald het sensibiliseren van mensen voor de problematiek van overmatig energieverbruik. Door een discussie over de distributie van de spaarlampen start de eerste ronde van de actie met enige vertraging in augustus 2004.

De uitvoering verloopt anders naargelang mensen aangesloten zijn bij een gemengde of zuivere intercommunale. Gezinnen die bij een gemengde intercommunale zijn aangesloten, ontvangen eerste een brief waarin de maatregel wordt aangekondigd en toegelicht. Vervolgens vindt via de post de bedeling van spaarlampen plaats. Wanneer men niet thuis is, deponeert de postbode een bon in de brievenbus waarmee men de spaarlamp kan afhalen op het postkantoor. Voor gezinnen aangesloten bij een zuivere intercommunale is de procedure dat men steeds de lamp zelf moet gaan afhalen in het postkantoor.

In 2005 beslist de minister een tussentijdse evaluatie uit te voeren. De resultaten van een enquête bij de doelgroep van het beleid zijn niet bemoedigend. Van de respondenten zegt 31% nog nooit gehoord te hebben van de beleidsmaatregel, 35% heeft geen brief of bon gekregen en 60% verklaarde geen spaarlamp te hebben ontvangen. Van de 40% die wel al een lamp kreeg had 25% die de lamp ongebruikt heeft weggelegd, 12% personen die aangeven dat ze de lamp om technische redenen niet kunnen gebruiken, en 10% die nog geen idee heeft over het eventueel gebruik van de lamp.

Op basis van de resultaten van de tussentijdse evaluatie komt de tweede ronde van het beleidsinitiatief ter discussie te staan. De implementatie van de maatregel loopt duidelijk mank, bovendien komt aan het licht dat bijna de helft van de gezinnen de verkregen spaarlamp (nog) niet gebruikt. Uiteindelijk beslist men om toch door te gaan met de uitvoering, maar men wijzigt de wijze van uitvoering. In januari 2006 werd beslist te starten met de 2^{de} ronde. Gezinsleden kregen nu een kortingsbon van 5 euro, er was geen bedeling meer van spaarlampen via de post. Er werden 2,5 miljoen kortingsbonnen voorzien, die in de relevante handelszaken konden worden ingeruild. Bovendien lanceerde men een mediacampagne om de

mensen te sensibiliseren voor de problematiek en de beleidsactie nogmaals onder de aandacht te brengen. De coördinatie komt in handen van het nieuwe Vlaamse energiebedrijf Eandis, namens de gemengde intercommunales.

Dit voorbeeld toont aan dat een ex ante evaluatie van de uitvoering problemen had kunnen voorkomen. Indien men de procedure van uitvoering en de rol daarin van de doelgroep kritisch had bekeken, was men mogelijk vooraf al tot de conclusie gekomen dat de bedeling via het postkantoor niet de optimale oplossing was. In een tijd van moderne communicatiemiddelen passeren we minder langs het postkantoor. De openingstijden stroken ook niet steeds met de werkuren. Bovendien is het aannemelijk dat heel wat werkende mensen niet speciaal voor spaarlamp langs het postkantoor passeren.

5.4.5.5 Kritieke slaagfactoren

- Betrekken van uitvoerders

Het is belangrijk dat de personen die met de eigenlijke uitvoering belast zijn betrokken worden bij de evaluatie. Zij zijn immers meestal het best geplaatst om vanuit hun concrete ervaringen hindernissen of knelpunten in het uitvoeringsproces onder de aandacht brengen.

- Beschikbare tijd t.a.v. besluitvormingsproces

De meerwaarde van een ex ante evaluatie van de uitvoering van het beleid werd hierboven uiteengezet. Net als bij elke ex ante evaluatie speelt echter het tijdsaspect een belangrijke rol om de meerwaarde te benutten. Het testen van de uitvoering op beperkte schaal neemt een bepaalde tijd in beslag die men vooraf moet incalculeren in het besluitvormingsproces; de resultaten van de evaluatie moeten immers nog kunnen omgezet worden in een verbetering van het geplande beleidsinitiatief. Dat kunnen specifieke procedures zijn, maar mogelijk ook meer fundamentele aanpassingen wanneer blijkt dat de praktische uitvoerbaarheid problematisch is.

5.4.6 Verwachte effecten en impact

5.4.6.1 Opzet en vraagstelling

Een ex ante evaluatie kan ook aandacht hebben voor de beleidsimpact. Prospectief wordt gekeken naar de verwachte effecten. In tegenstelling tot een ex post evaluatie staat niet de vraag naar doelbereiking of beleidseffectiviteit (attributie van effecten) centraal. Dergelijke vragen kunnen immers pas gesteld worden nadat het beleid uitgevoerd is.

De opzet van een ex ante evaluatie van de beleidsimpact is het voorspellen, schatten of berekenen van effecten van gepland beleid en hun kenmerken: de aard, omvang, tijdstip, tijdsduur en intensiteit. Een

belangrijk verschilpunt met een ex post evaluatie is dan ook de onzekerheid over de effecten: men kan ze nog niet meten, enkel ramen.

5.4.6.2 Meerwaarde

De planning van beleid gebeurt met het oog op het realiseren van een verandering ten aanzien van een maatschappelijke behoefte of probleem, of het voorkomen daarvan. Voor het antwoord op de vraag in welke mate een beleidsinitiatief effect zal hebben op korte termijn en wat de impact ervan zal zijn op langere termijn, speelt de graad van zekerheid over het optreden van de effecten en nauwkeurigheid over de omvang en intensiteit van effecten een belangrijke rol.

De beslissing om een beleidsinitiatief uit te voeren kan echter afhangen van gegevens die deze onzekerheid en onnauwkeurigheid althans voor een deel wegnemen. Een ex ante evaluatie van verwachte effecten en impact kan dergelijke informatie verschaffen.

Ze kan kaderen binnen verschillende motieven. Uiteraard kan dergelijke evaluatie de planning van beleid ondersteunen, door na te gaan wat het beleid aan verandering teweeg kan brengen. Men kan ook beter inzicht krijgen op welke intensiteit van het beleid vereist is om een bepaald effect of impact op langere termijn te verkrijgen. De intensiteit kan men sturen door de inzet van middelen die men aan het beleidsinitiatief verbindt.

Anderzijds kan dergelijke evaluatie - naar analogie met een toetsing op de relevantie van het beleid - een voorafname zijn op een ex post verantwoording voor doelbereiking en beleidseffectiviteit.

Deze vorm van ex ante evaluatie kan ook bijdragen tot het identificeren van gewenste of ongewenste neveneffecten.

5.4.6.3 Aanpak en technieken

De aanpak bij dit type van ex ante evaluatie is geënt op de werkwijze bij *forecasting* of voorspelling, waarbij men prognoses maakt en scenario's opstelt. Forecasting is het gericht verschaffen, systematiseren en gebruiken van toekomstgerichte informatie die voor beleidsbeslissingen bruikbaar is.

Daarbij maakt men een traditioneel onderscheid tussen explorerende en normatieve forecasting. Verkennende forecasting reikt toekomstverkennende informatie aan ten behoeve van *logisch mogelijke*, alternatieve toekomstige situaties en tracht aan te geven hoe men vanuit de bestaande toestand kan evolueren naar deze mogelijke situaties. Normatieve forecasting verschaft alternatieve toekomstbeelden, echter vanuit een subjectief en normatief beoordeelde bestaande toestand en *wenselijk geachte* toekomstige situaties. Men tracht aan te geven hoe men de discrepanties kan wegwerken en evolueren van de bestaande naar de wenselijke situatie. Binnen bepaalde sectoren en afhankelijk van het soort

gegevens kan men werken met simulaties van mogelijke beleidsacties om na te gaan wat de effecten zijn.

Een ex ante beleidsevaluatie van de verwachte effecten en impact is verwant met maar niet gelijk aan forecasting. Bij de evaluatie start men vanuit een concreet beleidsinitiatief dat gepland is of dat in de ontwerpfase zit. Forecasting en toekomstverkenning hebben enkel betrekking op de maatschappelijke omgeving. De verwachte effecten en impact van een beleid situeren zich echter ook in deze maatschappelijke omgeving; de effecten en impact van een beleidsinitiatief kunnen door deze context beïnvloed worden en omgekeerd.

Gegeven bovenstaande parallel in opzet en aanpak, zijn verschillende technieken die ingezet kunnen worden bij forecasting ook relevant voor een ex ante beleidsevaluatie van verwachte effecten en impact. De hoofdfunctie van deze technieken in het kader van evaluatie is het maken van analyses. Ze kunnen informatie aanreiken over toekomstige effecten die als basis kan dienen om beoordelingen te maken over beleidsopties of scenario's.

- Tijdreeksen
- Inputoutput modellen
- Delphi-techniek
- Impact matrix
- Cross impact matrix

5.4.6.4 Voorbeeld: multimodaal verkeersmodel

Voor de planning van het mobiliteitsbeleid - en daarmee samenhangend het ruimtelijke ordeningsbeleid - op Vlaams, provinciaal en lokaal niveau, werd een multimodaal verkeersmodel ontwikkeld door de Administratie Wegen en Verkeer van de Vlaamse Overheid. Aan de hand van deze simulatiemodellen over het personen- en goederenverkeer wordt getracht de evolutie te voorspellen bij voortzetting van het huidige beleid en bij een gewijzigd mobiliteitsbeleid ten gevolge van mogelijke beleidsopties of geplande maatregelen.

Aan de basis van de mobiliteitsontwikkeling liggen tal van mechanismen. Voor het personenvervoer hebben zij te maken met de maatschappelijke, economische en ruimtelijke organisatie waardoor allerlei activiteiten op afzonderlijke plaatsen worden uitgeoefend. De verplaatsingen die men doet zijn sterk individueel gebonden en worden beïnvloed door o.a. de leeftijd, de rolverdeling binnen het gezin en de levensstijl die men zich aanmeet. Voor het goederenvervoer geldt een andere dynamiek die nauw gerelateerd is aan de economische ontwikkeling en de technologie. Productieprocessen zijn meer en meer verdeeld in verschillende stadia en vinden op verschillende locaties, in gespecialiseerde bedrijven plaats, met meer transport tot gevolg. Het verkorten van doorlooptijden, inspelen op onverwachte marktveranderingen en de grotere vraag naar maatwerk hebben een grote invloed op de aard en de hoeveelheid transport. Ook de aard van goederen (meer half-/eindfabricaten i.p.v. grondstoffen) zorgen

voor andere transportstromen. Ook de Europese gedereguleerde markt zonder handelsbelemmeringen zorgt voor een verdergaande Internationalisering van de productie- en transportstromen.

Het model wordt gevoed met een aantal basisparameters waaronder de ruimtelijke verdeling van wonen en werken. Het grondgebied dat overeenkomt met het planniveau wordt opgedeeld in een aantal zones. Deze zijn onderling verbonden via een netwerk van wegen, spoor-, tram- en buslijnen. Voor elke zone worden gegevens opgetekend over de bevolking, de tewerkstelling, de scholen en de winkels. Enquêtes over het verplaatsingsgedrag, gegevens van de recentste volkstelling en verschillende databanken liggen aan de basis van deze berekeningen.

Het model berekent het aantal verplaatsingen tussen de verschillende zones en de route waarlangs gereden wordt. Omdat het om een multimodaal verkeersmodel gaat, wordt ook bepaald met welk vervoermiddel de verplaatsing gemaakt wordt (auto, trein, tram, bus, fiets); de 'modal split'. De betrouwbaarheid van het model wordt nagegaan door de resultaten ervan voor een bestaand basisjaar te vergelijken met werkelijke tellingen. Om dit goed te kunnen doen, is het noodzakelijk de parameters van het model regelmatig aan te passen (openbaar vervoer, wegen, bevolking en tewerkstelling per zone, regelgeving, inkomen ...).

Er zijn drie niveaus waarop een model kan uitgebouwd worden, elk met een eigen finaliteit:

Macro	Nagaan van effecten door simulatie van strategische beleidsbeslissingen op lange termijn op een groot gebied (gewest of provincie). Op deze schaal wordt niet naar individuele verkeersdeelnemers gekeken, maar naar verschuivingen in patronen.
Meso	Op dit tussenniveau kunnen effecten van oplossingen op een ruime omgeving geëvalueerd worden. Naast effecten op doorstroming van verkeer kan ook rekening worden gehouden met gevolgen van andere ontwikkelingen en maatregelen die de mobiliteit beïnvloeden.
Micro	Inzoomen op de verkeersafwikkeling van kruispunten of een cluster van kruispunten. Het gedrag van de individuele verkeersdeelnemer wordt gesimuleerd. Eventuele effecten van de aanpak die verder reiken dan het kleinschalige gebied zijn echter niet toonbaar.

Tabel 8: Detailniveaus van het multimodaal verkeersmodel

Op basis van het macro-model werd in het trendscenario (bij ongewijzigd beleid) een groei voorzien van het aantal verplaatsingen tussen 2000 en 2010 met 9% in Vlaanderen. Binnen Vlaanderen en het Brussels Hoofdstedelijk Gewest zal het aantal personenkilometer met de wagen met 21% stijgen tegenover 6% voor het openbaar vervoer. De gemiddelde reistijd met de auto neemt toe tot 33 minuten (een stijging met 17%) en de snelheid neemt af tot 11km/h (een daling met 12%). Het goederenvervoer neemt toe met 28%. Het wegverkeer blijft de sterkste groeier met 31% en versterkt zijn aandeel in de modal split tot 74%. De binnenvaart groeit met 22% en behoudt zijn aandeel in de modal split (13%). Het vervoer per spoor stijgt met 15%, en daalt qua aandeel in de modal split naar 13%.

Bronnen: <http://www.uitweg.be/nummer-03/pagina2.htm> ,
<http://viwc.lin.vlaanderen.be/mobiliteit/printsamenvatting.htm>

5.4.6.5 Kritieke slaagfactoren

- Beschikbaarheid en kwaliteit van inputdata

Een belangrijke voorwaarde voor het gebruik van technieken met het oog op scenario-opbouw en het gebruik in een ex ante evaluatie van verwachte effecten en impact, is de beschikbaarheid en

kwaliteit van de data die als input in het model of als basis voor het scenario dienen. Zonder deze data verkrijgt men immers geen resultaat.

- Correcte omgang met een model als instrument

Modellen zijn geen glazen bollen; evenmin reiken ze oplossingen aan. De resultaten zijn afhankelijk van de input (cf. supra) en ze moeten nog steeds geïnterpreteerd worden. Een model is ook niet kritisch: onbetrouwbare inputdata leveren onbetrouwbare resultaten op met betrekking tot toekomstige situaties en evoluties.

- Effecten van gemodelleerde maatregelen of van de modelleringswijze?

Bij de ontwikkeling van elk model moet men steeds waakzaam zijn of de resultaten het gevolg zijn van de maatregelen die men simuleert (modelleert), dan wel van de wijze waarop men het model heeft opgebouwd. Het analysereniveau is daarbij belangrijk: een model op macroniveau zal, omwille van de vereenvoudiging en veralgemening, vertekende resultaten geven wanneer men inzoomt op microniveau.

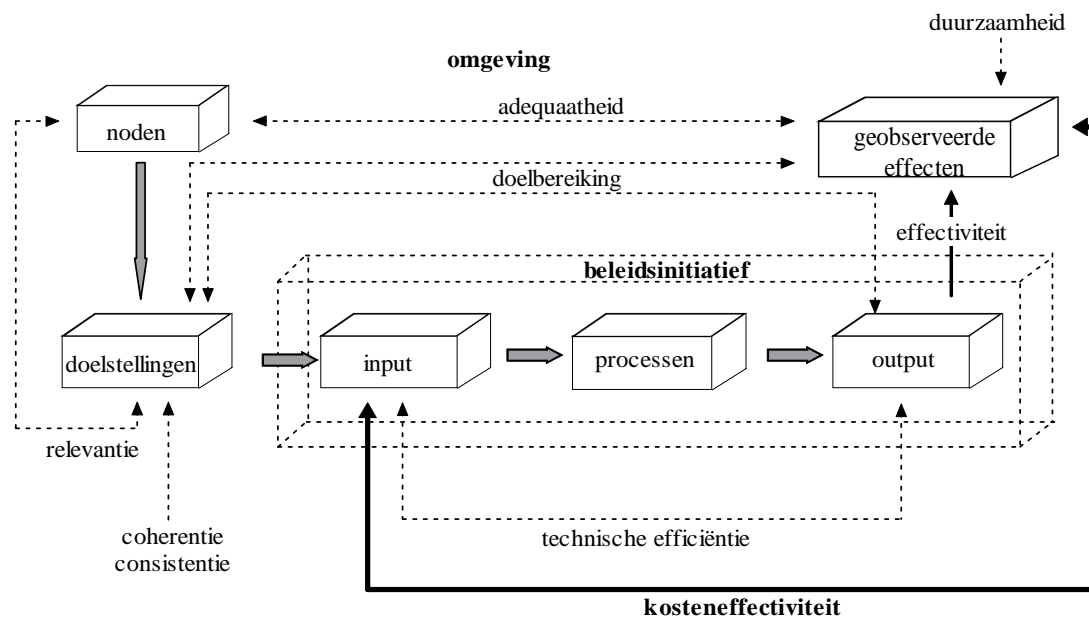
5.4.7 Kosten en baten en kosteneffectiviteit

5.4.7.1 Opzet en vraagstelling

Een ex ante evaluatie met een focus op de verwachte effecten kan in scope uitgebreid worden door ook de kosten in rekening te brengen. Centraal in de evaluatie staat de afweging van de kosten, de voordelen en de nadelen van het beleidsinitiatief.

De voor en nadelen verwijzen naar de effecten, terwijl de kosten verwijzen naar de financiële input die nodig is voor de uitvoering en realisatie van de beleidsoutput.

Centrale vraag is of het beleidsinitiatief de investering waard is. Afhankelijk van de context kan men trachten om alles in monetaire termen te vertalen. Wanneer dit moeilijk is, kan men ook nagaan hoe de kostprijs van een beleid zich verhoudt tot het saldo van positieve en negatieve effecten. Wanneer het beleidsinitiatief een beperkt aantal effecten genereert, kan ook een coëfficiënt van kosteneffectiviteit ontwikkelen als beoordelingsmaatstaf. Dergelijke coëfficiënt beantwoordt dan de vraag hoeveel het bereiken van een beleidseffect van bepaalde omvang kost.



Figuur 31: Kosteneffectiviteit als evaluatiecriterium

5.4.7.2 Meerwaarde

Een ex ante evaluatie die niet alleen de toekomstige effecten in kaart brengt, maar ook de afweging met de middelen maakt, verschaft een totaalbeeld op het beleidsinitiatief. Dit ondersteunt de keuze over het al dan niet uitvoeren van een beleid. Het totaalbeeld maakt de keuze transparanter. Dit is des te meer van belang wanneer het beleidsinitiatief een aanzienlijke investering van overheidsgelden vergt. Het uitvoeren van een kosten-batenanalyse wordt ook interessant wanneer de baten en/of kosten zich spreiden over een langere tijdshorizon. Vaak is immers het zo dat het zwaartepunt van de kosten in het heden ligt, terwijl een groot deel van de baten zich in de toekomst manifesteert.

5.4.7.3 Aanpak en technieken

Voor de afweging van kosten en effecten hanteert men monetaire of semi-monetaire technieken. Men tracht alle of zoveel mogelijk kosten en baten in geld trachten uit te drukken. Daarmee kan men gemakkelijk tot een synthesesoordeel komen: het saldobedrag van kosten en baten.

De technieken verschillen onderling in het bereik van effecten die worden meegenomen. Effecten kunnen verschillen op het vlak van hun wenselijkheid (gewenst / positief – ongewenst / negatief) en belangrijkheid (hoofdeffect – neveneffect). De richting hangt samen met de wenselijkheid (gewenst – ongewenst). Neveneffecten zijn dus niet per definitie negatief.

Beoordeling van het effect	Belangrijkheid van het effect	
	<i>hoofdeffect</i>	<i>neveneffect</i>
<i>positief / gewenst</i>		
<i>negatief / ongewenst</i>		

Tabel 9: Soorten effecten

Gewenste hoofdeffecten liggen vervat in de doelstellingen van een beleidsinitiatief. Eén vertrekpunt kan zijn om alle identificeerbare effecten mee te nemen. In de context van een afweging van voor- en nadelen kan men ernaar streven om het ganse spectrum van mogelijke effecten in kaart te brengen. In de mate waarin ze voorzienbaar zijn dienen naast de hoofdeffecten ook de neveneffecten, en naast de gewenste of positieve effecten ook de ongewenste of negatieve effecten in de vergelijking meegenomen te worden.

Wanneer men een sterkere koppeling met de doelstellingen van het beleid wil aanhouden, zal men een relatief beperkt aantal effecten in rekening brengen. Beide werkwijzen hebben voor- en nadelen. Respectievelijk spreken we over volgende technieken:

- Kosten-batenanalyse
- Kosteneffectiviteitanalyse

Deze technieken worden uitgebreid besproken in deel 3. Deel 4 beschrijft een uitgebreide case waarin ook een kosten-batenanalyse werd ondernomen.

5.4.7.4 Kritieke slaagfactoren

- Monetaire vertaalbaarheid

De omzetting naar monetaire termen van alle weerhouden baten of effecten is niet vanzelfsprekend. Vaak zal dit enkel kunnen via assumpties en bij consensus over een bepaalde manier van geldelijke waardering. Wanneer de hoofdeffecten van het beleidsinitiatief moeilijk of helemaal niet in geldtermen vertaalbaar zijn, stelt zich een probleem om een kosten-batenanalyse uit te voeren. Dit geldt minder voor een analyse van de kosteneffectiviteit daar deze ook niet-monetaire factoren kan meenemen.

- Ongrijpbare effecten

Een fundamenteeler probleem stelt zich wanneer belangrijke effecten van het beleid niet grijpbaar of tastbaar zijn. Dergelijke effecten kunnen moeilijk of niet uitgedrukt worden in een meeteenheid, laat staan vertaald worden naar monetaire termen. De levenskwaliteit bijvoorbeeld is moeilijk te meten. Vaak zal men dan benaderende indicatoren of proxy's hanteren. Toch zal men vaak knelpunten ervaren met de geldelijke omzetting van deze proxy's.

- Consensus over verdiscontering

De waarde van een geldelijke som in de toekomst is niet dezelfde als de actuele waarde ervan. Men zal toekomstige kosten en baten moeten omzetten naar een gelijke basis via verdiscontering. Belangrijk voor de toepassing van monetaire technieken in een evaluatie is dat er consensus wordt bereikt tussen de stakeholders over deze waarderingsregels.

5.4.8 Afwegen van beleidsalternatieven

5.4.8.1 Opzet en vraagstelling

Een andere belangrijke finaliteit van een ex ante evaluatie kan gelegen zijn in het afwegen van beleidsalternatieven. Dergelijke ex ante evaluatie is niet gericht op het verder ontwikkelen en verbeteren van een gepland beleidsinitiatief. De opzet is het ondersteunen van de beleidskeuze tussen mogelijke beleidsinitiatieven.

De centrale vraagstelling luidt dan: welk van de voorliggende beleidsinitiatieven is het beste? Meteen dringt zich de vraag op wat 'beste' hier dan betekent. Voor het beantwoorden van de vraag zal men inderdaad ook moeten kunnen terugvallen op evaluatiecriteria als vergelijkingscriteria.

Het afwegen van beleidsalternatieven kan op strategisch niveau of op operationeel niveau gebeuren. Op strategisch niveau zal men verschillende opties voor de fundamentele aanpak van het beleid vergelijken. Zo heeft men bijvoorbeeld de keuze tussen een preventief of repressief beleid om kleine criminaliteit te bestrijden. In de praktijk zal men vaak een combinatie preventieve en repressieve maatregelen uitwerken. Toch kan men, zoals zo vaak in de politiek, omwille van budgettaire beperkingen gedwongen worden om prioriteiten te stellen.

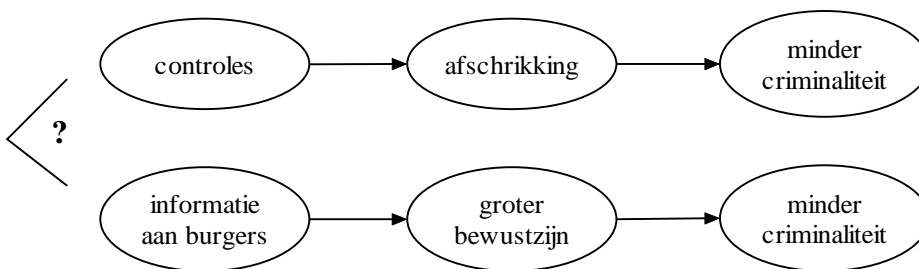
Op operationeel niveau zal men –binnen een strategische keuze met betrekking tot de aanpak- varianten vergelijken. De verschilpunten tussen de alternatieven zijn dan minder fundamenteel van aard, en zullen eerder betrekking hebben op de uitvoering van het beleid of op enkele specifieke kenmerken van de opties. Stel dat men in het bovenstaande voorbeeld kiest voor een preventieve aanpak, in het bijzonder via een campagne die de burgers attent moet maken op kleine criminaliteit en hoe ze zelf het risico op handtasdiefstal, auto-inbraak e.d. kunnen verkleinen. Voor de concrete uitvoering van het beleid kan men kiezen tussen het gebruik van affiches in het straatbeeld, de verspreiding van brochures, het aanspreken door de politie van mensen op straat, enz.

Andere voorbeelden van strategische keuzes tussen beleidsalternatieven, zijn bijvoorbeeld de keuze tussen accijnsverhoging op brandstof of het goedkoper maken van abonnementen op openbaar vervoer, de keuze tussen boetes op leegstand of renovatiepremies.

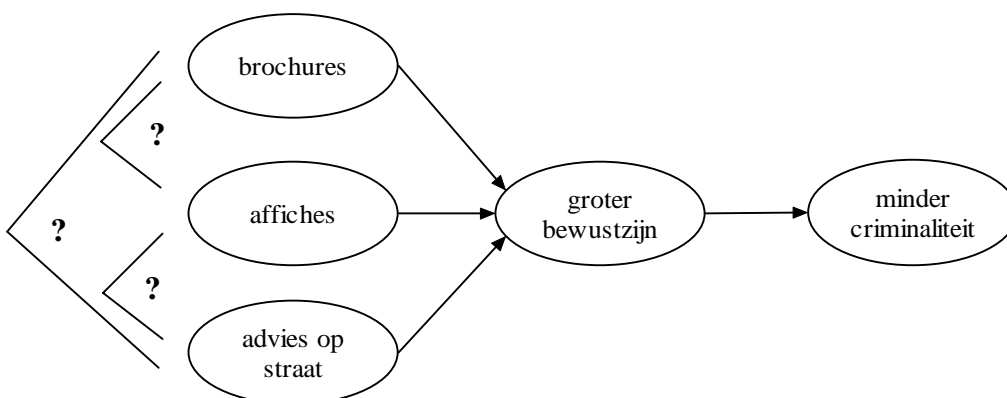
Voorbeelden van operationele beleidskeuzes zijn de keuzes tussen verschillende tracés van een spoorlijn of autoweg, tussen verschillende verkeersremmende constructies, de uitbesteding of uitvoering in eigen beheer van een groendienst of containerpark, enz.

Vertaald naar het veranderingsmodel kan men stellen dat strategische keuzes tussen beleidsalternatieven betrekking hebben op verschillende determinanten die beïnvloed kunnen worden met het beleid om dezelfde beleidsresultaten of effecten te bereiken. De operationele keuzes hebben eerder betrekking op de varianten van eenzelfde beleidsactie of output.

Onderstaande schema's bevatten een voorbeeld toegepast op de criminaliteitsbestrijding. In functie van hetzelfde effect (minder criminaliteit), kan men ervoor kiezen om potentiële daders af te schrikken, of het bewustzijn van risico's bij potentiële slachtoffers te vergroten. Afhankelijk van de strategische keuze richt het beleid zich op een andere determinant of hefboom en is een andere beleidsacties of output (controles vs. informatie aan burgers aangewezen. Stel dat men kiest om de burgers meer attent te maken. De tweede figuur geeft binnen deze strategische optie de operationele keuze aan die men moet maken. Hoe gaat men de burgers informeren? Via brochures huis aan huis, via affiches in de winkelstraten, of door het rechtstreeks aanspreken van burgers (via politieagenten).



Figuur 32: Strategische keuze: op welke determinanten richt het beleid zich?



Figuur 33: Operationele keuze: welke variant van de beleidsoutput?

Bovenstaande keuzeopties sluiten mekaar niet per definitie uit. Maar in de praktijk zal men vaak omwille van diverse redenen keuzes moeten maken, zowel op strategische als op operationeel niveau.

5.4.8.2 Meerwaarde

Wanneer men met het beleid een antwoord wil bieden op een maatschappelijk probleem of nood, zijn er vaak verschillende mogelijkheden om eenzelfde doel te bereiken. Het maken van beleidskeuzes is een cruciale stap in de beleidscyclus. De voorbereiding en bepaling van beleid dient ook zoveel mogelijk ondersteund te worden met informatie die enerzijds keuzeopties kan aanbieden en uitklaren, en anderzijds het maken van deze keuzes zelf kan vergemakkelijken. Het op een systematische en evaluatieve manier vergelijken van keuzeopties op voor- en nadelen (sensu lato) is dan een noodzakelijke en aangewezen bestuurspraktijk.

5.4.8.3 Aanpak en technieken

Het afwegen van beleidsalternatieven zal verschillen naargelang het om strategische dan wel om operationele opties gaat. Op strategisch niveau zal men te gronde met verschillende veranderingsmodellen en beleidstheorieën te maken hebben waaruit men kan of moet kiezen. In dat geval dient de evaluatie zich toe te spitsen op de achterliggende logica van de alternatieven. Vooral de keuze van de beleidsacties en de determinanten zal verschillen. De beleidsresultaten of effecten zullen wel gemeenschappelijk zijn over de alternatieven heen. Men kan nagaan en vergelijken welke determinanten de sterkste impact zullen hebben en het meest bijdragen aan de beoogde effecten. In functie daarvan kan men een keuze maken op het niveau van de determinanten en in tweede instantie tussen de beleidsacties waarmee de gekozen determinanten beïnvloed kunnen worden. De vergelijking van veranderingsmodellen gebeurt dus aan de hand van het achterwaarts redeneren (cf. supra). De vergelijking en afweging start vanaf gemeenschappelijke effecten en splitst dan uit naar alternatieve determinanten en bijhorende beleidsacties.

Op operationeel niveau vindt afweging plaats tussen gelijkende beleidsacties waaruit men zal kiezen om dezelfde determinanten te beïnvloeden. Op dit niveau zijn dus zowel de effecten als de determinanten zijn gemeenschappelijk tussen de alternatieven. Enkel de beleidsacties verschillen.

Om dergelijke varianten of operationele beleidsalternatieven met elkaar te vergelijken zijn tal van criteria denkbaar, waaronder deze die in de voorgaande paragrafen al aan bod kwamen. Een vergelijking op kosten en baten of op kosteneffectiviteit is in het bijzonder nuttig wanneer de af te wegen beleidsalternatieven gelijkende varianten zijn van eenzelfde beleidsinitiatief of output.

Heel wat vergelijkingspunten kunnen dan ook uit de strategische beleidsdoelstellingen (gericht op effecten) afgeleid worden. Zo kunnen alternatieven onderling vergeleken worden in de mate dat ze bijdragen aan elk van een bredere set van effecten die het beleid nastreeft. Dit is in het bijzonder

interessant wanneer het niet voor de hand ligt dat deze effecten compatibel zijn. Zo kan men alternatieve tracés voor een goederenspoorlijn bijvoorbeeld gaan vergelijken op economische prestatie (vervoerscapaciteit, verbindingstijd, e.d.) en milieuvriendelijkheid. Ook proceskenmerken kunnen als evaluatiecriteria geselecteerd worden, bv. de realisatietermijn.

De afweging van beleidsalternatieven kan een absoluut of relatief karakter hebben. In het eerste geval gebeurt het evalueren aan de hand van externe normen. Deze stellen bepaalde waarden op de toetsingscriteria voorop. Vervolgens gaat men per alternatief en per criterium de waarde van het alternatief vergelijken met de vooropgestelde waarde of norm. In het tweede geval werkt men zonder externe normen. Men gaat de alternatieven dan onderling vergelijken aan de hand van de respectieve waarden op de criteria.

Het inzetten van monetaire technieken om beleidsalternatieven te vergelijken biedt het voordeel dat men kan men een duidelijke maatstaf ontwerpt om tot een syntheseoordeel per alternatief en een vergelijkingspunt voor de onderlinge afweging van alternatieven te komen: het saldobedrag van kosten en baten, de kosteneffectiviteitsratio. Er zijn echter ook randvoorwaarden en beperkingen verbonden aan deze technieken (cf. supra en deel 3).

Men kan ook niet-monetaire technieken aanwenden. Men kan de alternatieven dan paarsgewijs gaan vergelijken. Om tot een eindoordeel per alternatief te komen en de alternatieven onderling af te wegen kan men ook een rangorde aanbrenge. Deze technieken kunnen een grote variëteit aan evaluatie- en vergelijkingscriteria hanteren. Ze vallen onder de gemeenschappelijke noemer van multicriteria-analyse.

De eerste stap omvat het waarderen van elk alternatief per toetsingscriterium. De tweede stap bestaat dan in het aanbrenge van een rangorde in de alternatieven, per toetsingscriterium. Ook hier zijn verschillende werkwijzen mogelijk, afhankelijk van de aard van de gegevens en de manier waarop men waarden toekent. Het 'waarderen' kan op tweeërlei wijze gebeuren: absoluut of relatief (cf. supra).

Om een rangorde aan te brengen heeft men een viertal mogelijkheden:

- Classificeren door toekenning van een gradatie

Deze wijze van waarderen is absoluut. Men beoordeelt elk alternatief via een extern referentiekader en komt tot een oordeel waarbij men het alternatief in een bepaalde klasse indeelt. De rangorde is deze van de klassen, binnen een klasse is er geen rangorde. Bijvoorbeeld: A - B - C, goed - matig - slecht.

- Rangschikken zonder algemeen label of meeteenheid

Deze wijze van waarden is relatief: elk alternatief wordt beoordeeld via een vergelijking met de andere alternatieven en komt tot een rangorde van alternatieven. De afstanden tussen twee alternatieven hebben op zichzelf geen absolute betekenis; ze zijn niet geijkt.

- Score toekennen

Deze wijze van waarden kan zowel absoluut als relatief zijn: men kent een waarde toe aan elk alternatief en komt tot een rangorde van alternatieven. De afstanden tussen twee alternatieven hebben ook een betekenis. Het voornaamste nadeel van deze werkwijze is echter dat het referentiekader vaak moeilijk gevalideerd kan worden doordat men met impliciete criteria werkt en de weging van criteria onduidelijk of arbitrair is.

- Positioneren op een metrische schaal

Deze wijze van waarden is absoluut: men kent een waardehoeveelheid toe in de mate waarmee ze kwantitatief gemeten wordt. Omdat dergelijke schaal geijkt is, leveren de afstanden tussen de alternatieven ook relevante informatie op.

Een laatste stap bestaat erin tot een syntheseoordeel te komen per beleidsalternatief en een finale rangorde van alternatieven. Een synthese over meerdere criteria is echter geen sinecure, doordat men vaak verschillende meetniveaus over de criteria heen heeft. De weging van de verschillende toetsingscriteria is ook zeer belangrijk. Hiertoe bestaan er multicriteria-technieken (cf. infra: deel 3).

➤ Multicriteria-analyse

5.4.8.4 Kritieke slaagfactoren

- Beschikbaarheid en kwaliteit van data

Om alternatieven te kunnen vergelijken op een of meerdere criteria heeft men voor elk alternatief basisgegevens nodig. De toepassing van monetaire technieken of multicriteria-analyse staat of valt met de beschikbaarheid van deze data. Bovendien is ook de kwaliteit van deze zogenaamde inputgegevens een aandachtspunt. Onbetrouwbare data vormen een slechte vergelijkingsbasis.

5.5 Ex post evaluatie

5.5.1 Inleiding

Het ex post evalueren van het beleid gebeurt vanuit een heel andere invalshoek als bij een ex ante evaluatie van beleid. Het voorwerp van de evaluatie is nu immers een bestaand beleid, een beleid dat al tot uitvoering is gebracht. Ex post evaluatie hangt dus samen met de laatste fase in de klassieke beleidscyclus, namelijk de beleidsbeoordeling. Voor ex post evaluaties is dan ook finaliteit verschillend van ex ante evaluaties.

De twee voornaamste motieven die ten grondslag liggen aan een ex post beleidsevaluatie zijn het afleggen van verantwoording en beleidsleren. Het zwaartepunt of de focus van ex post evaluaties is de beleidsimpact.

ex ante	ondersteunen van de beleidsplanning	beleidsinhoud ----- beleidsimpact
tussentijds	verbeteren van de uitvoering institutionele ontwikkeling	
ex post	verantwoording beleidsleren	

Tabel 10: Koppeling van evaluatiemotieven en -types

Dat neemt echter niet weg dat ook de beleidsinhoud mede voorwerp kan zijn van een ex post evaluatie. Zo kan men in een eerste stap de structuur en logica van het beleid uitklaren om vervolgens de scope van een ex post evaluatie te bepalen. Ook de processen van de beleidsuitvoering kunnen geëvalueerd worden in een ex post evaluatie. Zo kan men lessen trekken met het oog op het verbeteren van het management, een derde motief. Indirect kunnen ex post evaluaties ook ondersteuning bieden voor beslissingen in het kader van een volgende beleidscyclus. Ex post evaluaties kunnen met andere woorden onder alle vier voornaamste evaluatiemotieven geplaatst worden.

Analoog met ex ante evaluatie kunnen we ook met betrekking tot ex post evaluaties enkele verdere indelingen maken. Een eerste opdeling is gebaseerd op een eerder fundamenteel verschil in vraagstelling en focus van de ex post evaluatie. Voornamelijk het onderscheid tussen het evalueren van de mate van doelbereiking en het evalueren van de bijdrage van het beleid aan de doelbereiking is van wezenlijk belang. Het is terug te brengen op de twee essentiële vraagstellingen die een overheid zich voortdurend kan en moet stellen en die we in de inleiding van dit boekdeel hebben aangehaald. Het betreft de vragen “Doen we de dingen goed?” en “Doen we goede dingen?”.

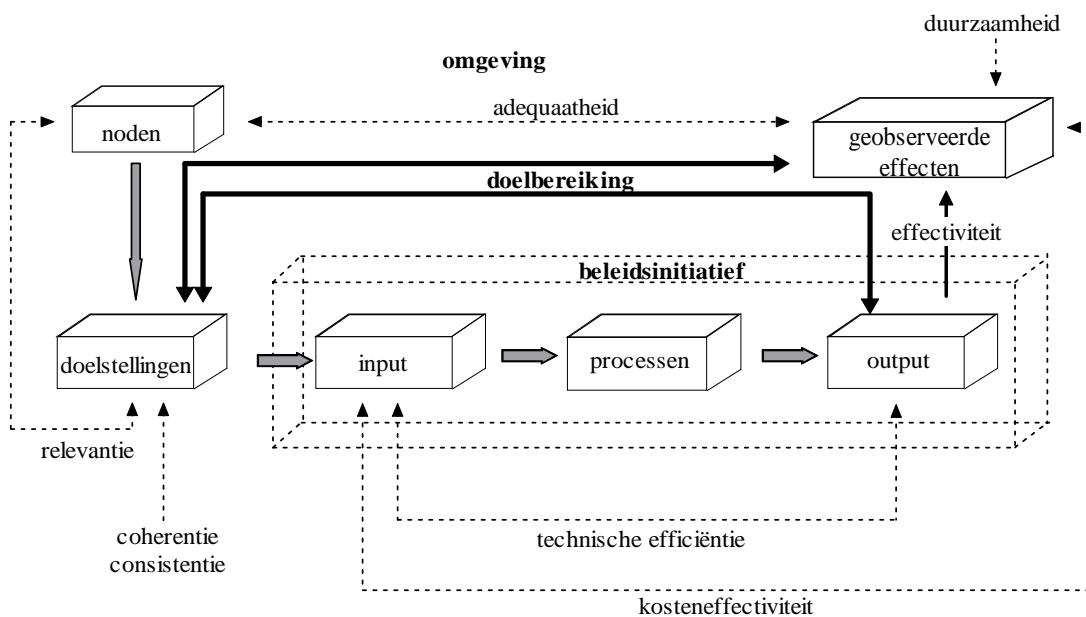
In de eerstvolgende sectie gaan we dieper in op het evalueren van de mate van doelbereiking. We staan stil bij het begrip doelbereiking en we geven aan wat de mogelijkheden en beperkingen van dergelijke evaluatie zijn.

Daarna komt de beleidseffectiviteit of de eigenlijke beleidsimpact aan bod. Dit vormt als het ware het kernthema van beleidsevaluatie. We gaan dieper in op de verschillende strategieën om effectiviteit vast te stellen.

5.5.2 Doelbereiking

5.5.2.1 Opzet en vraagstellingen

Een evaluatie naar de resultaten van het beleid kan in de eerste plaats opgebouwd worden rond de vraag in welke mate de vooropgestelde doelstellingen werden bereikt. Deze vraag kan men beantwoorden zonder de oorzaken van het al dan niet bereiken van de doelstellingen mee te nemen in de analyse. De mate waarin geconstateerde veranderingen inderdaad toe schrijven zijn aan de ontplooiende beleidsacties en in welke mate andere factoren die geen band met het beleid hebben daarin een rol spelen, blijft dan buiten beschouwing. Bijgevolg kan dit type van evaluatie ook geen gefundeerde uitspraak doen over dat aspect. Het evaluatiecriterium dat centraal staat is doelbereiking, zonder meer. De vraagstelling die hier centraal staat is dus beschrijvend van aard en heeft geen aspiraties om verbanden tussen factoren aan te tonen, laat staan om deze verbanden te verklaren.



Figuur 34: Doelbereiking als evaluatiecriterium

Wanneer we deze vraagstelling positioneren in het conceptueel kader van de beleidstheorie, zien we dat het veranderingsmodel hier in beeld komt, zij het slechts gedeeltelijk. We focussen namelijk op het eerste zicht enkel op de laatste component van dat model, de beleidsresultaten of effecten. Nochtans hoeft dit niet de enige component te zijn die hier onder de loep wordt genomen: ook de eerste component, de beleidsacties of output, kunnen in deze evaluatie meegenomen worden. Dat laatste heeft te maken met

het bestaan van twee categorieën van doelstellingen waarop we al eerder hebben gealludeerd. Strategische doelstellingen worden in het veranderingsmodel gereflecteerd in de component beleidsresultaten en hebben betrekking op de effecten van het beleid. Operationele doelstellingen vinden we in het model terug in de component beleidsacties en zijn outputgericht geformuleerd. Het evalueren op doelbereiking kan bijgevolg drie varianten aannemen:

- evalueren op operationele doelbereiking
- evalueren op strategische doelbereiking
- evalueren op operationele en strategische doelbereiking

Bovenstaand onderscheid is niet zonder belang. Het verschil tussen operationele en strategische doelstellingen heeft immers ook te maken met een 'grens' tussen overheid en omgeving. Daarmee bedoelen we niet dat operationele beleidsdoelstellingen enkel betrekking zouden hebben op de overheid zelf, los van de omgeving. In tegendeel, operationele doelstellingen hebben meestal ook een uitwerking in de maatschappelijke omgeving. Ze verwijzen immers zoals gezegd naar de output of de prestaties van de overheid, bijvoorbeeld de bouw van sociale woningen, controles van oppervlaktewaterkwaliteit of de organisatie van een cultuurmanifestatie.

Wel is het zo dat er een rechtstreekse band bestaat tussen het overheidsoptreden en de operationele doelstellingen, die er niet is ten aanzien van doelstellingen op strategisch niveau. Deze laatste verwijzen immers naar de maatschappelijke effecten en deze kunnen naast het overheidsoptreden ook beïnvloed worden door externe factoren, waar de overheid minder of geen vat op heeft. Met andere woorden, de overheid bepaalt zelf welke output zij realiseert, maar heeft niet op dezelfde wijze controle over de effecten die ze nastreeft. In de volgende sectie gaan verder in op de keuze tussen de drie aangehaalde varianten van doelbereiking. Deze keuze hangt samen met het motief van waaruit wordt geëvalueerd. Het evaluatiemotief vertelt ons ook meer over de wijze waarop evalueren op doelbereiking een meerwaarde kan verschaffen.

5.5.2.2 Meerwaarde

Het bepalen van de mate van doelbereiking kan een meerwaarde bieden in het kader van een drietal evaluatiemotieven. Een eerste motief de verbetering van (het management van) de beleidsuitvoering. Dit geeft meteen aan dat evalueren op doelbereiking niet exclusief onder de zogenaamde ex post evaluaties valt, maar ook tussentijds en dus al tijdens de implementatiefase kan gebeuren. Doelstellingen hangen immers vaak samen met een zekere tijdshorizon, dit wil zeggen het bereiken van een doelstelling wordt geprojecteerd naar een bepaald tijdstip in de toekomst. Vaak werkt men tijdens het uitvoeringsproces

gradueel naar de te bereiken doelstelling. Men kan daarbij tussentijdse doelstellingen formuleren. Evalueren op doelbereiking tijdens de uitvoering van het beleid kan dus nuttige informatie opleveren voor het management van organisatie(s) die instaan voor de beleidsuitvoering. Het proces van de beleidsuitvoering wordt dan opgevolgd aan de hand van informatie over de mate waarin men op schema zit om de doelstellingen te halen binnen de voorziene tijdsperiode. Door de uitvoering te monitoren kan het management indien nodig tijdig bijsturen op bepaalde aspecten van de implementatie. Wanneer het monitoringsysteem aangeeft dat er een probleem rijst, kan het management als eerste stap ook een tussentijdse evaluatie van de uitvoeringsprocessen ondernemen om de problematiek nader te onderzoeken en verbeterpunten te identificeren.

Een tweede motief voor het bepalen van de mate van doelbereiking is verantwoording. Aan de hand van dergelijke gegevens kan het management dat verantwoordelijk is voor de beleidsuitvoering verantwoording afleggen ten aanzien van de beleidsmakers. Dit gebeurt aan de hand van een ex post evaluatie op doelbereiking.

Ten derde is informatie over de mate waarin de doelstellingen worden bereikt die het beleid formuleerde, onontbeerlijk in het kader van beleidsleren. Om de vraag te beantwoorden waarom en hoe beleid al dan niet slaagt, zijn data over de mate van doelbereiking noodzakelijke vertrekgegevens. Een evaluatie op doelbereiking is echter op zich onvoldoende in het kader van beleidsleren. We komen op dit aspect verderop nog terug.

Welke variant(en) van een evaluatie op doelbereiking hangen nu normaliter samen met elk van deze motieven om te evalueren? De combinaties aangegeven in de onderstaande tabel zijn richtinggevend en vergen nadere toelichting.

<i>Variant</i>	Operationele doelbereiking (outputmeting)	Strategische doelbereiking (effectmeting)	Operationele en strategische doelbereiking
<i>Evaluatiemotief</i>			
Implementatiemanagement	X	-	-
Verantwoording	X	-	?
Beleidsleren	-	-	(X)

Tabel 11: Evaluatiemotieven en evaluatie op doelbereiking

In het kader van implementatiemanagement zal het bepalen van de mate van doelbereiking hoofdzakelijk betrekking hebben op de operationele doelstellingen. Deze zijn outputgericht en het uitvoeringsproces is primair gericht op het realiseren van een bepaalde output (dienstverlening, investeringen, ...).

Bij verantwoording als motief zijn de zaken complexer. Theoretisch kan men meerdere varianten van evalueren op doelbereiking verwachten, deels afhankelijk of het gaat om interne of externe verantwoording. Bij interne feedbacklijnen zullen voornamelijk operationele doelstellingen het voorwerp van de evaluatie zijn. Wanneer er externe feedbacklijnen zijn, kan men verwachten dat ook strategische doelstellingen bekeken worden en de evaluatie op doelbereiking naast een outputmeting ook een

effectmeting omvat. Immers, de finaliteit van de overheid ligt niet vervat in haar operationele doelstellingen, maar wel in de strategische doelstellingen. Ze hebben betrekking op de beoogde effecten in de omgeving of de laatste component van het veranderingsmodel. Het lijkt dan ook aannemelijk dat naast informatie over de output van de overheid, ook informatie over de evolutie in de maatschappelijke context wordt verzameld en gerapporteerd in het kader van externe verantwoording.

De bestuurlijke praktijk toont wellicht een dubbelzinnig beeld over effectmeting in het kader van verantwoording. Aan de ene kant is er de hypothese dat wanneer deze verantwoording gekoppeld is aan beslissingen over de allocatie van financiële middelen, effectmeting eerder zal worden vermeden. Een nadere analyse van beheersovereenkomsten als instrument in de sturings- en verantwoordingsrelaties binnen de overheid kan meer duidelijkheid brengen over de mate waarin deze hypothese gegrond is. Aan de andere kant overschat men mogelijk te vaak effectmetingen als verantwoordingselement. Het nut van het meten van effecten staat buiten kijf doordat het gaat om de eindresultaten waarop het beleid is gericht. Echter het louter samenbrengen van gegevens over output enerzijds en effecten anderzijds is meestal ontoereikend om de verantwoording een bredere scope dan doelbereiking mee te geven.

In het kader van beleidsleren mag men verwachten dat een evaluatie op doelbereiking zowel oog heeft voor outputmeting als effectmeting. Echter, we hebben deze optie in de tabel pro memorie aangekruist, omdat het evalueren op de mate van doelbereiking een noodzakelijke stap is voor beleidsleren maar niet voldoende. Beleidsleren omvat immers aspiraties die verder reiken dan enkel het vaststellen of de doelstellingen al dan niet gehaald worden. Beleidsleren is gebaseerd op causale verbanden, terwijl een evaluatie op doelbereiking enkel beschrijvende vragen kan beantwoorden.

Uit het bovenstaande blijken dan ook de fundamentele beperkingen van een evaluatie op doelbereiking. Gegevens over de mate van doelbereiking vormen nog geen aanwijzingen over beleidseffectiviteit. Precies dat laatste is een belangrijk evaluatiecriterium wanneer beleidsleren het centrale evaluatiemotief is. Het louter opvolgen van prestaties en effecten schiet tekort.

Maar ook in het kader van externe verantwoording levert een evaluatie op doelbereiking niet zonder meer voldoende en adequate gegevens aan.

Voor bepaalde verantwoordingssituaties en vanuit het motief beleidsleren kan men daarom een evaluatie op doelbereiking eerder beschouwen als een aanloop naar een evaluatie van beleidseffectiviteit. Deze laatste soort evaluatie komt in de volgende paragrafen aan bod.

5.5.2.3 Strategieën en technieken

Een evaluatie op doelbereiking kan ad hoc georganiseerd worden, maar men kan ook in een permanente opvolging voorzien, bijvoorbeeld in het kader van het management van de beleidsuitvoering. Een permanente opvolging van de mate van doelbereiking kan gebeuren aan de hand van een monitoringsysteem. In boekdeel 2 gaan we uitvoerig in op beleidsmonitoring. Hier bekijken we de stappen die men dient te nemen om de mate van doelbereiking vast te stellen. Er is een drietal stappen te doorlopen:

1. duidelijke en meetbare beleidsdoelstellingen
2. gegevensverzameling via monitoring
3. toetsing aan maatstaf voor doelbereiking

Een eerste stap is het concretiseren van de beleidsdoelstellingen. Het ligt immers voor de hand dat bij deze evaluatie de doelstellingen zelf de maatstaf zijn voor het evalueren. Zoals voorheen gesteld, is bij deze stap de interactie met de beleidsmakers als voornaamste stakeholders van groot belang. Mogelijk zijn de beleidsdoelen niet duidelijk genoeg geëxpliciteerd in de officiële beleidsdocumenten. Men kan dan teruggrijpen naar de techniek die werd uiteengezet in het kader van een ex ante evaluatie op de interne logica van het beleid. Deze houdt in dat men het veranderingsmodel van de achterliggende beleidstheorie probeert te reconstrueren. Dit veranderingsmodel weerspiegelt de vaak impliciete redeneringen achter de beleidsplanning en bevat de doelstellingen van het beleid onder de componenten beleidsacties en beleidsresultaten. Zeker voor een externe evaluator kan deze werkwijze nuttig zijn om het te evalueren beleid eerst in kaart te brengen.

Doelstellingen dienen te beantwoorden aan het zogenaamde SMART-principe: ze moeten specifiek, meetbaar, afgesproken, realistisch en tijdsgebonden zijn. Vervolgens dienen voor deze doelstellingen overeenstemmende indicatoren geformuleerd aan de hand waarvan de mate van doelbereiking gemeten kan worden. Ook voor het selecteren van indicatoren kan het SMART-principe als leidraad dienen. Verdere aandachtspunten zijn de afdeckingsgraad van de doelstellingen door de indicatorenset en de gevoeligheid van een indicator voor wijzigingen in het meetobject. Deze aspecten komen in deel 2 van de handleiding uitgebreider aan bod.

Een tweede stap betreft de datavergaring zelf. Belangrijk hierbij is een aantal eisen te respecteren die verbonden zijn aan het proces van datavergaring. Het verzamelen van gegevens dient in de eerste plaats zo efficiënt mogelijk te verlopen, met een minimale administratieve last. De selectie van informatiebronnen dient weloverwogen met een voorkeur om eerst de bestaande databronnen te raadplegen en te beoordelen op hun adequaatheid. De datavergaring zelf dient te beantwoorden aan de

principes van validiteit en betrouwbaarheid. Ook hier verwijzen we de lezer verder naar boekdeel 2 voor een uitdieping van deze aspecten.

De derde en laatste stap bij het evalueren op doelbereiking is het interpreteren van de verzamelde gegevens. Doordat het evalueren op doelstellingen eerder beschrijvend van aard is, beperkt de interpretatie van de gegevens zich tot de vraag met welke maatstaf voor doelbereiking men gaat vergelijken. Het concept doelbereiking heeft namelijk verschillende dimensies. Er zijn bijgevolg verschillende vergelijkingsbases denkbaar waarmee men een reële situatie na de beleidsuitvoering (B) kan vergelijken. Een vergelijkingsbasis of maatstaf bestaat ten eerste uit één of meerdere van onderstaande situaties:

- de uitgangssituatie voor de beleidsuitvoering (I)
- de beoogde of verwachte situatie na de beleidsuitvoering (E)
- de ideale, d.i. meest optimale situatie (O)

Ten tweede kan de mate van doelbereiking uitgedrukt worden in absolute of relatieve termen wanneer de gegevens gekwantificeerd zijn. Wanneer een verbetering het doel is, staan hieronder enkele voorbeelden van maten voor doelbereiking geformuleerd:

- (1) de bereikte verbetering: $B - I$
- (2) de bereikte procentuele verbetering $\frac{100(B - I)}{I}$
- (3) het verschil tussen de verwachte en bereikte situatie: $E - B$
- (4) het procentuele gedeelte van de verwachte verbetering, dat bereikt is: $\frac{100(B - I)}{E - I}$
- (5) het verschil tussen de bereikte en optimale situatie: $O - B$
- (6) het procentuele verschil tussen de verwachte en bereikte situatie: $\frac{100(E - B)}{E}$
- (7) het procentuele verschil tussen de bereikte en de verwachte verbetering: $\frac{100(E - B)}{E - I}$

Elk van deze maten hebben voor- en nadelen. Toch verschillen de eerste twee maatstaven van de andere op een belangrijk punt; ze hanteren enkel reële variabelen (I en B). De andere maten hebben een

normatieve component doordat hetzij de verwachte (E), hetzij de optimale situatie (O) onderdeel vormt van de vergelijkingsbasis. De waarde van de eerste twee maten is niet afhankelijk van de gestelde doelen (E) noch van de oorspronkelijke afwijking tussen de uitgangssituatie en de gewenste situatie (= E - I). Het resultaat is mogelijk het effect van het beleid, naast alternatieve verklaringen. Enkel het teken van het resultaat wordt beïnvloed door de soort verwachte verandering (vermindering of verhoging).

In de andere vermelde maten zit telkens een *normering* ingewerkt, aan de hand van de verwachte situatie (E) of de optimale situatie (O). Met andere woorden, de bereikte situatie wordt beoordeeld en deze waardering bepaalt mee het resultaat voor de waarde van deze maatstaven. De waarde van deze maten hangt dus af van het ambitieniveau of de hoogte van de gestelde doelen (E, waarbij $E \leq O$).

Welke maat voor doelbereiking is preferabel? Een randvoorwaarde voor de keuze van een maat is dat er gegevens beschikbaar zijn voor de variabelen die in de vergelijkingsbasis worden opgenomen. Idealiter wordt de gegevensverzameling dus afgestemd op de maat voor doelbereiking die men wil hanteren. Bij de selectie van een maat is het van belang vooraf te weten of deze ook van tel is voor een evaluatie op beleidseffectiviteit, d.i. de mate waarin de doelbereiking daadwerkelijk toe te schrijven is aan het beleid. In dat geval is een maat voor doelbereiking zonder normatieve component aangewezen. Beleidseffectiviteit refereert immers steeds naar een reële situatie waarbij men wil nagaan in welke mate het beleid heeft bijgedragen aan deze reële situatie. Wanneer in die analyse ook ambitieniveaus worden betrokken, zorgt dit voor extra complexiteit. Hoe het betrekken van normatieve elementen dergelijke evaluatie kan bemoeilijken, blijkt uit onderstaand voorbeeld.

Twee gemeenten voeren een gelijkaardig verkeersveiligheidsbeleid. De tabel bevat data over het aantal ongevallen bij de uitgangssituatie (I) en na de uitvoering van de maatregelen (B). Tevens geeft ze de beoogde of verwachte situatie na de uitvoering (E) aan.

# ongevallen naar situatie:	I	B	E
Gemeente X	10	4	6
Gemeente Y	10	4	3

Wanneer als maat voor doelbereiking het verschil tussen de beoogde en de bereikte situatie wordt genomen (E - B), geeft dit voor gemeente X de score 2 en voor gemeente Y de score -1. Met andere woorden: gemeente X heeft de doelstelling ruimschoots bereikt, terwijl het resultaat in gemeente Y onder de verwachtingen blijft.

Bij een verdere analyse van de doelbereiking in de twee gemeenten moet men dus een positieve en een negatieve score verklaren. Nochtans was de uitgangssituatie identiek, en is er geen verschil tussen de

reële situaties na de beleidsuitvoering noch in de wijze waarop het beleid hierop invloed heeft. Enkel het ambitieniveau verschilde en dat bepaalt de in dit voorbeeld scherp tegengestelde resultaten voor de mate van doelbereiking.

Met het oog op een latere evaluatie op beleidseffectiviteit zijn dus voornamelijk de eerste twee maten voor doelbereiking aan te bevelen: de bereikte verbetering en de bereikte procentuele verbetering. De andere maatstaven voor doelbereiking zijn daarom niet nutteloos. Maar men dient de waarden op deze maatstaven wel met de nodige omzichtigheid te interpreteren, precies omwille van de normatieve component.

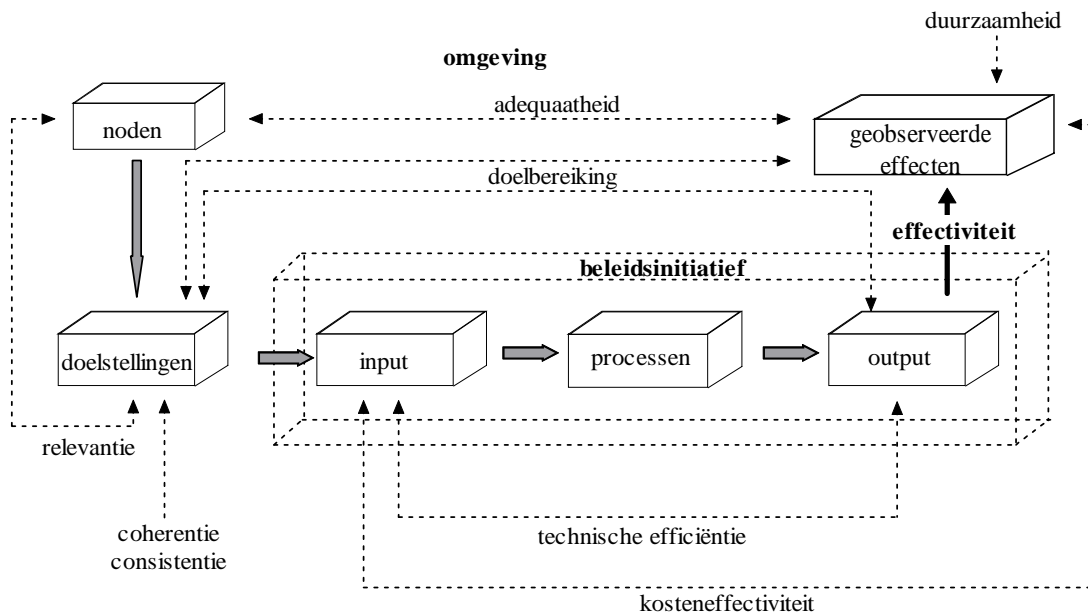
Soms voldoen doelen niet aan het SMART-principe en zijn ze bijvoorbeeld onrealistisch hoog of laag gesteld uit politieke overwegingen. Een andere tekortkoming kan zijn dat de doelen niet in meetbare termen zijn gesteld. Ook in dat geval is een vergelijking tussen enerzijds de ontwikkeling van een initiële naar een reëel bereikte situatie en een doel anderzijds wel mogelijk, doordat een doel steeds wel een richting zal aangeven (status quo, vermeerdering of vermindering). Bepaalde maten voor doelbereiking zijn niet hanteerbaar in dat geval, maar de eerste twee maatstaven uit de lijst zijn niet onmogelijk.

Een andere context doet zich voor wanneer de doelstellingen wijzigen tijdens de periode van de beleidsuitvoering. In dat geval verdient de definiëring van de uitgangssituatie bijzondere aandacht.

5.5.3 Beleidseffectiviteit: inleiding

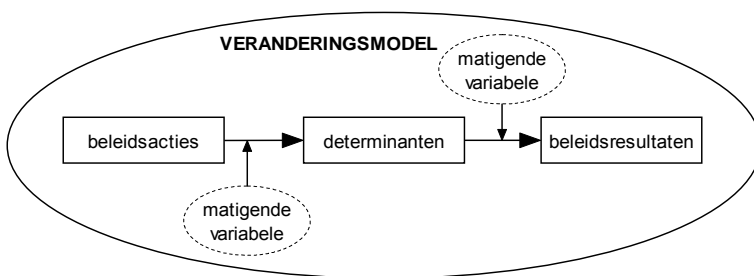
5.5.3.1 Opzet en vraagstellingen

Een tweede vraagstelling in het kader van de resultaten en impact van het beleid is: in welke mate heeft het beleid bijgedragen aan het bereiken van bepaalde doelstellingen? Dit is de vraag naar beleidseffectiviteit. Een beleid is effectief voor zover de vastgestelde effecten het gevolg is van het beleid. Immers, het bereiken van de doelstellingen kan ook toe te schrijven zijn aan andere factoren, naast het beleid zelf. Deze andere factoren spelen een rol in zogenaamde alternatieve of rivaliserende verklaringen.



Figuur 35: Beleidseffectiviteit als evaluatiecriterium

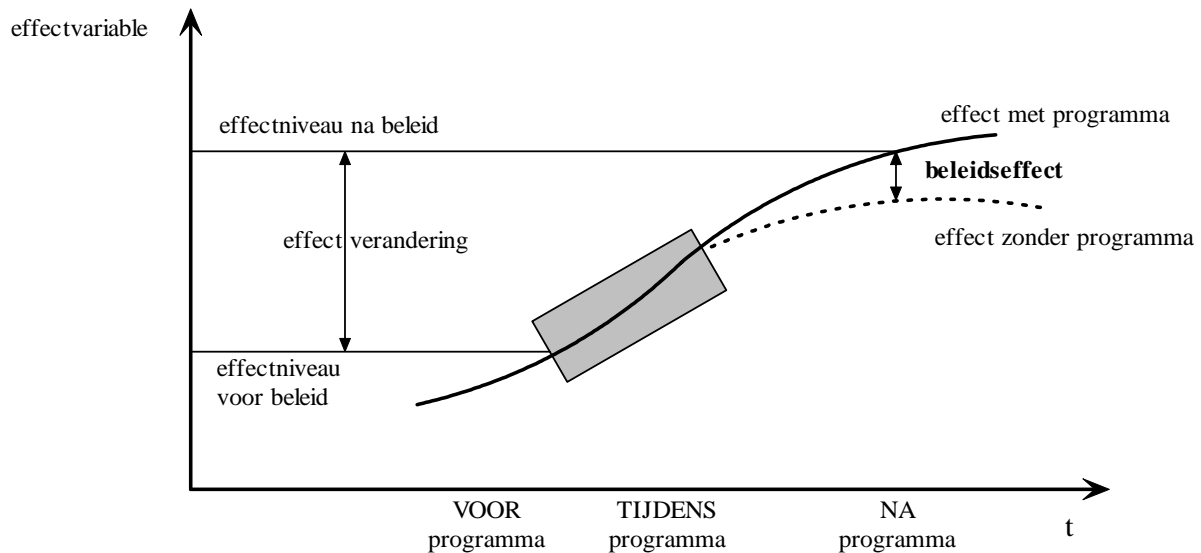
Met beleidseffectiviteit als centraal evaluatiecriterium, is de opzet van de evaluatie gericht op het vaststellen van verbanden en meer bepaald causale relaties. Causaliteit verwijst naar oorzaak en gevolg relaties, en in deze context naar relaties tussen het overheidsoptreden en de maatschappelijke effecten. Vertaald naar het conceptueel raamwerk van beleidstheorie, ligt de focus dus op het veranderingsmodel. Anders dan bij een evaluatie op doelbereiking gaat het niet om afzonderlijke componenten van dit model, maar om de dynamiek die in het model vervat ligt: de keten van beleidsacties over determinanten naar beleidsresultaten.



Figuur 36: Het veranderingsmodel

Om te kunnen spreken van een causale relatie van oorzaak A naar gevolg B moeten drie randvoorwaarden vervuld zijn. Ten eerste is er een bepaalde volgorde in de tijd: A moet B vooraf gaan. Ten tweede moet er samenhang zijn tussen A en B en wel zo dat er co-variatie is: meer A gaat meestal samen met hetzij meer B, hetzij minder B. Een laatste voorwaarde is dat er geen alternatieve verklaringen voor deze relatie tussen oorzaak en gevolg bestaan.

De vraag of beleid effect heeft is inderdaad de vraag of het een verschil bestaat tussen de situatie wanneer er beleid is en deze bij de afwezigheid van beleid. Deze “**policy on vs. policy off**” vergelijking kan schematisch als volgt worden voorgesteld.



Figuur 37: Effectniveaus, effectverandering en beleidseffect

De grafiek geeft de mate weer waarin een effectvariabele (Y-as) waarneembaar is doorheen de tijd (X-as). Een variabele is een meetbare factor die verschillende waarden kan aannemen. Een effect duidt op een verandering doorheen de tijd, ten gevolge van een oorzaak. Effecten hebben betrekking op een fenomeen, een kenmerk enz. Voorbeelden zijn waterkwaliteit, rookfrequentie, verkeersdrukke, klantentevredenheid, e.d. Een effectvariabele is dus een meetbare factor die betrekking heeft op een effect. Effecten kunnen gemeten worden aan de hand van indicatoren.

Het beleid streeft bepaalde effecten na: het probeert bepaalde variabelen of factoren te beïnvloeden. Beoogde effecten zijn dan terug te vinden in de beleidsdoelstellingen. Bovenstaande grafiek toont de evolutie van de (gemeten) waarden die een effectvariabele aanneemt doorheen de tijd. In dit tijdsverloop wordt een beleidsinitiatief (programma, maatregel) uitgevoerd.

De horizontale as geeft het tijdsverloop aan, waarbij drie fases werden onderscheiden: een periode die voorafgaat aan de beleidsuitvoering, een tijdssegment dat samenvalt met de duur van de beleidsuitvoering en een periode die volgt op de beëindiging van de beleidsuitvoering.

De doorlopende lijn geeft het effectniveau weer zoals vastgesteld op regelmatige tijdstippen via de indicator. Het gaat niet om een rechte horizontale lijn, maar om een golvende curve: dit geeft aan dat de waarde van de effectvariabele geen constante is, maar verandert ten gevolge van veel oorzaken en

omstandigheden die extern ten aanzien van het beleid zijn. Het niveau van de waterkwaliteit bijvoorbeeld zal fluctueren naargelang de graad van industriële activiteit in de regio en weersomstandigheden. Zo kan hevige regenval bijvoorbeeld de concentraties van schadelijke stoffen in het water verdunnen.

De mate waarin een bepaald effect optreedt, vertelt echter nog niets over de effectiviteit van het beleid. Anderzijds kunnen de mogelijkheden op dat vlak al wel afgelijnd worden aan de hand van extreme resultaten. Stel dat na afloop van een mediacampagne in het kader van rookpreventie men vaststelt dat alle tieners uit de doelgroep roken, dan kunnen we er vrij zeker van zijn dat de campagne geen groot succes was of zelfs contraproductief is geweest. Wanneer echter wordt vastgesteld dat geen enkele tiener meer rookt, mag men redelijkerwijs verwachten dat de campagne heeft gewerkt, vermits dergelijk resultaat boven de verwachtingen lag. Natuurlijk zullen dergelijke extreme resultaten zelden in de praktijk voorkomen en meestal zullen effectmetingen alleen niet zomaar geïnterpreteerd kunnen worden in termen van succes of falen.

Wanneer we effectmetingen voor en na de beleidsuitvoering vergelijken, krijgen we wel zicht op de ontwikkeling of evolutie. De ontwikkeling kan een status-quo vertonen of een verandering. Bij een verandering is de richting (+ of -) belangrijk. Maar ook op basis daarvan kan men nog niets afleiden met betrekking van het aandeel van het beleid in deze verandering. Het is hier ook dat de beperking van het evalueren op doelbereiking zich manifesteert (cf. supra).

De golvende stippellijn geeft het traject aan voor de effectvariabele dat men had vastgesteld wanneer er geen beleid was uitgevoerd. Door het vergelijken van de doorlopende curve en de stippellijn kunnen we afleiden dat het effectniveau in dit voorbeeld ook was gestegen zonder het beleid, maar niet zo sterk. Het effectniveau op het tijdstip na de beleidsuitvoering ligt hoger in de “policy on” situatie dan in de “policy off” situatie. Het verschil tussen beide niveaus is de netto meerwaarde die het beleid heeft op geleverd en die er niet was geweest zonder het beleid. Voor dit verschil kunnen we de term beleidseffect met recht en rede gebruiken. Het is dat deel van het effect dat effectief aan het beleid kan worden toegedicht.

Het schatten van beleidseffecten is de meest geavanceerde evaluatieopzet. De moeilijkheden liggen precies vevat in het verschil dat in het schema werd aangeduid tussen het effect dat daadwerkelijk optreedt en het effect dat zou hebben plaatsgevonden bij afwezigheid van het beleid. In de praktijk is het namelijk onmogelijk om tegelijkertijd de resultaten voor de “policy on” en “policy off” scenario’s te kennen voor een zelfde eenheid (rivier, stad, persoon, kruispunt, e.d.). De waterkwaliteit in de Schelde bijvoorbeeld kan niet tegelijkertijd gemeten worden in de situatie met en zonder de uitvoering van beleid dat de kwaliteit wil verhogen.

Er stelt zich dus een probleem. De effecten in het reële, “policy on” scenario kunnen we meten. Maar op één of andere wijze zullen we de effecten bij afwezigheid van beleid moeten kunnen achterhalen. Het “policy off” scenario is een hypothetisch scenario dat eerder gesimuleerd of gededuceerd moet worden, dan wel geobserveerd.

De moeilijkheidsgraad van de vraagstelling naar beleidseffectiviteit hangt van verschillende factoren af. De aard van het beleidsveld, de complexiteit van het beleid dat geëvalueerd wordt, en het tijdsverschil tussen het beleid en het moment dat de effecten optreden spelen allemaal een rol

5.5.3.2 Meerwaarde

De vraagstelling naar het al dan niet bestaan van beleidseffectiviteit is zonder twijfel één van de belangrijkste bij het evalueren van overheidsbeleid. Ze weerspiegelt één van de twee fundamentele vragen die een overheid zich kan stellen: “Doen we goede dingen?”. Die vraag is zomogelijk nog essentiëler dan de vraag: “Doen we de dingen goed?”. Want ze grijpt terug naar de *raison d’être* of bestaansreden van de overheid: een verschil maken voor de samenleving.

Het antwoord op de vraag of het beleid werkt of niet, is dan ook belangrijk in het kader van verantwoording. De vraag naar doelbereiking (cf. supra) volstaat in principe niet om de inzet van publieke middelen te verantwoorden. De overheid kan een aanzienlijke hoeveelheid resources inzetten. Het scenario waarbij de vooropgestelde doelstellingen worden bereikt zonder dat het beleid iets bijdraagt tot het halen van die doelstellingen, is niet denkbeeldig. Zekerheid over de bijdrage van het beleid aan maatschappelijke problemen draagt bij tot de legitimiteit van het beleid. De vraag naar beleidseffectiviteit is dan ook cruciaal in het kader van een democratische controle op het optreden van de overheid.

Een evaluatie op beleidseffectiviteit is ook functioneel vanuit het motief van beleidsleren. Op basis van dergelijke informatie kan een beslissing tot het verder zetten, het bijsturen of het stopzetten van een bepaald beleid onderbouwd worden. Ook levert dit type van evaluatie ondersteuning voor beleidsoverdracht: het overnemen van een bepaalde beleidsaanpak of toepassing van een beleidsinstrument in een andere sector of plaats, op basis van aangetoonde positieve resultaten elders. Ook wanneer een innovatief beleid wordt opgestart kan een evaluatie op beleidseffectiviteit een belangrijke meerwaarde opleveren.

5.5.3.3 Strategieën en technieken

De evaluator kan verschillende werkwijzen volgen om na te gaan of en in welke mate het beleid effectief is, dit wil zeggen: in welke mate het verantwoordelijk is voor de geobserveerde effecten.

Het verschil tussen de werkwijzen heeft zozeer betrekking is geen verhaal van andere accenten of een andere focus. In essentie gaat het om verschillende strategieën om beleidseffectiviteit aan te tonen. Elke strategie hanteert een andere benadering. We spreken dan ook over verschillende evaluatiedesigns:

- experimenteel design
- quasi-experimenteel design
- alternatief of kwalitatief design

Anders dan bij de subtypes van ex ante evaluatie gaat hier niet steeds om een vrijblijvende keuze naargelang de noden of interesse van de beleidsmakers en andere stakeholders. De keuze voor een evaluatiedesign in het kader van het evalueren van beleidseffectiviteit wordt mee gestuurd door praktische en ethische elementen (cf. infra).

Kenmerkend voor het experimentele en quasi-experimentele design is dat ze met vergelijkingsgroepen werken die de “policy-on” en “policy-off” situaties reflecteren. In het derde design werkt men niet met vergelijkingsgroepen. De werkwijze richt zich op het begrijpen hoe het beleid werkt en is bijgevolg meer kwalitatief van aard.

In de volgende paragrafen komen de verschillende strategieën aan bod. Vooraf dient er nog op gewezen dat de alternatieve strategieën verschillen in verklaringskracht. Het experimentele design biedt de sterkste verklaringskracht en dus de hoogste mate van zekerheid omtrent de effectiviteit van het beleid. In de praktijk zullen de randvoorwaarden voor het experimentele design niet steeds aanwezig zijn; het quasi-experimentele design vormt dan een terugvalpositie. Het alternatieve of kwalitatieve design omvat een diepteanalyse van beleidsprocessen en hoe deze bijdragen aan het bereiken van de beoogde effecten.

5.5.4 Experimenteel design

Het experimenteel design wordt wel eens de gouden standaard of het vlaggenschip genoemd wanneer het erop aan komt om causale verbanden te achterhalen. En dus wordt het ook als het sterkste evaluatiedesign beschouwd om beleidseffectiviteit vast te stellen.

5.5.4 1 Evaluatiekader: “policy on” vs. “policy off”

Bij een experimenteel design werkt men met twee of meerdere groepen waarbij men een onderscheid construeert tussen:

- een of meerdere groepen waarbij men beleid(salternatieven) toepast (de beleidsgroep);
- een groep waarbij het beleid niet wordt toegepast (de controlegroep).

Op deze wijze kan men de “policy on” en “policy off” situaties (cf. supra) gaan vergelijken en de beleidseffectiviteit gaan bepalen. De aard van de samenstellende eenheden van deze groepen hangt af van de betrokken beleidssector. Andere termen zijn analyse-eenheden of evaluanda. Het kan gaan om personen, bijvoorbeeld in het kader van een dienstverlening binnen de gezondheidszorg. De analyse-eenheden kunnen ook gemeenten zijn in het kader van subsidies voor een lokaal cultuurbeleid, of categorieën van bedrijven, kruispunten, oppervlaktewateren, enz.

5.5.4.2 Vergelijkingsbasis

Kenmerkend voor het experimenteel design is dat de groepen ‘at random’ of op basis van toeval worden samengesteld. Dit is tevens het belangrijkste verschilpunt met quasi-experimentele designs. Bij een at random samenstelling heeft elke eenheid evenveel kans om in de beleidsgroep of in de controlegroep terecht te komen.

De ‘at random’ of toevallige samenstelling van de groepen is namelijk de beste werkwijze om de kritische voorwaarde van equivalentie tussen de beleids- en controle groep te realiseren. Het begrip equivalentie is op te splitsen naar een aantal dimensies:

- Identieke samenstelling: de twee groepen vertonen dezelfde kenmerken die relevant zijn met betrekking tot het beleid en de effecten ervan.
- Identieke instelling: de twee groepen hebben a priori dezelfde mate van openstelling tot het beleid. Bij afwezigheid van het beleid in beide groepen zouden beide groepen eenzelfde effectverandering laten opmeten.
- Identieke ervaringen: doorheen de tijd zijn de twee groepen op dezelfde manier onderhevig aan tijdsgebonden processen, externe gebeurtenissen enz.

Theoretisch perfecte equivalentie waarbij elke eenheid in de ene groep identiek is met een eenheid in de andere groep is de praktijk niet haalbaar maar ook niet noodzakelijk. Het is voldoende wanneer aan de voorwaarde van equivalentie is voldaan op geaggregeerd niveau. Zolang verschillpunten tussen de twee groepen geen invloed hebben op de effecten die bestudeerd worden, spelen deze geen rol. In het andere geval kunnen echter fouten optreden in het vaststellen van de beleidseffectiviteit. De kern van de zaak is immers om het beleid als mogelijke verklaringsgrond voor de vastgestelde effecten af te zonderen van andere mogelijke oorzaken.

De procedure van toevallige toewijzing mag echter zelf niet geïmproviseerd zijn. Het dient een systematische werkwijze te zijn waarbij er een gelijke kans is voor elke eenheid om aan ofwel de beleidsgroep ofwel de controlegroep toegewezen te worden. De procedure moet garanderen dat externe factoren die naast het beleid invloed kunnen hebben op de relevante effecten in gelijke mate aanwezig zijn in beide groepen, met uitzondering van toevallige fluctuaties.

De evaluator moet dus een uitgesproken, op toeval gebaseerde, procedure hanteren om aan deze voorwaarde te voldoen. Vaak gebruikt men daarvoor tabellen met op toevalsbasis samengestelde nummers (door computers). De beslissing om een bepaalde eenheid aan een groep toe te wijzen is dan enkel gebaseerd op het volgende nummer in de tabel (bv. even nummer voor beleidsgroep en oneven nummer voor controlegroep).

Toevallige fluctuaties zouden echter vervagen tot nihil wanneer de procedure een oneindig aantal maal zou worden herhaald. Statistisch kan dan ook de kans worden berekend dat een bepaald verschil optreedt in een reeks van toevallige toewijzingen. Dit is belangrijk voor het zogenaamde testen op significantie. Daarbij gaat men na of een bepaald verschil in effect tussen de twee groepen waarschijnlijk door toeval of eerder door toedoen van het beleid optreedt. Omdat de aan- of afwezigheid van het beleid in een adequaat experimenteel design het enige verschil is naast een op basis van toeval voorkomend verschil, worden dergelijke beoordelingen de basis om beleidseffectiviteit vast te stellen.

De enige implicatie van de rol van toeval en het testen op significantie is dat de groepen van een bepaalde grootte moeten zijn. Statistische equivalentie neemt toe met de grootte van de groepen. Vergelijking dit met het tossen op kop of munt: de kans op kop of munt is statistisch 50/50. Men zal in de praktijk deze verdeling beter benaderen wanneer men de munt 1000 keer opgooit dan wel slechts 4 keer. Voor verdere details hierover verwijzen we de lezer naar gespecialiseerde literatuur over statistiek.

5.5.4.3 Meten en vergelijken

Vervolgens wordt het beleid toegepast bij de ene groep, en niet bij de andere groep. In een uitgebreide variant zullen verschillende beleidsalternatieven op verschillende groepen worden toegepast, terwijl bij één groep geen beleid wordt toegepast. Vervolgens gaat men effectmetingen uitvoeren bij de zogenaamde experimenteergroep (met beleid) en de zogenaamde controle groep (zonder beleid), waarbij men verschilpunten gaat toeschrijven aan de toepassing van het beleid.

Onderstaand schema geeft de werkwijze weer bij de effectmetingen en vergelijking tussen beleidsgroep en controlegroep.

	Effectmetingen		
	Voor	Na	Vershil
Beleidsgroep	B1	B2	$B = B2 - B1$
Controlegroep	C1	C2	$C = C2 - C1$

Tabel 12: Effectmetingen in een experimenteel design

Het beleidseffect = $B - C$, waarbij:

- B1, C1 = metingen van de effectvariabele vóór het beleidsinitiatief werd toegepast, respectievelijk bij de beleids- en controlegroep;
- B2, C2 = metingen van de effectvariabele nadat het beleidsinitiatief werd toegepast, respectievelijk bij de beleids- en controlegroep;
- B, C = de effectverandering voor respectievelijk de beleids- en controlegroep.

Voor elke groep wordt de waarde van effectvariabele opgetekend voor en na de beleidsuitvoering in de beleidsgroep. Vervolgens wordt voor elke groep de gemiddelde effectverandering berekend. Wanneer aan de voorwaarde van equivalentie is voldaan, valt uit het resultaat van de controlegroep af te leiden wat het resultaat voor de beleidsgroep was geweest indien ook zij geen toepassing van het beleid had gekregen. Wanneer het resultaat (effectverandering) van de controlegroep wordt afgetrokken van het resultaat (effectverandering) van de beleidsgroep, is de restwaarde een rechtstreekse schatting van het gemiddelde beleidsinvloed op de effectvariabele, d.i. het beleidseffect.

Echter, het verschil tussen beide groepen bevat ook een toevalsfactor die voortspuit uit de procedure toevallige toewijzing (cf. supra). Het numerieke verschil tussen de gemiddelde effectscores kan dan ook niet zomaar worden geïnterpreteerd als het beleidseffect. Het verschil moet nog getest op statistisch significantie. Conventionele technieken hiervoor zijn de t-test, variantieanalyse of covariantieanalyse.

In bovenstaand schema werd uitgegaan dat het beleidseffect af te lezen is op basis van een voor en na situatie ten aanzien van de beleidsuitvoering. In bepaalde gevallen is dergelijke meetprocedure onmogelijk. Bij de evaluatie van de effectiviteit van een preventiecampagne rond tienerzwangerschappen is zwangerschap de relevante effectvariabele. De campagne is echter vanzelfsprekend gericht op tieners die nog niet zwanger zijn. Zowel de beleids- en controlegroep bestaan uit meisjes die voor de start van de campagne nog niet zwanger zijn. De waarde van de effectvariabele is gelijk aan nul. De voormeting levert dus geen verschillende waarden op. Enkel de nameting van het aantal zwangerschappen kan een verschil opleveren tussen beide groepen (met en zonder campagne).

De inschatting van netto beleidseffecten via het experimenteel evaluatiedesign kan ondersteund worden door twee strategieën voor gegevensverzameling.

Een eerste betreft het uitvoeren van meerdere metingen van de effectvariabele en dit zowel voor als na de toepassing van het beleid dat wordt geëvalueerd. De algemene regel is dat des te meer metingen van de effectvariabele worden uitgevoerd, des te meer accuraat de schatting zal zijn van het beleidseffect.

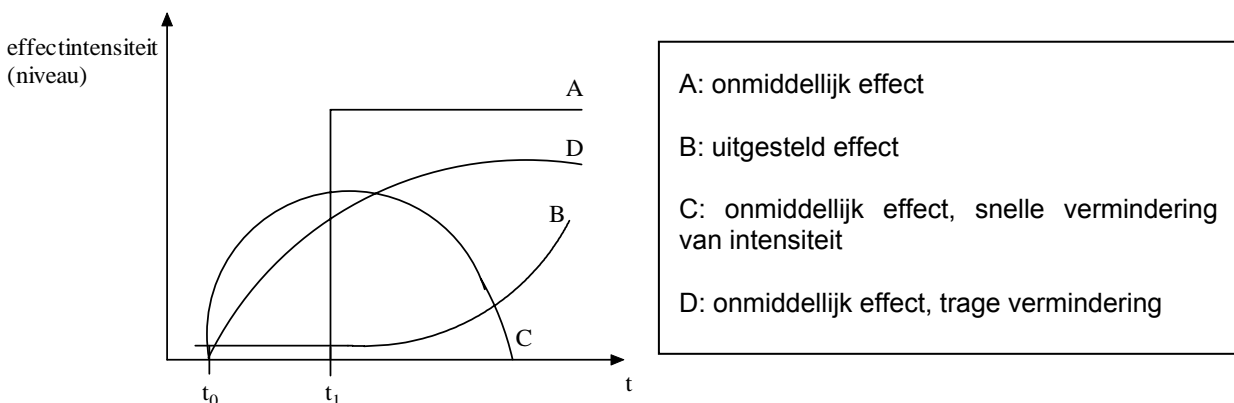
Een tweede strategie is om metingen periodiek te herhalen tijdens en na de uitvoering van het beleid. Het helpt de evaluator om zinvolle informatie aan te reiken over hoe beleidsinterventies doorheen de tijd werken. Dit kan belangrijke informatie opleveren ook voor de beleidsmakers. Neem het voorbeeld van een avondcursus volwassenenonderwijs die de overheid inricht of subsidieert. Wanneer men tot de vaststelling komt dat de vorming de meeste effecten produceert tijdens de eerste vier weken van een zes weken durende cursus, dan kan het inkorten van de opleiding een te overwegen keuze zijn die kostenbesparend is zonder sterk te tornen aan de effectiviteit van het beleidsprogramma.

Periodieke nametingen zullen daarnaast ook kunnen bijdragen tot inzichten over het optreden en het verloop van beleidseffecten. Afhankelijk van de concrete context, kunnen effecten immers opgedeeld worden in verschillende soorten.

Beleids effecten -voor zover ze daadwerkelijk optreden - kunnen namelijk verschillen naar:

- tijdstip: onmiddellijk zichtbaar vs. uitgesteld
- duur: kort vs. lang
- intensiteit: laag vs. hoog

De combinatie van deze kenmerken resulteert in een soort effect dat een bepaalde evolutie volgt. Onderstaande figuur toont enkele voorbeelden. De beleidsuitvoering loopt van t_0 tot en met t_1 .



Figuur 38: Effecttypes op basis van tijdstip, duur en intensiteit

Kennis over het soort effect dat men kan verwachten is bijgevolg belang voor de monitoring van effectvariabelen. Tevens is het essentieel om een accurate beoordeling te maken van beleidseffecten. Ook hierin speelt de beleidstheorie en achterliggende kennis en ervaringen met betrekking tot de specifieke beleidscontext een belangrijke rol.

De aangehaalde strategieën voor gegevensverzameling duiden op hun beurt ook op het belang van monitoring van beleid. Deel 2 van de handleiding gaat op dit aspect uitgebreid in.

5.5.4.4 Randvoorwaarden

De meerwaarde van een evaluatie op beleidsimpact via het experimenteel design wordt verregaand erkend. De werkwijze kan op blijvende aandacht rekenen omdat het de meest solide methode is om beleid op effectiviteit te evalueren. Dergelijke evaluaties zijn ook haalbaar onder de aangewezen omstandigheden.

Ondanks dat ze de sterkste garanties bieden naar de geldigheid van de resultaten, steekt men lang niet bij alle impactevaluaties van wal met het 'vlaggenschip'. De meest solide methode vraagt dan ook een secure toepassing en de kostprijs kan oplopen wanneer ze op grote schaal wordt toegepast. Ze vergt een niet te onderschatten investering in tijd en expertise, en naargelang de beleidssector en het concrete evaluandum, ook de samenwerking van mogelijke doelgroepen en dienstverleners. Gelet op deze randvoorwaarden, dient de keuze voor het experimenteel design weloverwogen beslissing te zijn. Daarbij dient men ook in te schatten in welke mate het te evalueren beleid prioriteit heeft voor beleidsmakers. Wanneer er van hun kant weinig aandacht is, of wanneer men bij voorbaat sterke aanwijzingen heeft dat het beleid - ongeacht de bevindingen van de evaluatie- niet zal worden voortgezet, is het eerder raadzaam om deze veeleisende werkwijze niet te volgen.

5.5.4.5 Beperkingen op het hanteren van het experimenteel design

Los van bovenstaande randvoorwaarden, is het hanteren van een experimenteel design niet steeds een vanzelfsprekende optie. Er zijn immers ook enkele elementen aan te wijzen die het gebruik deels of geheel in de weg kunnen staan.

Een eerste element is van praktische aard en heeft betrekking op de fase waarin het beleid zich bevindt. Het is niet denkbeeldig dat men voor een innovatief beleid of demonstratieproject een evaluatie van de beleidsimpact wil organiseren aan de hand van een experimenteel design. Dergelijke ex ante toepassing is mogelijk wanneer er voldoende middelen zijn voorzien. Ze kan nuttig informatie opleveren in het kader van een beslissing om het beleid wel of niet uit te breiden naar een grotere schaal of om er een permanent beleid van te maken.

Echter van zodra een beleid is gepland en in de implementatiefase zit, dient men er op te letten om het experimentele design niet te snel toe te passen. De impactvraagstelling kan immers leiden tot resultaten die minder relevant zijn vooraleer het beleid een zekere stabiliteit en maturiteit in werking heeft. Immers, het gebeurt vaak dat vroeg in de uitvoeringsfase bepaalde aspecten van de implementatie nog bijgestuurd worden om het beleid en de uitvoering ervan verder te verfijnen. Wanneer men op dat moment reeds een experimentele evaluatie opzet en er veranderen tijdens de evaluatie nog tal van

factoren, komt dat de bruikbaarheid van de evaluatieresultaten niet ten goede. Veranderingen tijdens de uitvoeringsfase kunnen de facto leiden tot verschillende varianten van het beleid. Daardoor is het achteraf niet meer duidelijk welke 'versie' van het beleid welke effecten voortbracht. Beleidsevaluaties met een experimenteel design worden dus best gereserveerd voor ex post evaluaties van een beleid dat tamelijk stabiel is en waarvan de implementatie tijdens de duur van de evaluatie geen ingrijpende veranderingen ondergaat.

Ten tweede kan een experimenteel evaluatiedesign als inadequaaf beoordeeld worden vanuit ethische overwegingen. In het bijzonder wanneer het om personen gaat, kan de procedure van toevallige toewijzing aan de beleids- en controlegroep door stakeholders gezien worden als arbitrair en als het onterecht niet toekennen van voordelen aan personen in de controlegroep. In deze zienswijze staat men wel open voor het experimenteren met beleid wanneer dit klaarblijkelijk positieve effecten kan meebrengen. Maar men beschouwt het als schadelijk wanneer men potentiële goede dienstverlening niet openstelt voor zij die er nood aan hebben. Dit wordt als onethisch beschouwd. De keerzijde van dit dilemma kan als volgt worden geargumenteed: meestal is het niet op voorhand geweten of een beleid effectief is en daarin ligt precies de reden voor een experiment. Vanuit dit opzicht worden de personen van de controlegroep niet weerhouden van een tussenkomst waarvan de gunstige invloed vaststaat, maar worden ze misschien wel gespaard van een tijdsverlies door een beleid dat niet effectief is.

In bepaalde gevallen zullen stakeholders afkerig staan tegenover de procedure van toevallige toewijzing omdat er een kans bestaat dat bepaalde personen nadeel zullen hebben. Neem bijvoorbeeld een evaluatie van de werking van de financiële bijstand die de overheid voorziet voor gezinnen van een niet nader gedefinieerde inkomenscategorie. Voor de duur van de evaluatie kan de onderhoudskost van gezinnen oplopen afhankelijk van het 'lot' of ze tot de controlegroep behoren of niet. De evaluator kan wel voorzien in een vergoeding voor een eventuele meerkost maar dit gegeven zelf kan het gedrag van deze personen en dus de resultaten van de evaluatie beïnvloeden.

De sterkste ethische dilemma's betreffen doorgaans de omstandigheden van de controlegroep. Wanneer van conventionele dienstverlening bewezen is dat ze effectief is, zou het onethisch zijn om deze dienstverlening te onthouden met het oog op het uittesten van alternatieven. Zo is het ondenkbaar om kinderen geen wiskundelessen te geven omdat ze tot een controlegroep worden toegewezen ter vergelijking met een andere groep kinderen die een nieuw wiskundecurriculum volgen. In dergelijke gevallen, is de hamvraag echter niet of het nieuwe curriculum beter is dan het bestaande, maar wel of het beter is dan het bestaande. In deze optiek is dan ook een vergelijking tussen een groep die het nieuwe curriculum krijgt en een controlegroep die het klassieke curriculum volgt de aangewezen experimentele setting.

Ethische bezwaren kunnen ook ontstaan door de procedure van toevallige toewijzing te hanteren wanneer de middelen schaars zijn en de vraag naar dienstverlening niet volledig kan beantwoord worden. Slechts een deel van de in aanmerking komende personen kan genieten van de dienstverlening. In dergelijk geval kan echter ook beargumenteerd worden dat de toevallige toewijzing de meest billijke methode is om hierin te beslissen vermits alle personen een gelijke kans krijgen. Wanneer de onzekerheid over de effectiviteit van het beleid groot is, kan dit inderdaad een oplossing bieden. Echter, wanneer beleidsverantwoordelijken bij voorbaat overtuigd zijn van de effectiviteit van het beleid, ondanks het gebrek aan empirische bevindingen die dit staven, kunnen zij sterk gekant zijn tegen een aselechte keuze. Ze kunnen dan erop aandringen dat de meest behoeftige personen prioriteit krijgen bij de dienstverlening.

Een derde beperking op het gebruik van een experimentele opzet met aselechte toewijzing is dat de uitvoering van het beleid of de dienstverlening tijdens de evaluatie op kritieke punten kan afwijken van de dagdagelijkse praktijk. Des lager het routinegehalte in de beleidsuitvoering, des te groter de kans op afwijkingen. Wanneer het gewicht van de afhankelijke beleidsbeslissing het toelaat, kan een aanpak voor dit probleem erin bestaan om twee rondes in de evaluatie te voorzien. Waarbij de dienstverlening eerst gebeurt als onderdeel van het evaluatieonderzoek en daarna in de dagdagelijkse setting. Op deze wijze kunnen bepaalde administratieve procedures geëvalueerd getoetst worden in het licht van het criterium van effectiviteit.

Een vierde beperking betreft de middelen die rechtstreeks aangewend kunnen worden voor de beleidsevaluatie. Hoewel het experimenteel design de sterkste garanties kan bieden op het vlak van de geldigheid van de resultaten, heeft deze aanpak een hogere kostprijs en vraagt ze meer tijd in vergelijking met alternatieve methoden. Zoals we eerder al stelden: wanneer het beleid dat geëvalueerd wordt bij voorbaat een minieme kans heeft om voortgezet te worden of weinig aandacht van stakeholders krijgt, is het maar de vraag of dit evaluatiedesign de investering waard is.

Een laatste beperking die we hier willen vermelden, vormt de kans dat het experiment tijdens de uitvoering zelf vertekend wordt doordat er relevante verschillen optreden tussen de beleidsgroep en de controlegroep. Dit kan leiden tot een mogelijke bias of vertekening van de inschatting van beleidseffecten.

5.5.5 Quasi-experimenteel design

5.5.5.1 Evaluatiekader: “policy on” vs. “policy off”

Uit de voorgaande paragraaf is af te leiden dat het experimenteel evaluatiedesign het meest solide is om de mate van beleidseffectiviteit te bepalen. De praktische en/of ethische bezwaren zullen echter in vele gevallen de evaluator noodzaken om een andere strategie te volgen. Deze alternatieve strategie wordt ook evaluatie met een quasi-experimenteel design genoemd. Het principe voor het evaluatiekader blijft hetzelfde: het vergelijken van een policy on met een policy off situatie.

Het voornaamste verschilpunt ligt echter in de constructie van de vergelijkingsbasis. In quasi-experimenteel design wordt de vergelijkingsbasis niet op toevalsbasis of ‘at random’ opgesteld.

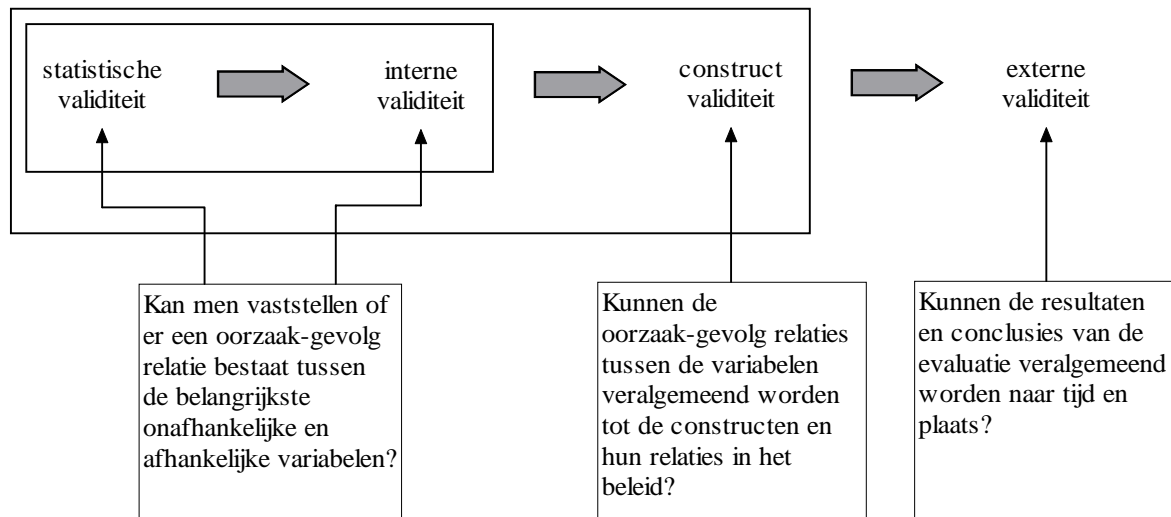
Quasi-experimentele designs zijn er echter in meerdere soorten naargelang de wijze waarop de vergelijkingsbasis dan wel wordt geconstrueerd.

5.5.5.2 Het bias probleem

Gemeenschappelijk voor quasi-experimentele designs is echter de problematiek van bias of vertekening. Immers, omdat de vergelijkingsbasis niet wordt opgebouwd aan de hand van at random samenstelling van een beleids- en controlegroep, komt de equivalentie van de policy-on en policy-off situatie tussen de momenten van de voor- en nametingen in het gedrang. Equivalentie is nochtans essentieel om het beleid te kunnen afzonderen van alternatieve verklaringsgronden voor de geobserveerde effecten of veranderingen.

Beperkingen op het vlak van equivalentie leiden tot een bias of vertekening in de schatting van de mate van beleidseffectiviteit. Dergelijke bias kan op twee manieren zich manifesteren: het beleidseffect - dat deel van het effect dat effectief toegeschreven kan worden aan het beleid- wordt systematisch ofwel onderschat ofwel overschat. Bovendien kan vooraf de richting van de afwijking niet altijd achterhaald worden. In het ergste geval is de conclusie over beleidseffectiviteit totaal verkeerd.

Deze bias vormt met andere woorden een bedreiging voor de geldigheid of validiteit van het evaluatiedesign. Wanneer men het ganse evaluatieproces bekijkt, kan men verschillende soorten van validiteit onderkennen. Ze zijn ook onderling verbonden, zoals weergegeven in onderstaand figuur.



Figuur 39: Types van validiteit doorheen het evaluatieproces

In deel 2 van de handleiding wordt dieper ingegaan op deze types van validiteit en de factoren die de geldigheid in het gedrang kunnen brengen. Het bias probleem waarvan hier sprake in de context van een quasi-experimenteel evaluatiedesign, vormt in het bijzonder een bedreiging voor de interne validiteit. Ze heeft betrekking op het vaststellen van causale relaties tussen onafhankelijke variabelen (het beleid) en afhankelijke variabelen (effecten).

Vanwaar komt dergelijke bias? Ze ontstaat door een proces waarvan de invloeden niet allemaal of volledig gekend zijn. Er zijn verschillende oorzaken denkbaar. Een bias kan ten eerste ontstaan door de procedure zelf waarmee de analyse-eenheden (personen, locaties, organisaties, ...) aan groepen worden toegewezen. Maar eenheden in beide groepen kunnen ook verschillen vertonen op kenmerken die men vooraf niet in acht neemt maar die relevant zijn omdat ze de effecten beïnvloeden, los van al dan niet toepassing van het beleid. We spreken dan van een *selectiebias*. Enkele voorbeelden maken dit duidelijk.

Stel dat men wil nagaan of een nieuwe aanpak van dienstverlening vruchten afwerpt. Om de beleids- en controlegroep samen te stellen zou men kunnen rekening houden het feit of personen uit de doelgroep vrijwilliger zijn om mee te werken. Door zich als vrijwilliger op te geven hebben de personen zelf bepaald in welke groep ze terecht komen. De selectiebias is dan te omschrijven als elk verschil tussen vrijwilligers en niet-vrijwilligers dat zou blijken bij een effectmeting wanneer geen van beide groepen de dienstverlening volgens de nieuwe aanpak zou ontvangen. Omdat we onvoldoende kennis hebben van alle relevante verschillen tussen vrijwilligers en niet-vrijwilligers, zijn we in mogelijkheden beperkt om de aard en draagwijdte van de vertekening te bepalen.

Een selectiebias ontstaat echter vaak op een meer subtiele wijze, zonder dat we op een doelbewuste wijze analyse-eenheden selecteren waarbij het beleid wordt toegepast of niet. Een evaluatie van het drugspreventiebeleid in scholen kan een school met een innovatieve aanpak vergelijken met een andere school waar deze aanpak niet wordt gevolgd maar die voor de rest qua profiel vergelijkbaar is. De kinderen uit deze tweede school kan men dus selecteren als de controlegroep. Op het einde van het schooljaar kan men het druggebruik vergelijken tussen de twee leerlingengroepen. Laten we zelfs aannemen dat de mate van druggebruik hetzelfde is in beide scholen bij het begin van het schooljaar. Hoe kan de evaluator met zekerheid stellen dat wanneer in geen van beide scholen een preventieprogramma zou gelopen hebben, de resultaten qua druggebruik op het einde van het schooljaar dezelfde zouden zijn in de twee scholen? Er spelen immers heel wat persoonlijke, culturele en economische factoren mee in de schoolkeuze. Deze factoren bepalen mee of kinderen in de ene of de andere school terecht komen als leerling. Een deel van deze factoren kan eveneens een rol spelen voor de mate waarin kinderen in contact komen met drugs tijdens het schooljaar. Naarmate dit inderdaad zo is, zal er een selectiebias of vertekening zitten in elke beoordeling van het effect van het preventiebeleid gebaseerd op een onderlinge vergelijking van de twee scholen.

Er zijn nog *andere mogelijke oorzaken* van een bias in de beoordeling van beleidseffectiviteit, los van de procedure waarmee men de beleids- en controlegroep samenstelt. In deel 2 van de handleiding wordt ingegaan op alle mogelijke bedreigingen voor de interne validiteit van een evaluatiedesign, geïllustreerd met concrete voorbeelden. We vermelden hier slechts enkele oorzaken, in het bijzonder deze waarvoor een op maat aangepast quasi-experimenteel design rechtstreeks remedies kan aanreiken (cf. infra).

Analyse-eenheden kunnen tijdens de evaluatie en dus tussen de voor- en nameting op de effectvariabele wegvallen uit hun groep: ofwel verdwijnen ze letterlijk uit de groep, ofwel is het niet meer mogelijk om een nameting uit te voeren. Dit verschijnsel heet drop-out en kan zowel in de beleids- als controlegroep optreden. Het probleem van drop-out is dat wegvallende eenheden uit beide groepen meestal zullen verschillen op factoren die - naast de aan- of afwezigheid van het beleid- eveneens het effect kunnen beïnvloeden. Daardoor zal de vergelijkbaarheid van de overgebleven eenheden in de twee groepen ook veranderd zijn. De gevolgen zijn dezelfde als bij het optreden van een selectiebias.

Bij twee andere oorzaken is het temporele aspect van belang: algemene trends en uitzonderlijke gebeurtenissen. Algemene trends kunnen het reële beleidseffect versterken of afzwakken. Lange termijn trends of autonome ontwikkelingen kunnen ertoe leiden dat een bepaald effect was opgetreden ook wanneer het beleid niet was uitgevoerd. Dat effect kan in dezelfde richting of in tegengestelde richting werken als het beleidseffect. Wanneer men bestaande trends niet identificeert of onderkent, komt men tot een over- of onderschatting van de effectiviteit van het beleid. Plotse, uitzonderlijke gebeurtenissen kunnen interfereren met de uitvoering van het beleid en bijgevolg ook tot een vertekening leiden in de beraming van een beleidseffect.

Samengevat, een bias kan optreden op het moment dat de analyse-eenheden worden toegewezen aan de beleids- en controlegroep, maar ook daarna. De problematiek komt erop neer dat er een verschil ontstaat tussen analyse-eenheden, waarmee men de policy-on en policy-off situaties vergelijkt, op relevante kenmerken die hun waarde op de effectvariabele beïnvloeden naast het feit of ze behoren tot de beleids- of controlegroep.

Het moet de lezer duidelijk geworden zijn dat een bias of vertekening belangrijke gevolgen kan hebben voor beslissingen die op basis van evaluatieresultaten en -conclusies genomen worden. Bijzonder aandachtspunt voor de evaluator is dan ook om het potentieel voor dergelijke bias te minimaliseren van bij de uitwerking van het quasi-experimenteel evaluatiedesign. In de volgende sectie bespreken een aantal werkwijzen. Ze hebben in de praktijk hun deugdelijkheid bewezen, maar zullen nooit de garanties van een experimenteel design kunnen voorleggen.

5.5.5.3 Vergelijkingsbases

Hoe kan de evaluator het gevaar op een bias minimaliseren in een quasi-experimenteel design? Er bestaan verschillende mogelijkheden die we hieronder overlopen. Het gaat om verschillende werkwijzen om een adequate vergelijkingsbasis te creëren.

Net als met het strategische verschil tussen een experimenteel en quasi-experimenteel design zal de evaluator niet altijd vrij kunnen kiezen tussen de opties voor een quasi-experimenteel design. Anderzijds biedt elk van de onderstaande opties wel een oplossing op maat om een van de vermelde oorzaken van een bias te vermijden of te remediëren.

- *Controlegroep samenstellen via matching*

Eén manier om een controlegroep samen te stellen, wanneer de 'at random' procedure onmogelijk is, steunt op het principe van matching. Deze werkwijze is gericht op het vermijden van een selectiebias.

Hierbij wordt doorgaans eerst de beleidsgroep gespecificeerd waarna de evaluator een controlegroep samenstelt. Daarbij wordt een zo sterk mogelijk 'match' met de beleidsgroep nagestreefd. Men selecteert eenheden waarop het beleid niet van toepassing zal zijn die overeenkomen met de eenheden in de beleidsgroep op een set van geselecteerde kenmerken.

Elk kenmerk waardoor de twee groepen zouden verschillen op de effectvariabele wanneer bij geen van beide het beleid zou worden uitgevoerd is daardoor relevant. In de mate waarin geen equivalentie tussen beide groepen bereikt wordt door matching op kenmerken die - naast het beleid zelf- het effect beïnvloeden, zal de vaststelling van het beleidseffect vertekend worden.

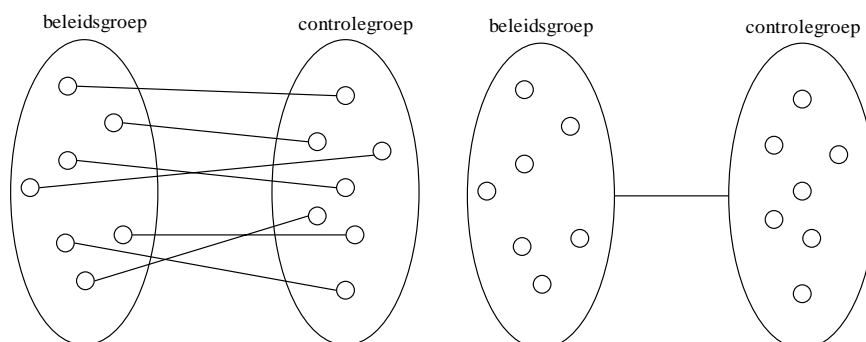
Bij matching zijn volgende twee elementen van belang:

- de keuze van de kenmerken of variabelen waarvoor matching moet gebeuren
- de matching procedure

De *keuze van de relevante kenmerken* dient de evaluator te maken op basis van voorkennis of theoretische inzichten in de sociale mechanismen die spelen. Een accurate beleidstheorie speelt hier dus weer een belangrijke rol (cf. supra). Bijzondere aandacht moet uitgaan naar variabelen die potentieel verband houden met het selectieproces dat eenheden toekent aan de beleids- of controlegroep. Wanneer matching voor deze variabelen onmogelijk is, dient men ze toch te identificeren en te meten. Dit laat toe om ze mee te nemen in de data-analyse om een mogelijk overblijvende selectiebias te verkennen, en eventueel statistisch te controleren (cf. infra).

Het is echter niet altijd noodzakelijk om de groepen te matchen op elk kenmerk dat de literatuur of expertise als relevant aanduidt; vaak zullen sommige variabelen correleren. Zo hangen de scores van leerlingen op intelligentietesten en hun schoolcijfers doorgaans nauw met elkaar samen. Het volstaat dan om één van de twee indicatoren en de bijhorende dataset te gebruiken wanneer de intellectuele capaciteit een relevant kenmerk is.

De *procedure van matching* kan op individueel of op geaggregeerd niveau gebeuren. Bij geaggregeerde matching gebeurt de overeenstemming van de eenheden niet case per case, maar wordt de verdeling binnen beide groepen vergelijkbaar gemaakt voor elke weerhouden relevante variabele. Individuele matching is verkieslijk wanneer een breed gamma van kenmerken voor matching worden gebruikt. Het nadeel is de grotere tijdsinvestering. Bovendien is individuele matching moeilijker te realiseren naarmate het aantal kenmerken waarop de match moet gebeuren stijgt.



Figuur 40: Individuele of geaggregeerde matching

Onbetrouwbare data kunnen ten grondslag liggen aan het optreden van kunstmatige statistische fenomenen zoals regressie naar het gemiddelde. Stel dat eenheden van de beleidsgroep en controlegroep zich in verschillende uiteinden van de verdeling bevinden voor een relevante variabele door onbetrouwbare, inconsistente meetprocedures. In dergelijk geval zal matching vooral mogelijk zijn tussen die eenheden uit beide groepen die qua 'score' het dichtst bij elkaar aanleunen en die zich dus rond het algemene gemiddelde bevinden. Bijgevolg kan de vergelijking van de groepen op basis van de metingen echter pseudo-verschillen opleveren die niet gerelateerd zijn aan beleidseffecten.

Hoe adequaat matching ook wordt toegepast, er kunnen steeds kritieke verschillen overblijven tussen de beleids- en controlegroep. Het is daarom raadzaam om dergelijke variabelen toch te kennen en te meten, waardoor ze statistisch gecontroleerd kunnen worden zelfs indien er geen matching mogelijk is. Statistische controles (cf. infra) hebben doorheen de tijd de techniek van matching in grote mate deels vervangen en aangevuld. Toch blijft de techniek van matching op een aantal variabelen erg aangewezen. Dit is in het bijzonder het geval wanneer wordt gewerkt met kleine geaggregeerde groepen en wanneer de te evalueren eenheden (personen, plaatsen, organisaties, ...) bijzondere kenmerken vertonen.

- *Gelijkstellen van beleids- en controlegroep via statistische procedures*

In de praktijk zal de controlegroep vaak noodgedwongen samengesteld worden op basis van contextgebonden relevante factoren en pragmatische overwegingen. De kans op een bias is dan niet weg te denken. Het is van belang dat men de relevante verschillen tussen de groepen kent en kan meten. Men kan met behulp van statistische procedures voor deze verschillen controleren die anders tot vertekende schattingen en uitspraken over beleidseffectiviteit zouden leiden. Ook wanneer er verschillen ontstaan tijdens de evaluatie, bijvoorbeeld door drop-out (cf. supra), kan men trachten om deze statistisch te beheersen.

Men kan univariate analyses uitvoeren waarbij om de beurt voor één relevant kenmerk statistisch gecontroleerd wordt. Vaak zal men echter een multivariate analysemethode hanteren: men controleert daarbij voor meerdere kenmerken tegelijk. Het betreft zogenaamde regressieanalyse. In deel 3 van de handleiding wordt verder ingegaan op de werkwijze.

Het doel van deze statistische gelijkenschakeling is om initiële verschillen tussen de beleids- en controlegroep in rekening te brengen. Het meten en vergelijken van effecten gebeurt op dezelfde wijze als bij een experimenteel design (cf. supra).

	Effectmetingen		
	Voor	Na	Vershil
Beleidsgroep	B1	B2	$B = B2 - B1$
Controlegroep	C1	C2	$C = C2 - C1$

B en C zijn ook hier de effectveranderingen voor respectievelijk de beleids- en controlegroep. De statistische controle is erop gericht om het verschil in effectverandering (B - C) aan te passen door de proportie die toe te schrijven is aan de initiële verschillen in mindering te brengen. Het (eventueel) overblijvende verschil in effectverandering (B - C) is dan gelijk aan het reële beleidseffect of de beleidseffectiviteit.

De controlevariabelen in dergelijke statistische analyse hebben betrekking op:

- initiële kenmerken van de eenheden die gerelateerd zijn aan de effectvariabele

en/of

- kenmerken van de eenheden die rechtstreeks verband houden met de selectiebias.

Het geheel van (veronderstelde of vastgestelde) relaties tussen controlevariabelen en effectvariabelen kan worden vertaald naar een statistisch model. Voor de modellering van beide soorten controlevariabelen verwijzen we de lezer door naar meer gespecialiseerde literatuur over statistiek.

Soms zal de evaluator voorafgaand zekerheid hebben over welke variabel(e) men hanteerde voor de selectie van eenheden. In de lijn daarvan gebeurt de selectie van eenheden in de beleids- en controlegroep soms aan de hand van een drempelwaarde ('cutting point') op een bepaalde variabele. Eenheden met een score boven de drempelwaarde wijst men toe aan de beleidsgroep, eenheden met een lagere score komen terecht in de controlegroep. Zo richt men bijvoorbeeld het beleid op personen met de hoogste noden, op de meest ongevalgevoelige locaties, enz. In dergelijk geval kan de selectiebias ook statistisch gecontroleerd worden. In de praktijk worden bij beleidsinitiatieven echter niet altijd duidelijk geëxpliciteerde regels voorzien over de toepassingsvoorwaarden, d.w.z. bepalingen over welke eenheden in aanmerking komen en welke niet.

- *Reflexieve controles*

Een derde werkwijze in het kader van een quasi-experimenteel evaluatiedesign verschilt van de voorgaande doordat het zonder apart controlegroep werkt. De bepaling van beleidseffectiviteit wordt gebaseerd op informatie over de eenheden waarop het beleid werd toegepast. De beleidsgroep wordt daarvoor op zichzelf 'gereflecteerd' in de tijd, vandaar de benaming reflexieve controle. Een aantal auteurs spreekt echter onterecht van een 'non-experimenteel' design. Deze werkwijze omvat immers ook een beleidsinterventie waarbij men nagaat of er veranderingen optreden en meer bepaald ten gevolge van de beleidsactie. Om nuttige informatie over veranderingen of effecten te

verkrijgen, gebeuren de metingen op twee of meerdere tijdstippen, waarvan minstens één voorafgaandelijk aan de beleidsuitvoering.

De assumptie die men hier maakt is dat eenheden niet veranderd zijn met betrekking tot de effectvariabele in de periode tussen de verschillende metingen. Het verschil tussen de waarden op de effectvariabele voor en na de beleidstoepassing beschouwt men vervolgens als beleidseffect.

De eenvoudigste werkwijze is deze met één voor- en nameting, ook wel een *pre-post design* genoemd. Dit wordt hieronder weergegeven (O = meting, X = beleidsuitvoering).



Het nadeel hiervan is echter het reële gevaar voor een vertekening of bias van de schatting van het beleidseffect doordat in tussentijd andere factoren de effectvariabele kunnen beïnvloeden. Hoe langer de periode tussen de voor- en nameting, des te groter het gevaar op vertekeningen.

Een fenomeen waar men eveneens beducht voor moet zijn, wordt in de literatuur aangeduid met statistische regressie. Effectwaarden kunnen schommelen doorheen de tijd. Op lange termijn zullen ze echter tenderen naar het eigen gemiddelde. Metingen kunnen echter waarden opleveren die, wanneer ze met het lange termijn gemiddelde zouden worden vergeleken, als extreme (hoge of lage) waarden te bestempelen zijn. Omdat de waarden over tenderen of 'teruggaan' naar het statistische gemiddelde, zullen extreme waarden afgelost worden met meer 'normale' waarden. Alleen weet men niet steeds of men met extreme of normale waarden te maken heeft: men kent het gemiddelde niet en/of heeft nog data over een langer tijdssegment. Daarin ligt de kern van het probleem bij het bovenstaande pre-post design. De eerste meting kan een extreme waarde opleveren, waardoor er veel kans is dat de tweede meting een waarde oplevert die aanzienlijk verschilt. In bepaalde gevallen zal men dan ook onterecht besluiten dat de verandering in waarden geheel aan de beleidsuitvoering is toe te schrijven, terwijl het de facto gaat om statistische regressie. Een beleidsinterventie kan natuurlijk voor een trendbreuk zorgen bij de effectvariabele, waardoor deze na het beleid systematisch hogere of lagere waarden (naargelang de context) zal laten optekenen. Alleen, een simpel pre-post design zal daarover geen uitsluitsel kunnen brengen.

Dit pre-post design is dan ook te beperkt voor het bepalen van beleidseffectiviteit. Enkele auteurs spreken van een naïeve werkwijze. Deze aanpak is inderdaad geschikt voor monitoring van effecten, weliswaar zonder de aspiratie om uitspraken te doen over een beleidseffect, d.w.z. over de mate waarin de geobserveerde effectverandering toe te schrijven is aan het beleid.

Het pre-post design kan echter versterkt worden wanneer men veelvuldige metingen van de effectvariabele kan doen die de tijdsperiodes vóór en na de beleidstoepassing overspannen. Het werken met *tijdreeksen* is de sterkste vorm van reflexieve controle.



Tijdreeksen kunnen via monitoring opgebouwd worden. Ze maken het mogelijk om trends in kaart te brengen. Daardoor kan men de evolutie van de effectvariabele beter interpreteren. Eenmalige uitschieters in de effectwaarden kunnen immers duiden op de invloed externe ad hoc gebeurtenissen. Mogelijk was een trend al ingezet vooraleer het beleid werd uitgevoerd. Dergelijke informatie is belangrijk om te kunnen inschatten wat het effect was geweest wanneer er geen uitvoering van het beleid was geweest. Dergelijke factoren (trends of ad hoc gebeurtenissen) zouden een beoordeling van een beleidseffect gebaseerd op het simpele pre-post design kunnen vertekenen. Tijdreeksen verschaffen meer informatie over wat er gebeurd was zonder beleidstoepassing.

Het inzicht in de effectveranderingen en de accuraatheid van de beoordeling van beleidseffecten kan nog verbeterd worden. Zo kan men ten eerste ook externe factoren, waar men van weet of vermoedt dat ze het effect kunnen beïnvloeden, in de tijd opvolgen. Fluctuaties in de waarden van de effectvariabele kunnen dan afgezet worden tegen de evolutie van deze externe factoren. In het kader van ongevallenregistraties kan men bijvoorbeeld ook het weertype monitoren.

Wanneer men beschikt over relatief lange tijdreeksen van de periode vóór de beleidstoepassing, kan men lange termijn trends en eventuele seizoensgebonden variaties modelleren. Hiermee kan men de trend projecteren naar de termijn tijdens en na de beleidstoepassing. Vervolgens kan men nagaan of de geobserveerde effecten na de beleidstoepassing significant afwijken van deze projectie.

Ten tweede kan men naast de vergelijking tussen de situaties voor en na de uitvoering van het beleid, nog extra vergelijkingspunten inbouwen.

Zo kan men effectgegevens uitsplitsen aan de hand van bepaalde kenmerken van de eenheden. Vaak geeft dit een gedetailleerder inzicht in de effectverandering. Het kan belangrijk zijn om bepaalde structurele verschillen te ontdekken waardoor men de conclusies over beleidseffecten kan nuanceren en bijgevolg nauwkeuriger kan formuleren. Zo kan men bijvoorbeeld personen opdelen naar leeftijdscategorie of inkomen, gemeenten naar hun verstedelijkingsgraad, organisaties naar het aantal werknemers of de economische sector waartoe ze behoren, enz.

Daarnaast kan men comparatieve tijdreeksen ontwikkelen door meerdere vergelijkbare cases of analyse-eenheden in de tijd op te volgen. Belangrijk is dat het hier in de praktijk dan meestal gaat om een vergelijkbare groep dan wel een controlegroep in de strikte betekenis van het woord. Dit heeft alles te maken met de mate waarin men vat heeft op de vergelijkingsgroep. Bij een experimenteel design of bij bovenstaand quasi-experimenteel design probeert men via een systematische procedure een controlegroep samen te stellen (at random of matching). Wanneer dit adequaat gebeurt, heeft men genoeg aan een eenmalige voor- en nameting bij de twee groepen om relevante verschillen (policy on vs. policy off) te achterhalen. In bepaalde gevallen zal men wel een aantal criteria kunnen hanteren om vergelijkbare eenheden te selecteren, maar men zal weinig vat hebben op de vergelijkingsgroep tussen de twee metingen.

Zo zou men Vlaanderen bijvoorbeeld kunnen vergelijken met andere Europese regio's om de luchtkwaliteit op te volgen. Daarbij dient men aandacht te schenken aan het zoeken naar vergelijkbare regio's aan de hand van een aantal criteria. In het voorbeeld kan dit het bruto regionale product zijn, de aanwezigheid van bepaalde chemische verwerkingsnijverheid, verkeersintensiteiten enz. Het milieubeleid in Vlaanderen zal verschillen van dat in de vergelijkbare regio's. Met andere woorden: in de vergelijkingsgroep heeft men geen policy-off situatie, maar een afwijkende policy-on situatie. Ook in andere regio's worden milieumaatregelen getroffen.

Aan de hand van comparatieve tijdreeksen kan men de effectveranderingen bij de invoering van nieuwe maatregelen in de beleidsgroep aftoetsen tegen de trends in de vergelijkingsgroep. Maar omdat de vergelijkbaarheid steeds belangrijke beperkingen zal hebben, kan men nooit met dezelfde graad van zekerheid uitspraken doen over het beleidseffect als ingeval van een at random of via matching samengestelde controlegroep. Meerdere metingen voor en na de beleidsingreep zullen echter wel betere indicaties opleveren.

Zo kan men de luchtkwaliteit van Vlaanderen monitoren tegen het licht van een bepaalde beleidsmaatregel die wordt ingevoerd. De evolutie van de Vlaamse luchtkwaliteit kan men vergelijken met de trends van vergelijkbare regio's. De conclusies gebaseerd op de vergelijking zullen niet met zekerheid de effectiviteit van het Vlaamse beleid kunnen beoordelen. Maar de comparatieve gegevens over de vergelijkbare regio's kan men wel afwegen tegen mogelijke andere factoren die de trend van de effectvariabele voor Vlaanderen kunnen verklaren, bijvoorbeeld het fenomeen van statistische regressie (cf. supra). Schematisch kan de werkwijze van comparatieve tijdreeksen als volgt worden voorgesteld.

Beleidsgroep	○	○	○	X	○	○	○
Vergelijkingsgroep	○	○	○		○	○	○

Tabel 13: Comparatieve tijdreeksen voor een beleids- en vergelijkingsgroep

Tijdreeksen bevatten vaak niet enkele de periode voor en na de beleidsuitvoering. Vaak zal men ook monitoren tijdens de duurtijd van het beleid. Een beleidsmaatregel blijft bovendien vaak een hele tijd van kracht. Het is dan interessant om informatie te hebben over de duurzaamheid van geobserveerde effecten of veranderingen. Ook hier men met of zonder een vergelijkingsgroep werken. De X staat hier voor één en hetzelfde beleidsinitiatief dat eenmalig ingevoerd wordt.

Beleidsgroep	○	○	○	X	○	X	○	X	○	X	○	X	○
(Vergelijkingsgroep)	○	○	○	○	○	○	○	○	○	○	○	○	○

Tabel 14: Tijdreeksen opbouwen tijdens de uitvoering van het beleid

Tijdreeksen hebben ook nut in het kader van meer complexe beleidsinitiatieven. Wanneer men maatregelenpakket uitwerkt in functie van eenzelfde doelstelling, is het vaak moeilijk, zonet onmogelijk, om het effect van elke maatregel apart te gaan bepalen. Een deel van de oplossing is om de verschillende maatregelen gefaseerd in te voeren. In combinatie met monitoring kan men de evolutie vóór en na elke maatregel in kaart brengen, zoals hieronder schematisch weergegeven: X_1 duidt op de introductie van de eerste maatregel, X_2 op de invoering van een tweede maatregel.

Beleidsgroep	○	○	○	X_1	○	○	X_2	○	○	○
--------------	---	---	---	-------	---	---	-------	---	---	---

Tabel 15: Comparatieve tijdreeksen bij complex beleid

Tijdreeksen lenen zich tot bevattelijke grafische voorstellingen van evoluties en trends.

Ze kunnen de aandacht richten op belangrijke schommelingen in waarden op effectvariabelen, maar ook voor eventuele controlevariabelen die men opvolgt. Het is echter oplettend om te snel conclusies over beleidseffecten te formuleren enkel gebaseerd op een snelle screening van tijdreeksen. Het is raadzaam om steeds een statistische analyse toe te passen om dergelijke 'evidente' veranderingen te staven. Pas dan kan men een beoordeling maken met een gekende graad van zekerheid.

5.5.5.4 Randvoorwaarden en beperkingen

De quasi-experimentele methode kan men beschouwen als de 'second-best' optie om beleidseffectiviteit te bepalen. De meerwaarde van dergelijk design, indien accuraat toegepast, is dat het uitzicht geeft op

betrouwbare en valide schattingen van beleidseffecten terwijl er een grotere mate van flexibiliteit is om aan te passen aan de beleidscontext die geëvalueerd wordt en die niet steeds compatibel is met de randvoorwaarden voor een experimenteel design.

Het kritieke punt blijft het vermijden of controleren van de bias of vertekening van de resultaten. De evaluator dient voor elke evaluatie apart de nodige aandacht te besteden aan de beleidscontext, de relevante variabelen om op een accurate wijze de beleidseffectiviteit te bepalen. Tevens dient men de stakeholders te informeren over mogelijke beperkingen of assumpties met betrekking tot de evaluatieresultaten.

5.5.6 Alternatief of kwalitatief design

Het experimenteel en quasi-experimenteel design zijn benaderingen om (de mate van) beleidseffectiviteit vast te stellen. Soms zal het echter niet mogelijk zijn om (geheel of gedeeltelijk) de invloed van het beleid af te zonderen van alternatieve verklaringen voor de vastgestelde effectveranderingen. Of men heeft de beleidseffectiviteit wel kunnen vaststellen maar men wenst ook na te gaan *waarom* beleid al dan niet geslaagd is.

In beide gevallen heeft men nood aan inzicht in de werking van het beleid en andere verklarende factoren. Daarvoor dient men kwalitatieve technieken aan te wenden, zoals case studie, diepte-interviews met beleidsuitvoerders, focusgroepen etc.

Literatuur

Afdeling Beleidsanalyse, Nederlands Ministerie van Financiën (1989). *Heeft beleid effect? Enkele benaderingen voor de evaluatie van bestaand beleid*. 's Gravenhage: SDU.

Bressers J.Th.A. en Hoogerwerf A. (1991). *Beleidsvaluatie*. Alphen aan den Rijn: Samsom H.D. Tjeenk Willink.

Chen H.-T. (2005). *Practical Program Evaluation*, Sage: Thousand Oaks.

Donaldson S.I. (2003). "Theory-Driven Evaluation in the New Millennium". In: Donaldson S.I. and Scriven M. *Evaluation Social Programs and Problems: Visions for the new millennium*. New Jersey: Lawrence Erlbaum Associates, Publishers.

Donaldson S.I. and Lipsey M.W. (2006). "Roles for theory in contemporary evaluation practice: Developing practical knowledge". In: Shaw I., Greene J. and Mark M. (Eds.) *Handbook of Evaluation*. Thousand Oaks (Ca): Sage.

European Commission (1999). *The Means Collection – Evaluating socio-economic programmes*. Luxembourg: Office for Official Publications of the European Communities.

McDavid J.C. & Hawthorn L.R.L. (2005). *Program Evaluation and Performance Measurement, an Introduction to Practice*, Thousand Oaks, CA: Sage.

Rossi P.H., Lipsey M.W. and Freeman H.E. (2004). *Evaluation – A Systematic Approach*. 7th edition. Thousand Oaks, CA: Sage.

Stern E. (ed.) (2005). *Evaluation Research Methods. – Volume III*. London: Sage.

Wholey J.S., Hatry H.P. and Newcomer K.E. (eds.) (2004). *Handbook of Practical Program Evaluation*. San Francisco, CA: Jossey-Bass.

Worthen B.R., Sanders J.R. and Fitzpatrick J.L. (2003). *Program Evaluation – Alternative Approaches and Practical Guidelines*. New York: Addison Wesley Longman.

6. Evaluatiemanagement en –capaciteit

6.1 Verantwoordelijkheden en budget

Belangrijk voor het ganse evaluatieproces is het uitklaren van verantwoordelijkheden van de verschillende betrokkenen. Dit geldt in de eerste plaats aan de vraagzijde.

De beslissing om te evalueren kan vervat liggen in een wettelijk of reglementair kader (wet, decreet, beheersovereenkomst). In alle andere gevallen is een evaluatie het voorwerp van een ad hoc beslissing vanwege een opdrachtgever. De opdrachtgever is doorgaans een politieke institutie (parlement, regering, minister) of het topmanagement van een administratie die verantwoordelijk is voor de planning en/of uitvoering van het beleid. Het kan ook een ad hoc instantie zijn; een interministerieel comité, regionale commissie. Ook het Rekenhof kan optreden als opdrachtgever. In het geval van interbestuurlijke samenwerkingsverbanden, zullen doorgaans de betrokken partners gezamenlijk de beslissing nemen.

Ingeval van een interne evaluatie is de opdrachtgever niet noodzakelijk dezelfde actor als de uitvoerder van het beleid. Een minister, lid van de bestendige deputatie, burgemeester of schepen kan bijvoorbeeld zijn of haar administratie opdragen om een evaluatie uit te voeren. Dit blijft een interne evaluatie omdat de administratie behoort tot de uitvoerende macht.

Ingeval van een externe evaluatie geldt dit vanzelfsprekend ook. Bovendien is de opdrachtgever niet noodzakelijk ook de reële uitbesteder en gesprekspartner voor de externe evaluator. Een minister kan zijn administratie de opdracht geven om een evaluatie te laten uitvoeren door externen. Vaak zal een cel of afdeling binnen de administratie dan de planning en opvolging van de evaluatie voor haar rekening nemen. Het is belangrijk dat een externe evaluator duidelijkheid heeft over wie de opdrachtgever is en wie de concrete gesprekspartner is voor het uitvoeren de evaluatie.

Zowel ingeval van een interne als een externe evaluatie is het beschikken over een indicatief budget belangrijk. Het budget kan zowel top-down als bottom-up bepaald worden, of door een combinatie van beide benaderingen.

Het budget vaststellen in verhouding tot het te evalueren beleidsinitiatief is kenmerkend voor de top-down methode. Het is moeilijk om een 'typisch' percentage aan te geven; dit hangt bijvoorbeeld samen met de aard van het programma, project of maatregelenpakket of de individuele maatregel. Er bestaan praktijkvoorbeelden waarbij de evaluatiekost 10% tot zelfs de helft bedraagt van de kost van de geëvalueerde maatregel. Dergelijke percentages zijn te verantwoorden wanneer het gaat om zeer innovatieve of experimentele beleidsacties. In dat geval is het echter relevanter om de evaluatiekost uit te

drukken in relatie tot de kost van de veralgemeende invoering van het beleidsprogramma waartoe men zal beslissen als het proefproject succesvol blijkt.

Bij de bottom-up methode zal men de verwachte kost van elke evaluatievraag schatten. Een alternatieve indeling die hierbij gebruikt kan worden, is deze tussen algemene kosten (structurering van de evaluatie, rapportering) en de kosten verbonden aan dataverzameling en -analyse. Deze laatste categorie kan uitgesplitst worden naar elk van de toegepaste evaluatietechnieken. In boekdeel 3 geven we per evaluatietechniek een relatieve aanduiding van de kostprijs. De tijd nodig voor het toepassen van een bepaalde techniek is niet altijd recht evenredig met de kostprijs. De meest solide techniek is niet noodzakelijk de duurste.

Wanneer een evaluatie wordt uitbesteed is het raadzaam om in het bestek de verplichting op te nemen dat offertes de kostprijs uitsplitsen naar kostensoorten, bijvoorbeeld personeel, werking en overhead. Een meer gedetailleerde uitsplitsing kan geënt worden op de fasen van het evaluatieproces: voorbereiding, dataverzameling, data-analyse en rapportering.

In de praktijk zal de kostprijs variëren naargelang het concrete geval. Het budget wordt vaak bepaald door het combineren van de vermelde top-down en bottom-up methoden. Volgende factoren zijn daarbij in acht te nemen:

- het aantal evaluatievragen;
- het aantal evaluatievragen dat een diepgaande analyse vergt;
- de verwachtingen met betrekking tot de aannemelijkheid van de bevindingen;
- de kwaliteit van bestaande monitoringsystemen en databanken;
- de tijdsbesteding aan opvolging en de mate waarin gegevens binnen de administratie moeten gegenereerd en/of verwerkt worden.

6.2 De status van de evaluator

De status van de evaluator is moet bij elke evaluatie een belangrijk aandachtspunt zijn, en meer bepaald de onafhankelijkheid. Onafhankelijkheid staat niet los van de geloofwaardigheid en de meerwaarde van een evaluatie. Dit geldt zowel bij interne als externe evaluaties. De omstandigheden waarin evaluatieteams opereren vertonen een grote variëteit. Als evaluatieteams zeer nauw samenwerken met de te evalueren instanties ontstaat, zoals in de literatuur omschreven, een 'kritische vriend'. Deze vorm is het meest voorkomend in ex ante evaluaties gericht op de verbetering van het beleid en management.

Vaak zal in dit geval de onafhankelijkheid al minder duidelijk afgelijnd zijn. Opdrachtgever en evaluator maken afspraken afhankelijk van de context. De onafhankelijkheid van de evaluator ten opzichte van het te evalueren beleid kan zowel vanuit een structureel als een mentaal perspectief worden benaderd.

Een totaal onafhankelijke evaluatie of waarde vrije evaluatie bestaat echter niet. Een evaluator is altijd onderworpen aan een hele resem invloeden. Een aantal factoren beïnvloeden de onafhankelijkheid van de evaluator. Sommige factoren zijn beheersbaar, andere zijn moeilijk te controleren. Bepaalde factoren spelen impliciet een rol, andere manifesteren zich expliciet. Hieronder vermelden we een aantal mogelijke invloeden waaraan een (interne of externe) evaluator of evaluatieteam onderworpen kan zijn.

- Eigen visies, waarden en normen van de evaluator

In de interpretatie van de analyseresultaten en de formulering van conclusies speelt de evaluator een belangrijke rol. De evaluator staat niet los van eigen ervaringen, opleiding en persoonlijke waarden en normen. Het is belangrijk om zich daarvan bewust te zijn. De data-analyse dient volgens objectieve procedures en technieken te verlopen. De geformuleerde conclusies moeten steeds beargumenteerd worden vanuit de resultaten van de analyse en hun toetsing met het referentiekader, d.w.z. de geëxpliciteerde evaluatiecriteria, normen en standaarden.

Evaluatoren kunnen een persoonlijke sympathie of affiniteit hebben voor het beleid of met de betrokken doelgroep. De inschakeling van interne de selectie van externe evaluatoren gebeurt vaak op basis van hun substantiële kennis van een bepaald beleidsdomein, hun contacten en hun evaluatie-ervaring. De rapportering dient dergelijke elementen duidelijk te vermelden.

- Bijsturingen en cours de route

Hoewel het belangrijk is om de evaluatiescope vooraf vast te leggen, is het niet ondenkbaar dat tussentijdse bevindingen consequenties hebben op het verdere verloop van het evaluatieonderzoek. In het extreme geval wordt het evaluatiemotief aangepast en verschuift de klemtoon bijvoorbeeld van de verbetering van het management en de uitvoering naar beleidsleren. De evaluator dient dan het ganse evaluatiedesign bij te sturen. Dit brengt vaak een trade-off tussen wenselijkheid en haalbaarheid met zich mee.

- Vertrouwen tussen opdrachtgever en evaluator

Naast de contractuele relatie die ook de financiering (ingeval van een externe evaluatie) regelt, dient er een wederzijds vertrouwen tussen opdrachtgever en evaluator te bestaan doorheen het ganse evaluatieproces. Goede afspraken over fasering in de aanpak en communicatie zijn belangrijke hefboomen om dit vertrouwen te bewerkstelligen. Ook ethische codes kunnen daartoe bijdragen. We komen verderop nog terug op het communicatieaspect.

- Politieke context

Evaluatie van beleid vindt nooit in een politiek vacuüm plaats. Tal van organisaties, zoals politieke partijen, drukkingsgroeperingen en anderen oefenen invloed uit op het beleid. Deze groepen hebben verschillende en soms tegenstrijdige belangen en prioriteiten. Beleidsverantwoordelijke staan onder druk om de meerwaarde van hun beleidskeuzes en uitgevoerd beleid aan te tonen.

Laatstgenoemd spanningsveld kan leiden tot politieke druk op het evaluatieteam. Politici worden geconfronteerd met een snel veranderende en complexe handelingscontext, gekenmerkt door machtsverhoudingen en concurrerende belangen. Dit vormt vaak de aanleiding tot pogingen om invloed uit te oefenen op het design en de uitvoering van een beleidsevaluatie. Dergelijke invloed zal vaak optreden binnen de normale onderhandelingsmarges. In de praktijk zijn er evenzeer situaties waarbij de politieke druk de aanvaardbare grenzen overschrijdt. Evaluatieteams kunnen anticiperen op dergelijke druk door vooraf de politieke belangen en invloeden in kaart te brengen. Ze kunnen zo rekening houden met het conflictgehalte bij de uitwerking van hun evaluatiedesign door bepaalde aanpak op het vlak van het betrekken van en feedback naar de verschillende stakeholders. We zetten een aantal adviezen voor de (interne of externe) evaluator op een rij.

- Probeer de gedachtegang van de betrokken politici te begrijpen en te achterhalen wat deze ze winnen of te verliezen hebben bij het evaluatieproces
- Wijs de opdrachtgever erop dat de interpretatie van de bevindingen leidt tot een aantal praktische aanbevelingen ter verbetering van het beleid en management.
- Probeer te duidelijkheid te scheppen over de evaluatiecriteria die de machtigste politieke stakeholders hanteren om het beleid te beoordelen. Door deze criteria mee te nemen in het design, kunnen hun vragen zeker een antwoord krijgen. Dit impliceert niet dat daarnaast nog andere criteria worden opgenomen in het evaluatiedesign.
- Voorzie een structuur of feedbackmechanisme waardoor de belangrijkste politici en stakeholders betrokken zijn bij de planning en opvolging van het evaluatieproces. Op die manier ligt de verantwoordelijkheid bij de opdrachtgevers om zelf een consensus te bereiken over zowel de inhoud als het resultaat van de evaluatie.
- Schrijf het evaluatierapport met aandacht voor de formulering van de bevindingen, zeker wat de verantwoordelijkheden voor het falen of het succes van een beleid betreft. Review zelf de laatste versie alvorens deze over te maken aan de opdrachtgevers. Elk van de bevindingen en conclusies dient beargumenteerd en gestaafd te zijn.

Bovenstaande adviezen hebben niet de aspiratie om de politieke invloed te doen verdwijnen, maar ze kunnen bijdragen bij het anticiperen op en kanaliseren van politiek druk. Het leren omgaan met deze mogelijke invloeden zal de kwaliteit en de aanvaardbaarheid van de evaluatieresultaten ten goede komen. Politieke invloed mag in geen geval de kwaliteit van de evaluatie hypothekeren.

6.3 Opvolging van de evaluatie

6.3.1 Communicatie tussen de evaluator en de stakeholders

Tal van redenen kunnen aangehaald worden waarom het management van een evaluatieproces voortdurende interacties vereist tussen alle betrokken partners. Goede communicatie is één van de kritieke succesfactoren in het evaluatieproces. Communicatie draagt bij tot:

- het maken van duidelijke afspraken over de finaliteit en inhoud van de evaluatie
- het opbouwen van vertrouwen tussen opdrachtgever en evaluator
- een accurate beeldvorming over het evaluandum
- de identificatie van relevante informatiebronnen
- de betrokkenheid van de stakeholders bij het verloop van de evaluatie
- de legitimiteit van de bevindingen
- het eigenaarschap van de evaluatieresultaten
- de bruikbaarheid van de evaluatie

Het is dus van belang om verschillende rapportering –en overlegprocedures op voorhand duidelijk af te spreken met de betrokken actoren. Zowel opdrachtgevers, andere stakeholders als het evaluatieteam dienen consensus te bereiken over de gebruikte communicatiekanalen die de eerlijkheid en de onpartijdigheid van het evaluatieproces verzekeren. Zo creëert men ook een terugvalpositie voor het oplossen van mogelijke conflicten die tussen de verschillende partijen kunnen ontstaan.

Hieronder geven we een aantal praktische tips die nuttig zijn in de uitbouw van de communicatiekanalen en interactiemogelijkheden.

- Klare afspraken met de opdrachtgever over verantwoordelijkheden en aanpak rond communicatie en het betrekken van andere stakeholders, als vertrekpunt

- Stakeholders in een vroeg stadium voorbereiden op het evaluatieproces

Het ontwikkelen van een gunstig 'evaluatieklimaat' kan door het informeren van alle betrokken stakeholders over het doel en de meerwaarde van de evaluatie. Vaak bestaat er van nature een associatie van evaluatie met controle, die echter in veel gevallen onterecht is. Door hierover klaarheid te scheppen kan potentiële tegenstand vermeden of verminderd worden en de interesse en inbreng van stakeholders benut worden.

- Aanmoedigen van stakeholders om hun inbreng te doen

Evaluatoren en opdrachtgevers kunnen de deelname van de verschillende stakeholders niet opleggen doch enkel aanmoedigen en ervoor zorgen dat iedereen de kans krijgt om in meer of mindere mate betrokken te zijn. Volgende acties dragen daartoe bij:

- o Duidelijkheid verschaffen over de timing van het evaluatieproces en de vorm van de evaluatierapportage.
- o Informatie verschaffen over de concrete kanalen voor communicatie en inspraak
- o Aanmoedigen van constructieve kritiek vanwege de stakeholders en het benutten van de voorziene kanalen
- o Creatie van een begeleidingscommissie of stuurgroep waarin de opdrachtgevers en belangrijkste stakeholders vertegenwoordigd zijn (cf. infra).
- o Koppeling van lange termijndoelstellingen met 'quick wins' of tussentijdse doelstellingen zodat stakeholders via een aantal snelle resultaten hun motivatie en betrokkenheid kunnen handhaven.

6.3.2 Stuurgroepformule

Deze sectie behandelt de vraag op welke manier een organisatie een begeleidingscommissie of stuurgroep voor het opvolgen van een evaluatie kan organiseren. In het eerste deel beantwoorden we de vraag, welke voordelen de creatie van een stuurgroep biedt voor de uitvoering van een evaluatie, terwijl het tweede deel de leden van een stuurgroep en hun functie verduidelijkt. De stuurgroepformule kan zowel bij interne als externe evaluaties toegepast worden. Vooral bij externe evaluaties vormt de stuurgroep een essentieel contactorgaan voor de evaluator.

In normale omstandigheden heeft de stuurgroep verantwoordelijkheden en taken vóór, tijdens en na het evaluatieproces. De betrokkenheid van de belangrijkste stakeholders via de creatie van een stuurgroep heeft zijn nut in de praktijk al veelvuldig bewezen. Vooral de aanwezigheid van de relevante institutionele en andere stakeholders die verantwoordelijk zijn voor de belangrijkste beleidsresultaten is van belang.

Een stuurgroep biedt tal van voordelen. De oprichting van een stuurgroep maakt de volgende elementen mogelijk:

- Een betere aanvaarding van de evaluatie door de stakeholders, daar een stuurgroep vertrouwensrelaties opbouwt en versterkt.
- Makkelijkere toegang tot en een beter begrijpen van de verschillende feiten en fasen die plaatsvinden gedurende de (geplande, lopende of voorbijge) uitvoering van het beleid.
- Kansen voor de optimalisering van het gebruik en het verhogen van de leerkansen voor de gebruikers als resultaat van de interacties tussen de verschillende aanwezige stakeholders.
- Conclusies en aanbevelingen, die alle aangehaalde inzichten en bedenkingen omsluiten.
- Een snelle verspreiding van de conclusies en het in rekening brengen van de aanbevelingen op een snelle en informele wijze.
- Grotere kansen dat de aanbevelingen en conclusies daadwerkelijk leiden tot acties en verandering.

Een stuurgroep omvat traditioneel een viertal categorieën van mensen:

- Het strategische management van de verantwoordelijke organisaties. Voorbeelden hiervan zijn organisaties die de nodige fondsen voorzien, de beleidscel van de administratie en indien van toepassing de verschillende administratieve of bestuurlijke niveaus die betrokken zijn bij het beleidsontwerp. Een multi-level benadering om het strategisch management te betrekken is zeer belangrijk in een zeer complex beleid waarbij ook verschillende geografische dimensies een belangrijke invloed uitoefenen.
- Het operationeel management ofwel de verantwoordelijken voor de concrete uitvoering van het beleid. Het betreft de afdelingshoofden van uitvoerende diensten, die toch nog een zekere afstand hebben en een helikopterperspectief op hun afdeling en rol.
- De stakeholders uit het beleidsveld, die rechtstreeks of onrechtstreeks met het beleid en de beleidseffecten in contact komen. Zo kunnen verschillende belangengroepen deel uitmaken van de stuurgroep.
- De externe experts vormen de laatste groep. Omwille van hun specifieke kennis van het beleidsveld ofwel hun methodologische kennis op het vlak van evaluaties kunnen ze een belangrijk klankbord vormen tijdens de ontwikkeling van het evaluatiedesign en de interpretatie van de resultaten. De inbreng van experts zal ook de discussie openstellen voor algemene aandachtspunten, die de evaluatie met zich meebrengt.

De belangrijkste doelen van de stuurgroep zijn het verzekeren van een hoogstaande en kwaliteitsvolle evaluatie en het werk van het evaluatieteam te faciliteren. Hieronder sommen we de belangrijkste taken van een stuurgroep op. Zoals reeds vermeld vervult een stuurgroep taken gedurende het hele evaluatieproces, gaande van het design tot en met de rapportering. Van bij de start is de stuurgroep betrokken bij het evaluatieproces. Vervolgens begeleidt het evaluatieteam de eigenlijke uitvoering van de evaluatie door het evaluatieteam op te volgen. Op die manier zorgt de stuurgroep ervoor dat het evaluatieteam de evaluatie uitvoert zoals overeengekomen in het bestek. Tot slot heeft de stuurgroep ook een belangrijke impact op het operationaliseren van de besluiten en aanbevelingen.

De stuurgroep draagt bij aan de volgende stappen in het evaluatieproces:

- Voorbereiding van het bestek (bij externe evaluatie)
- Verfijning van het evaluatiedesign in samenspraak met de evaluator
- Beslissingen in verband met toegang tot informatie
- Ondersteunen en opvolgen van het evaluatieproces en de evaluator
- De kwaliteit van de rapportage bewaken
- Verspreiding van de resultaten en de verdere opvolging

Bovenstaande elementen tonen duidelijk het belang van een goed samengestelde en functionerende stuurgroep aan. Een stuurgroep levert belangrijke bijdragen aan de uitvoering van evaluaties. Hierbij dient wel aandacht besteed te worden aan de onafhankelijkheid en evenwichtige samenstelling van de stuurgroep; het zijn kritieke succesfactoren voor de optimale werking van dit instrument. De stuurgroep mag onder geen beding in de plaats treden van het evaluatieteam bij de eigenlijke uitvoering van de evaluatie.

6.4 Rapportering en feedback

Evaluatoren die onvoldoende aandacht besteden aan de communicatie van de evaluatiebevindingen zijn op het einde van de rit vaak teleurgesteld. Een goed doordachte en toegepaste communicatiestrategie is een kritieke succesfactor opdat de evaluatie gebruikt wordt of invloed heeft. Daarvoor dient men te bepalen wat, waar, wanneer en hoe wordt gecommuniceerd. Ingeval de evaluatie wordt uitbesteed dient de externe evaluator duidelijke afspraken te maken met de opdrachtgever over de communicatie tijdens en na afloop van het evaluatieproces.

De communicatie van evaluator naar stakeholders is één element. De stakeholders kunnen ook feedback op deze communicatie leveren aan de evaluator. Deze tweerichtingsproces kan sterk bijdragen aan de

kwaliteit van de evaluatie en de rapportering van de bevindingen met het oog op de bruikbaarheid en invloed van de evaluatie in het kader van beleidsbeslissingen.

6.4.1 Algemene aandachtspunten

In elk geval dient men oog te hebben voor de diversiteit van het publiek, de variëteit van communicatiekanalen en de specifieke context van de evaluatie: het motief en gebruikszet, de politieke of persoonsgebonden gevoeligheid van de informatie, de status en de mate van eenduidigheid van de bevindingen, etc.

Het *publiek* is onder te verdelen in de opdrachtgever, de andere stakeholders die rechtstreeks of indirect belang hebben bij het beleid en de evaluatie ervan, en het ruime publiek. De communicatie zal naargelang de publiekscategorie verschillen: in hoeveelheid, focus, de mate van detail, frequentie en het kanaal.

Hoe meer een bepaalde publiekscategorie bij het evaluatieproces betrokken is, des te meer (gedetailleerde) informatie zij zal verwachten, over de verschillende aspecten van de evaluatie en op geregelde basis. Het brede publiek zal doorgaans geen boodschap hebben aan technische aspecten van de evaluatietechnieken die zijn gebruikt, maar voornamelijk geïnteresseerd zijn in de evaluatiebevindingen. Sommige 'derden' zoals andere onderzoeksinstituten kunnen wel sterke interesse betonen voor de methodische aspecten. Ook belanggroepen zullen vaak meer details willen vernemen over de gevolgde aanpak in de evaluatie om de gebruikswaarde vanuit hun perspectief te kunnen beoordelen. De communicatie zal logisch het meest intensief gericht zijn op de opdrachtgever en leden van de stuurgroep.

De evaluator dient ook omzichtig om te gaan met de informatie die men meedeelt. Bepaalde gegevens kunnen een zeer hoge politieke gevoeligheid hebben en zullen vertrouwelijk moeten behandeld worden. Dergelijke informatie wordt doorgaans niet buiten de cirkel van de sterkst betrokken stakeholders verspreid. Hetzelfde geldt voor gegevens over de privacy van individuele personen. Zulke gegevens, bijvoorbeeld verzameld in een enquête, zullen meestal niet eens herkenbaar voorgelegd worden aan de stuurgroep. Meestal zal men deze gegevens aggregeren zodat het niet of zeer moeilijk te achterhalen heeft welke respondent welke antwoorden heeft gegeven.

Belangrijk is dat men ook aangeeft wat de status van de informatie is: gaat het om tussentijdse of definitieve resultaten? Wat is de eventuele foutenmarge? Hebben ze betrekking op een deelanalyse? Zijn de conclusies overkoepelend of gebaseerd op slechts één van meerdere evaluatiecriteria? Dergelijke nuances en duidingen zijn cruciaal om de boodschap accuraat over te brengen.

Er is een brede waaier aan mogelijke *communicatiekanalen*:

- Persoonlijke discussies
- Korte communicaties: memos, e-mail, telefonisch, brochures, nieuwsbrief
- Stuurgroepoverleg
- Managementsamenvatting
- Tussentijdse rapporten
- Eindrapportage
- Presentaties
- Workshops met de stakeholders
- Persmededelingen

De evaluator zal constant moeten afwegen welk kanaal het meest geschikt is in functie van de combinatie van tijdstip, publiek en aard van de boodschap die men wil overbrengen.

6.4.2 Resultaten, conclusies en aanbevelingen

Een evaluatie is een proces gericht op het afleveren van een eindproduct. Onder meer aan de hand van een evaluatierapport kan het verloop en de uitkomst van dit proces meegedeeld worden (cf. infra). Als we spreken over het product van een evaluatie is het belangrijk om een onderscheid te maken tussen een trits van elementen:

- (1) resultaten
- (2) conclusies
- (3) aanbevelingen

Deze drie componenten staan in volgorde waarmee ze gegenereerd worden tijdens het evaluatieproces. De resultaten zijn de uitkomst van de standaardstap data-analyse. Conclusies zijn het product van de standaardstap beoordeling. Aanbevelingen ten slotte vormen een mogelijk maar niet noodzakelijk complement: het zijn adviezen en voorstellen gericht op eventuele beleidsbeslissingen. Hieronder bespreken we elk element wat meer in detail.

6.4.2.1 Resultaten

De resultaten zijn gekoppeld aan de data-analyse in het evaluatieproces. Resultaten zijn noodzakelijk om de evaluatievragen te beantwoorden. Ze beschrijven de situatie, geven in een ex ante evaluatie aan wat de verwachtingen zijn, en maken ex post de vergelijking van de verwachtingen met de realiteit, identificeren verklarende factoren voor het falen of welslagen van het beleid en geven met andere woorden aan of het beleid daadwerkelijk een verschil heeft gemaakt.

De resultaten hebben een grotere geloofwaardigheid en overtuigingskracht wanneer deze rechtstreeks voortvloeien uit de verzamelde data en gebaseerd zijn op solide analysetechnieken. Knelpunten voor de analyse hebben te maken met de onvolledigheid van data of tegenstrijdigheden op het vlak van verbanden en patronen. Toch is het belangrijk om de analyseresultaten te formuleren. Een professioneel ingestelde evaluator zal steeds de beperkingen en tekortkomingen van de gebruikte dataset en analysetechnieken expliciet vermelden en toelichten. Het toepassen van meerdere analysetechnieken kan tegenstrijdigheden aantonen doordat ze de dezelfde data op een andere wijze benaderen en verwerken. Soms kan men met deze werkwijze ook ogenschijnlijke tegenstrijdigheden opheffen; niet elke techniek kan bijvoorbeeld in dezelfde mate causale verbanden achterhalen.

Resultaten van de data-analyse kunnen meerdere functies vervullen:

- het geven van een *beschrijving* van een bepaald aspect van het evaluandum
- het *vergelijken* van het reële met het verwachte of gewenste
- het *identificeren* van een causaal verband en de precondities waarvoor het verband geldt, bv. redenen en verklaringen voor het welslagen of falen van beleid en relevante contextfactoren
- kenmerken van causale verbanden: de sterkte (impact), de richting (positief, negatief), de kans waarmee ze (kunnen) vóórkomen (probabiliteit), enzovoort.

De data-analyse gebeurt in functie van de evaluatievragen en –criteria. Resultaten van de data-analyse vormen dan ook het eerste deel van het antwoord op de evaluatievragen en toetsing aan de evaluatiecriteria. In functie hiervan zijn de conclusies complementair aan de resultaten.

6.4.2.2 Conclusies

Het formuleren van conclusies is geen technische materie. Conclusies zijn de uitkomst van de essentie van evalueren: het bepalen van de meerwaarde van het evaluandum. Conclusies worden gevormd door toetsing van de informatie bekomen uit de data-analyse aan de evaluatiecriteria, waarden en normen.

Een aantal technieken kunnen de evaluator helpen om tot een beoordeling te komen. Voorbeelden zijn expert panels, kosten-batenanalyse, multicriteria-analyse en impact assessments. De geschiktheid van de technieken hangt af van specifieke randvoorwaarden. Toch zullen deze tools eerder de input leveren voor het formuleren van de beoordelingen dan zelf beoordelingen aanbrenge. De voornaamste eigenschap van deze technieken is dat het aftoetsen aan het referentiekader met alle relevante evaluatiecriteria op een gestructureerde en systematische wijze ondersteunen. Het bepalen van de criteria, het relatieve gewicht van elk criterium en de normering blijven een menselijke keuze die uiteindelijk voortvloeit uit het primaat van de politiek.

Het is dus belangrijk dat het referentiekader duidelijk geëxpliciteerd is. Als aan deze voorwaarde is voldaan, heeft de evaluator de mogelijkheid en de verantwoordelijkheid om te komen tot conclusies. Professionalisme en ethische richtlijnen zijn daarbij onontbeerlijk.

Conclusies zijn gebaseerd op de verzamelde en geanalyseerde data. Als complement van de resultaten van de data-analyse vormen de conclusies de antwoorden op de gestelde evaluatievragen.

Conclusies zijn algemener van aard dan de resultaten: ze brengen een interpretatie en samenvatting van de resultaten en duiden op de implicaties van de vaststellingen. Vaak wordt er dus abstractie gemaakt van afzonderlijke maatregelen of interventies. Of de conclusies worden gepresenteerd in twee trappen. Men formuleert eerst specifieke conclusies over de afzonderlijke deelaspecten of met betrekking tot de aparte evaluatiecriteria. Daarna stelt men de overkoepelende of samenvattende eindbeoordeling op: de bevindingen in hoofdlijnen. Deze laatste kunnen bijvoorbeeld verwoord worden als: "sommige maatregelen zijn zeer efficiënt", "deze actoren hebben een belangrijke invloed gehad op de beleidsimplementatie", "bepaalde maatregelen hebben hun relevantie verloren", enzovoort.

Anderzijds dient de evaluator alle evaluatievragen te beantwoorden die deel uitmaken van de scope van de opdracht. Gebeurt dit niet, dan moet dit duidelijk geargumenteed worden. Alle conclusies die relevant en zinvol zijn voor de opdrachtgevers moeten opgenomen worden. Dit komt zeker tot uiting in situaties waar de evaluatie een aantal onverwachte negatieve effecten aan het licht brengt of als blijkt dat een bepaalde doelstelling niet langer haalbaar is.

Één van de belangrijkste regels voor het formuleren van conclusies is de koppeling met de data-analyse. Deze link dient zo expliciet mogelijk te zijn. De conclusies maken tevens melding van een aantal beperkingen en moeilijkheden op methodisch vlak die de data-analyse en de status van de resultaten beïnvloeden. Dit geeft de lezer inzicht in welke mate de conclusies eenduidig zijn, definitief, veralgemeenbaar, en dergelijk meer.

6.4.2.3 Aanbevelingen

Een derde en laatste mogelijke component van een evaluatie zijn de aanbevelingen. We gebruiken hier de voorwaardelijke wijs. In het theorie- en praktijkveld van beleidsevaluatie bestaat er namelijk geen

algemene consensus over de mogelijkheid en noodzakelijkheid: kan en moet een evaluator naast resultaten en conclusies ook aanbevelingen afleveren aan de opdrachtgever?

Michael Scriven formuleerde het als volgt: 'An evaluation without a recommendation is like a fish without a bicycle'. De stelling is zeker niet cryptisch bedoeld. In de visie van deze auteur zijn het twee aparte zaken. Sommige auteurs stellen inderdaad dat het formuleren van aanbevelingen bij voorbaat niet tot het takenpakket van de evaluator behoort. Ze halen daarvoor meerdere argumenten aan. Het vrijwaren van de onafhankelijkheid van de evaluator, in het bijzonder ingeval van externe evaluatie, is er één van. "Wat moet er gebeuren, wat zijn de next steps?", is geen vraag waarop de evaluator maar wel de beleidsmaker zelf dient te antwoorden. Bovendien, zo stellen deze auteurs, beschikt de evaluator doorgaans over te weinig informatie en expertise om deze vraag te beantwoorden.

Andere auteurs haken in op dat laatste argument en suggereren een oplossing: het betrekken van stakeholders bij de formulering van aanbevelingen. Stakeholders bezitten wel de nodige specifieke kennis en expertise. Aanbevelingen zijn volgens deze werkwijze de uitkomst van een interactief proces tussen het evaluatieteam en de belangrijkste stakeholders. Stakeholders kunnen als klankbord fungeren voor de evaluator en voorstellen toetsen op politieke en praktische haalbaarheid, op relevantie op basis van praktijkervaring. Ze kunnen alternatieve pistes voor handelen aanreiken die vervolgens door de evaluator worden uitgewerkt en verfijnd in aanbevelingen. Het schrijven van aanbevelingen is een delicate oefening; de wijze van formulering kan een grote impact hebben op hun latere toepassing of vertaling naar de beleidspraktijk. Dit element is een bijkomende reden voor de consultatie van de belangrijkste stakeholders bij het schrijven van aanbevelingen.

In de praktijk zal de opdrachtgever vaak geïnteresseerd zijn in concrete aanbevelingen vanwege de evaluator. In elk geval dienen aanbevelingen aan een aantal voorwaarden of kwaliteitseisen te voldoen. Centraal staat daar bij de functie van aanbevelingen: ze zijn bedoeld als voorstellen voor beslissingen over het geëvalueerde beleid. De aard van deze beslissingen kan zeer divers zijn. Algemeen heeft men de keuze tussen het onveranderd voortzetten, het bijsturen, het omgooien of het stopzetten van het beleid. We zetten de aandachtspunten op een rij.

- Aanbevelingen zijn *actiegericht*. Het moet dan ook gaan om praktisch haalbare suggesties om concrete problemen op te lossen of tegenstellingen weg te werken. Waar mogelijk, kan een schatting van de financiële en organisatorische impact toegevoegd worden. Hetzelfde geldt voor de realisatietermijn: een aantal aanbevelingen kunnen wellicht op korte termijn omgezet worden in de praktijk, terwijl andere adviezen betrekking hebben op de middellange en lange termijn. Dergelijke toegevoegde informatie maken de aanbevelingen tastbaarder voor de opdrachtgever, terwijl ze van de evaluator nauwkeurigheid vergen.

- Aanbevelingen worden geformuleerd *in de lijn van de focus van de evaluatie*. Een evaluatie ter ondersteuning van de beleidsplanning die de relevantie van de beleidsdoelstellingen toetst aan de noden, dient aanbevelingen te bevatten op het niveau van deze doelstellingen. Wanneer de implementatieaanpak centraal staat, moeten de adviezen een andere, meer procesmatige focus hebben.
- Aanbevelingen zijn *gekoppeld aan de resultaten en conclusies van de evaluatie*. Net zoals conclusies hun fundament moeten hebben in de resultaten van de data-analyse, moeten suggesties gebaseerd zijn op de resultaten en conclusies. Aanbevelingen zijn de laatste component van het product van een evaluatie. Dit product moet een samenhangend geheel zijn van resultaten, conclusies en aanbevelingen.
- Aanbevelingen zijn *aan een specifieke stakeholder of gebruiker gericht*. Het is belangrijk om adviezen te formuleren tot de actoren die er daadwerkelijk iets mee kunnen aanvangen. Net zoals bij het opstellen van de conclusies kan men in twee trappen werken: eerst algemene aanbevelingen die alle stakeholders aanbelangen, en vervolgens deze uitsplitsen naar concrete adviezen per stakeholder.
- Aanbevelingen zijn *handelingsvoorstellen die betrekking hebben op de bevoegdheden en verantwoordelijkheden van de geadresseerde stakeholders*. Een frequente frustratie bij beleidsmakers heeft te maken met aanbevelingen rond elementen waarop ze geen impact hebben. Bijvoorbeeld, een evaluatie van het functioneren van de politie in opdracht van de FOD binnenlandse zaken. In de loop van het onderzoek kan de evaluator stoten op elementen die eerder betrekking hebben op de gebrekkige werking van het gerecht en evaluator formuleert een aantal adviezen ter verbetering. De FOD binnenlandse zaken heeft echter geen directe impact daarop. Betekent dit dat de FOD binnenlandse zaken helemaal niets kan ondernemen te verbetering van het politieapparaat? Vanuit het standpunt van de FOD is de evaluator in deze evaluatie op een zijspoor beland. Los van het gegeven of de aangekaarte problemen bij het gerecht zich inderdaad stellen, zijn de aanbevelingen niet afgestemd op de opdrachtgever.
- Aanbevelingen kunnen ook als *keuzemenu* opgesteld worden. Meestal zijn immers meerdere wegen en opties naar verbetering of verandering van beleid.

Het is raadzaam de aanbevelingen te sorteren op belangrijkheid. Essentiële of strategische hefbomen waarmee een grote impact bereikt worden kunnen zo extra benadrukt worden en ontsnappen niet aan de aandacht temidden van de hele lijst van adviezen. Dergelijke indeling van aanbevelingen kan best worden gecombineerd met de groepering per stakeholder.

Het geheel van aanbevelingen zal ook variëren in ambitieniveau of ingrijpendheid. Om dit overzichtelijk te maken, kan de evaluator een onderscheid maken tussen een minimalistische en maximalistische visie of scenario.

De bovenstaande aandachtspunten maken duidelijk dat voor het formuleren van conclusies voldoende tijd moet worden uitgetrokken. Half werk op dit vlak reduceert de bruikbaarheid of mogelijke invloed van de hele evaluatie aanzienlijk. Om ze in de praktijk om te zetten, kan men werken in twee rondes om de aanbevelingen op te nemen in het evaluatierapport. In een eerste ronde stelt men een draft van de adviezen op. Men vraagt daarna feedback van de stakeholders. Het evaluatieteam kan met hen van gedachten wisselen over de betekenis van de resultaten, mogelijke verrassingen, onverwachte tegenstellingen en welke oplossingen daarvoor in aanmerking komen. Vervolgens kan de evaluator de aanbevelingen verfijnen of nuances aanbrengen. Het betrekken van degenen die de aanbevelingen uiteindelijk moeten omzetten naar de praktijk is zeer belangrijk. Deze werkwijze vergroot ook de betrokkenheid van de beleidsmakers en uitvoerders, leidt tot een groter draagvlak en grotere bruikbaarheid van de aanbevelingen gebaseerd op de evaluatie. Het evaluatieteam dient echter de logische koppeling van resultaten, conclusies en aanbevelingen te bewaken.

6.4.3 Het evaluatierapport

Elke evaluatie zal normaal ook uitmonden in een eindrapportering aan de opdrachtgever (verder: het evaluatierapport). Het evaluatierapport is in vele gevallen de finale neerslag van het evaluatieproces. Het is dan ook zaak dat de evaluator de boodschap op een correcte, leesvriendelijke en professionele manier overbrengt.

Omwille van het cruciale belang van een kwalitatief hoogstaand evaluatierapport als instrument in het kader van de bruikbaarheid en invloed van een evaluatie, gaan we dieper in op enkele belangrijke principes en onderdelen.

- Het rapport dient duidelijk en coherent geschreven te zijn. Het is belangrijk om de lengte van het rapport zelf te beperken tot de neerslag van de essentie voor elk van de stappen in het evaluatieproces. Uitwijdingen, toelichtingen, de uitwerking van analyses en worden best opgenomen in bijlagen of complementaire deelrapporten.
- De opname van een managementsamenvatting in het rapport is onontbeerlijk. Deze samenvatting is geschreven voor beleidsmakers en politici die meteen de hoofdlijnen van de aanpak, resultaten, conclusies en aanbevelingen moeten kunnen lezen.
- Een heldere structuur die alle stappen en componenten van het evaluatieproces en het product evenwichtig aan bod laat komen.
- Om de presenteerbaarheid en leesbaarheid van het geheel te optimaliseren is het raadzaam om figuren en tabellen op te nemen.

De ideale index van een evaluatierapport bestaat niet. Het evaluandum, de context van het evaluatieproces en de concrete verwachtingen van de stakeholders zullen dat mee bepalen. Toch kunnen we een aantal basisingrediënten voor een rapportage oplijsten. Ze hernemen de facto de belangrijke stappen en elementen van het evaluatieproces en omvatten de verschillende componenten van het product van de evaluatie.

- ✓ Managementsamenvatting
- ✓ Introductie
 - Opzet van de evaluatie
 - Contouren van het evaluandum en de evaluatiescope
 - Opdrachtgever en evaluator
 - Structuur van het rapport
- ✓ Evaluatiedesign
 - Beschrijving van het evaluandum
 - Scope van de evaluatie: evaluatievragen en -criteria
 - Aanpak van de dataverzameling
 - Aanpak van de data-analyse
 - Aanpak van de beoordeling
 - Achtergrondinformatie bij de voorgaande punten
- ✓ Resultaten
- ✓ Conclusies
- ✓ Aanbevelingen
- ✓ Bijlagen

Tabel 16: Bestanddelen van een evaluatierapport

- Managementsamenvatting

De lezer van een evaluatierapport heeft vaak niet de tijd om het hele rapport door te nemen. De managementsamenvatting biedt de lezer een compacte versie van het rapport met de essentiële informatie uit alle onderdelen van het rapport. Omdat heel wat geïnteresseerden zich uit praktische noodzaak moeten beperken tot deze samenvatting, dient de evaluator er veel aandacht aan te besteden.

- **Introductie**

De inleiding geeft de aanleiding, het motief, de betrokken actoren voor de evaluatie weer. De inleiding is de aangewezen plaats om de rationaliteit van de evaluatie aan te tonen. Waarin kadert de evaluatie? Wat is het evaluatiemotief? Wat wil de evaluatie bereiken? Ook een algemene beschrijving van het evaluandum en de ruwe afbakening of contouren van de inhoudelijke scope van de evaluatie komen hier reeds aan bod. Op basis van de inleiding dient de lezer de relevantie van het rapport voor zijn of haar informatiebehoefte te kunnen inschatten. Een element dat tevens deel uitmaakt van de inleiding is de 'leeswijzer': informatie die de lezer reeds wegwijs maakt in de teksten en onderdelen van het voorliggende rapport.

- **Evaluatiedesign**

In dit onderdeel van het rapport beschrijft het evaluatieteam het evaluatieonderwerp. Op die manier weet elke lezer welke elementen precies deel uitmaken van de evaluatie. Deze definiëring vormt de basis voor elke fase in het evaluatieonderzoek en dus ook voor elk onderdeel van het evaluatierapport. Het tweede element dat in dit deel van het rapport dient te zijn opgenomen is de vragenset. Welke vragen heeft de evaluator behandeld? Ook de assumpties die aan de basis liggen van de vragenkeuze dient hier te worden vermeld. Uit de vragen kan men ook de centrale evaluatiecriteria afleiden. Anderzijds bevat dit deel ook informatie over de aanpak van de uitvoering van de evaluatie: het verzamelen en analyseren van gegevens en de uiteindelijke beoordeling. Welke technieken heeft men gebruikt, welke bronnen? Aansluitend vermeldt men ook achtergrondinformatie, bijvoorbeeld over de knelpunten of beperking op het vlak van dataverzameling.

- **Presentatie van de resultaten**

Dit hoofdstuk bevat de resultaten van de gegevensanalyse en dient tevens als vertrekpunt voor de conclusies en aanbevelingen. Resultaten komen volledig tot hun recht wanneer de voorstelling op een duidelijke en overzichtelijke manier plaatsvindt. Hierbij kan het evaluatieteam gebruik maken van tabellen en grafieken om het geheel te verduidelijken. De interpretatie van de resultaten is even belangrijk als de dataverzameling: welke verbanden worden zichtbaar, welke patronen komen naar voor, welke 'waarden' zijn onverwacht, enz.

Een vaak gemaakte fout in evaluatierapporten is lezeronvriendelijk schrijven, vooral in het deel over de data-analyse. De koppeling met de initiële evaluatievragen moet duidelijk zijn doorheen heel het rapport. Verschillende werkwijzen zijn voorhanden: de vraag en antwoord methode, een bespreking per stakeholder of onderdeel van het onderzochte beleid. Basisidee is dat het evaluatieteam de lezer enige structuur aanbiedt.

- **Conclusies**

Een eerste element in dit deel zijn de criteria en standaarden op basis waarvan de conclusies tot stand zijn gekomen. Het expliciet vermelden is noodzakelijk. Het evaluatieteam kan de informatie presenteren op basis van het de tweespalt sterkten en beperkingen. Deze werkwijze laat toe dat de lezer een duidelijk inzicht krijgt in de rationaliteit achter het beoordelingsproces. Het maakt ook de overstap naar aanbevelingen helder.

- Aanbevelingen

Aanbevelingen vormen meestal het sluitstuk van het evaluatierapport. Vaak zullen ze in samenspraak met de opdrachtgever en/of andere stakeholders worden geformuleerd (cf. supra).

- Bijlagen

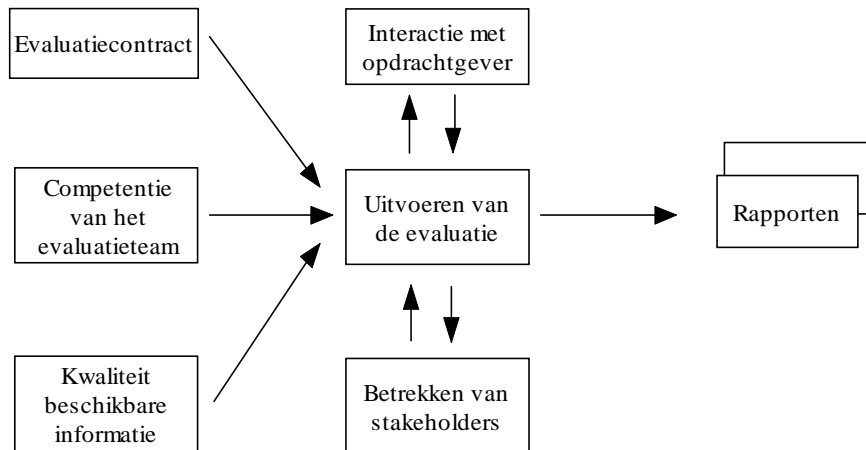
In bijlagen of afzonderlijke deelrapporten kunnen uitwijdingen, gedetailleerde uitwerkingen van analyses e.d. opgenomen worden. Wat geen noodzakelijke informatie voor de lezer is om de redeneringen en belangrijkste bevindingen doorheen het evaluatieproces te begrijpen, hoort thuis in de bijlagen.

De conclusies en aanbevelingen bieden de lezer de kans, een eigen opinie te vormen. Gebruikers zullen de conclusies gepresenteerd in het evaluatierapport vergelijken met hun eigen bevindingen. Op die manier kunnen evaluatierapporten mensen tot andere inzichten brengen. Een eerste stap naar beleidsverandering.

6.5 Kwaliteitsbewaking en meta-evaluatie

Doorheen het evaluatieproces moet kwaliteitsbewaking een permanent aandachtspunt zijn. Zowel de aantoonbare kwaliteit van de evaluatie als de perceptie van de kwaliteit door de stakeholders hebben een grote impact op de uiteindelijke bruikbaarheid en invloed van de evaluatie in het kader van beleidsbeslissingen. In deze paragraaf staan we stil bij de dimensies van kwaliteit en we overlopen welke criteria gehanteerd kunnen worden om een evaluatie te evalueren. Dat laatste wordt ook wel een meta-evaluatie genoemd.

Verschillende factoren spelen een rol naar de kwaliteit van een evaluatie. Onderstaand schema toont een aantal clusters van factoren:



Figuur 41: Factoren die de kwaliteit van een evaluatie beïnvloeden

Het evaluatiecontract (ingeval van een externe evaluatie) of de afsprakennota (ingeval van een interne evaluatie) vormt een belangrijk vertrekpunt. Hierin wordt de aanzet tot het evaluatiedesign gegeven, dat dan verder uitgewerkt kan worden in een aparte startnota. De competentie van de evaluator of het evaluatieteam is een andere cruciale factor om te komen tot een kwaliteitsvolle evaluatie. De competentie heet betrekking op kennis en vaardigheden en ervaring om het evaluatieproces op een professionele manier te doorlopen. De kwaliteit van de beschikbare informatie speelt uiteraard ook een essentiële rol: de betrouwbaarheid, geldigheid, relevantie en actualiteit van de gegevens bepalen mee de kwaliteit en dus bruikbaarheid van de evaluatie. Bij de uitvoering van de evaluatie komen heel wat aspecten kijken die cruciaal zijn: onder meer timing en focus zijn belangrijke aandachtspunten, maar ook de objectiviteit bij het formuleren van de conclusies. Een vlotte interactie tussen opdrachtgever en evaluator is een andere hefboom voor de kwaliteit van de evaluatie. Daarnaast is er de omgang met andere belangrijke stakeholders die een toegevoegde waarde tot het evaluatieproces kunnen leveren. De evaluator dient hun verwachtingen en verzuchtingen te integreren in het evaluatiedesign en waar mogelijk hun kennis te valoriseren via de dataverzameling of bij interpretatie van de analyseresultaten. Ten slotte is er de neerslag van evaluatieproduct: de rapportage. Aandachtspunten daaromtrent kwamen in voorgaande paragraaf al aan bod.

Wat betekent kwaliteit voor het evaluatieproces? Deze vraag is aanleiding voor menig debat en kent geen uniform antwoord. Gebaseerd op een aantal theoretisch inzichten en praktijkervaringen heeft een aantal evaluatieassociaties in het buitenland een lijst met evaluatiestandaarden opgesteld. Het is een normatief instrument zonder verplichtend karakter. Dergelijke standaarden zijn geldig voor evaluatie ongeacht het evaluandum, de institutionele context en de benadering (behalve voor personeelsevaluatie). Ze vormen een leidraad voor iedereen die betrokken is bij een evaluatie: evaluatoren, opdrachtgevers en andere stakeholders.

Bij wijze van voorbeeld hebben we hieronder de evaluatiestandaarden opgesteld door de Zwitserse evaluatieassociatie (SEVAL) opgenomen. De standaarden vallen in vier thematische groepen.

Bruikbaarheid	Deze standaarden garanderen dat de evaluatie gericht is op de informatiebehoeften van de vooropgestelde gebruikers van de evaluatie
Haalbaarheid	Deze standaarden verzekeren dat de evaluatie wordt uitgevoerd op een realistische, doordachte, diplomatische en kostenbewuste wijze.
Eigenaarschap	Deze standaarden verzekeren dat een evaluatie wordt uitgevoerd op een wettelijke en ethische manier en dat het welzijn van de stakeholders voldoende aandacht krijgt.
Accuraatheid	Deze standaarden verzekeren dat een evaluatie geldige en bruikbare informatie genereert en verspreidt.

Tabel 17: Thema's doorheen de evaluatiestandaarden van de Zwitserse evaluatieassociatie

Onder de vier thema's ressorteren respectievelijk 8, 3, 6 en 10 standaarden. Men heeft bewust geen gewichten toegekend noch aan afzonderlijke standaarden, noch aan de groepen. Deze keuze is ingegeven door de het idee dat de belangrijkheid van een standaard of groep varieert van evaluatie tot evaluatie. We overlopen de afzonderlijke standaarden, hun definities. Voor de thema's bruikbaarheid en accuraatheid vermelden we tevens bijhorende vragen die men kan stellen in het kader van kwaliteitsbewaking.

Bruikbaarheid

1. Identificeren van stakeholders

De personen die betrokken zijn bij en beïnvloed worden door een evaluatie worden geïdentificeerd zodat hun belangen en noden in rekening kunnen worden gebracht.

- *Zijn de verschillend stakeholders geïdentificeerd?*
- *Werden de noden van de stakeholders geanalyseerd?*

2. Verduidelijken van de evaluatiedoelstellingen

Iedereen die betrokken is bij de evaluatie zal verzekeren dat de doelstellingen van de evaluatie duidelijk zijn voor alle stakeholders.

- *Zijn het evaluatiemotief en de doelstellingen gekend en begrijpbaar?*
- *Liggen de doelstellingen in het verlengde van de noden en verwachtingen van de stakeholders?*

3. Geloofwaardigheid

Zij die evaluaties uitvoeren zijn zowel competent als betrouwbaar; dit zal helpen verzekeren dat de bevindingen van de evaluatie de grootste mogelijke mate van aanvaardbaarheid en geloofwaardigheid krijgen.

- *Voldoet de competentie van het evaluatieteam aan de verwachtingen?*
- *Ondersteunt het evaluatieteam de stakeholders opdat ze het evaluatieproces begrijpen en erin kunnen participeren?*

4. Scope en selectie van informatie

De scope en de selectie van informatie die werd verzameld maakt het mogelijk om pertinente vragen te stellen over het voorwerp van de evaluatie. Deze scope en selectie houdt ook rekening met de belangen en noden van de opdrachtgevende partij, en tevens deze van andere stakeholders.

- *Beantwoordt het evaluatiedesign aan het evaluatiemotief en de verwachtingen van de opdrachtgever en stakeholders?*

5. Transparantie van waardeoordelen

De achterliggende redenering en standpunten waarop de interpretatie de evaluatieresultaten berust, worden zo beschreven dat de basis voor de waardeoordelen duidelijk is.

- *Is het referentiekader om tot conclusies te komen geëxpliciteerd?*
- *Zijn de conclusies geënt op de resultaten van de data-analyse?*

6. Begrijpbaarheid en helderheid van rapportage

Evaluatierapporten beschrijven het voorwerp van de evaluatie, met inbegrip van diens context, doelstellingen, de gestelde vragen, de gevolgde aanpak, alsook de bevindingen van de evaluatie - op zulke manier dat de meest pertinente informatie beschikbaar is en gemakkelijk begrijpbaar.

- *Is het eindrapport bevattelijk opgesteld?*

7. Tijdige rapportage

Significante tussentijdse resultaten, alsook finale rapporten, worden beschikbaar gesteld voor de bedoelde gebruikers, zodanig dat ze tijdig kunnen gebruikt worden.

- *Wordt de tijdslijn van het evaluatieproces gerespecteerd?*
- *Zijn de resultaten, conclusies en aanbevelingen tijdig beschikbaar?*

8. Evaluatie impact

De planning, uitvoering en presentatie van de evaluatie moedigt stakeholders aan om zowel het evaluatieproces op te volgen als de evaluatie te gebruiken.

- *Welke mechanismen en werkwijze worden voorzien die toelaten om het evaluatieproces op te volgen?*

- *Welke momenten en kanalen worden voorzien om de evaluatiebevindingen kenbaar te maken en over te dragen naar de opdrachtgever en andere stakeholders?*

Haalbaarheid

1. Praktische procedures

Evaluatieprocedures worden ontworpen zodanig dat de vereiste informatie wordt verzameld zonder onnodige verstoring van het evaluandum of van de evaluatie zelf.

2. Anticiperen op politieke leefbaarheid

Bij de planning en uitvoering van de evaluatie houdt men rekening met uiteenlopende standpunten en verschillende belangen. Dit gebeurt met het oog op het verkrijgen van medewerking en het vermijden van pogingen door een of andere groep om evaluatieactiviteiten te beperken of de resultaten te verstoren of te misbruiken.

3. Kosteneffectiviteit

Evaluaties genereren informatie waarvan de waarde de productiekost verrechtvaardigt.

Eigenaarschap

1. Formeel geschreven akkoord

De verantwoordelijkheden van de partijen die overeenkomen om een evaluatie uit te voeren worden bepaald in een schriftelijk akkoord om de contracterende partijen te verplichten de overeengekomen voorwaarden te vervullen, of indien niet het geval is, het akkoord te heronderhandelen.

2. Verzekeren van individuele rechten en welzijn

Evaluaties worden gepland en uitgevoerd zodanig dat de rechten het welzijn van individuen worden beschermd en gerespecteerd.

3. Respecteren van menselijke waardigheid

Evaluaties worden zo georganiseerd opdat de contacten tussen de deelnemers gekenmerkt worden door wederzijds respect.

4. Volledige en evenwichtige beoordeling

Evaluaties zijn volledig en evenwichtig wanneer ze de sterktes en zwaktes van het evaluandum waarderen en presenteren, op zulke wijze dat sterktes uitgebouwd kunnen worden en probleemgebieden aangepakt kunnen worden.

5. Bevindingen beschikbaar maken

De betrokken partijen verzekeren dat de bevindingen van de evaluatie beschikbaar worden voor alle geëffecteerde personen en tevens voor alle anderen die een geëgitimeerde aanspraak hebben om ze te ontvangen.

6. Belangenconflicten aangeven

Belangenconflicten worden openlijk en eerlijk aangepakt zodat ze zo min mogelijk het evaluatieproces en de conclusies compromitteren.

Accuraatheid

1. Precieze beschrijving van het voorwerp van de evaluatie

Het evaluandum moet duidelijk en nauwkeurig beschreven, gedocumenteerd en eenduidig geïdentificeerd worden.

- *Bevat het evaluatierapport alle elementen van het te evalueren beleid?*

2. Analyseren van de context

De omgevingsfactoren die een invloed hebben op het evaluandum worden geïdentificeerd.

- *Beschrijft het rapport de context van het beleid?*

3. Precieze beschrijving van doelstellingen, vragen en procedures

De nagestreefde doelstellingen, gestelde vragen en gehanteerde aanpak in de evaluatie worden voldoende precies omschreven en gedocumenteerd zodat ze zowel geïdentificeerd als beoordeeld kunnen worden.

- *Is de scope eenduidig en duidelijk geformuleerd en consistent opgebouwd?*
- *Is het evaluatiedesign te rechtvaardigen in functie van het beantwoorden van de evaluatievragen en het vullen van het evaluatiemotief?*
- *Is de scope voldoende flexibel geformuleerd om tijdens het evaluatieproces eventueel nog vragen toe te voegen?*
- *Voorziet het design in voldoende mogelijkheden opdat het evaluatieteam voldoende en relevante informatie kan verzamelen om de voorliggende vragen te beantwoorden?*

4. Betrouwbare bronnen van informatie

De gebruikte informatiebronnen worden voldoende precies beschreven zodat hun adequaatheid kan beoordeeld worden.

- *Heeft het evaluatieteam de juiste situaties of organisaties gekozen om valide vergelijkingen te kunnen maken?*

- *Heeft men bepaalde keuzes in de dataverzameling toegelicht?*

5. Geldige en betrouwbare informatie

Om de validiteit en betrouwbaarheid van de interpretatie te verzekeren, is het nodig om de procedures hiervoor te selecteren, te ontwikkelen en te gebruiken.

- *Is het proces van dataverzameling goed beschreven?*
- *Heeft het evaluatieteam de meest geschikte technieken toegepast?*

6. Systematisch controle op fouten

De verzamelde, geanalyseerde en gepresenteerde informatie in een evaluatie is systematisch gecheckt op fouten.

- *Is de verzamelde informatie verifieerbaar?*

7. Kwalitatieve en kwantitatieve analyse

Kwalitatieve en kwantitatieve informatie wordt systematisch en passend geanalyseerd in een evaluatie, op zulke manier dat de gestelde evaluatievragen daadwerkelijk beantwoord kunnen worden.

- *Beantwoordt de evaluatie de belangrijkste vragen?*
- *Zijn de resultaten logisch af te leiden uit de data-analyse?*

8. Omstandige conclusies

De bereikte conclusies worden expliciet uitgewerkt zodat stakeholders ze kunnen begrijpen en beoordelen.

- *Zijn de weergegeven bevindingen duidelijk?*
- *Worden de bevindingen voldoende becommentarieerd?*

9. Neutrale rapportering

Rapportage is vrij van bias door persoonlijke gevoelens of voorkeuren van gelijk welke bij de evaluatie betrokken partij; evaluatierapporten brengen de conclusies op een neutrale manier.

- *Liggen de conclusies in het verlengde van de evaluatievragen?*

10. Meta-evaluatie

De evaluatie zelf wordt geëvalueerd op basis van de bestaande (of andere relevante) standaarden, zodat de evaluatie passend wordt uitgevoerd, en zodat stakeholders op het einde de sterktes en zwaktes van de evaluatie kunnen inschatten.

Een totale kwaliteitsbewaking heeft niet alleen tot doel de kwaliteit van het geleverde werk door het evaluatieteam te beoordelen tijdens en na het evaluatieproces. Het kan ook alle mogelijke moeilijkheden waarmee de evaluatie wordt geconfronteerd in kaart brengen. Daaruit kan men lessen trekken voor toekomstige evaluaties en in het bijzonder voor het management van deze evaluaties.

Onderstaand voorbeeldschema bevat een aantal van de kwaliteitsstandaarden en criteria. Opdrachtgevers en lezers van het evaluatierapport kunnen het evaluatieproces en/of het rapport scoren op basis van dergelijk schema. Hoe meer elementen van de vermelde aandachtspunten terugkeren in de aanpak en het product van de evaluatie, des te hoger de kwaliteit van de evaluatie. Onder het schema presenteren we een mogelijke werkwijze om de totaalscore te berekenen. Volgens dit voorbeeldschema zijn evaluaties met een totaalscore tussen 0 en 2 ondermaats van kwaliteit. Een totaalscore van 5 of meer wijst op een kwaliteitsvolle evaluatie.

1 Identificatie van de stakeholders			
<input type="checkbox"/> Duidelijke identificatie van de doelgroep <input type="checkbox"/> Consulteren van de stakeholders om hun noden te identificeren <input type="checkbox"/> Stakeholders engageren voor het hele evaluatieproces <input type="checkbox"/> Evaluaties staan steeds open voor nieuwe stakeholders			
<input type="checkbox"/> 4 Zeer goed <input type="checkbox"/> 3 Goed <input type="checkbox"/> 2 Matig <input type="checkbox"/> 1 Slecht			
2 Evaluatieteam			
<input type="checkbox"/> Aantrekken van competente evaluatoren <input type="checkbox"/> Aantrekken van evaluatoren die tegemoet komen aan de noden en vragen van de stakeholders <input type="checkbox"/> Ondersteunen van stakeholders opdat ze het evaluatieproces begrijpen <input type="checkbox"/> Gepast reageren op kritiek en suggesties vanwege de stakeholders			
<input type="checkbox"/> 4 Zeer goed <input type="checkbox"/> 3 Goed <input type="checkbox"/> 2 Matig <input type="checkbox"/> 1 Slecht			
3 Evaluatiescope			
<input type="checkbox"/> Prioriteit geven aan de meest belangrijke vragen <input type="checkbox"/> Voorzien van de nodige flexibiliteit om vragen tijdens de evaluatie toe te voegen <input type="checkbox"/> Voldoende informatie verzameld om de meerwaarde van het beleid te achterhalen <input type="checkbox"/> Duidelijke definiëring van de scope			
<input type="checkbox"/> 4 Zeer goed <input type="checkbox"/> 3 Goed <input type="checkbox"/> 2 Matig <input type="checkbox"/> 1 Slecht			
4 Verdedigbaar evaluatiedesign			
<input type="checkbox"/> Consistente en eenvormige opbouw <input type="checkbox"/> Het gebruik van geschikte methoden en technieken <input type="checkbox"/> Geschikte keuze van het onderzoeksobject <input type="checkbox"/> De verzameling van alle noodzakelijke data			
<input type="checkbox"/> 4 Zeer goed <input type="checkbox"/> 3 Goed <input type="checkbox"/> 2 Matig <input type="checkbox"/> 1 Slecht			
5 Betrouwbare data			
<input type="checkbox"/> Gebruik maken van de meest aangewezen bronnen <input type="checkbox"/> Documenteren en weergeven van de informatiebronnen <input type="checkbox"/> Het inzetten van een hele waaier technieken en methoden <input type="checkbox"/> Verklaar het nut van de verzamelde informatie			
<input type="checkbox"/> 4 Zeer goed <input type="checkbox"/> 3 Goed <input type="checkbox"/> 2 Matig <input type="checkbox"/> 1 Slecht			
6 Analyse van de informatie			
<input type="checkbox"/> Weergave en motivering van de gebruikte methoden en technieken <input type="checkbox"/> Het gebruik van verschillende analytische technieken om de consistentie en de herhaalbaarheid van de bevindingen te bepalen <input type="checkbox"/> Rapportering van de beperkingen van de analyses en interferenties <input type="checkbox"/> Verifieer de accuraatheid van de bevindingen			
<input type="checkbox"/> 4 Zeer goed <input type="checkbox"/> 3 Goed <input type="checkbox"/> 2 Matig <input type="checkbox"/> 1 Slecht			
7 Resultaten			
<input type="checkbox"/> De resultaten zijn logisch af te leiden uit de data - analyse <input type="checkbox"/> Duidelijk verband tussen de resultaten en de geformuleerde hypothesen <input type="checkbox"/> Betrokkenheid van stakeholders bij de interpretatie van de analyses <input type="checkbox"/> Duidelijke en eenvormige presentatie van de resultaten			
<input type="checkbox"/> 4 Zeer goed <input type="checkbox"/> 3 Goed <input type="checkbox"/> 2 Matig <input type="checkbox"/> 1 Slecht			
8 Conclusies			
<input type="checkbox"/> Beperken van de conclusies tot diegenen die overeenstemmen met de vragen zoals vermeldt in het bestek <input type="checkbox"/> Weergeven van de informatie die de conclusies ondersteunt en verklaar waarom deze niet werden weerhouden <input type="checkbox"/> Betrokkenheid van de stakeholders in de formulering van de conclusies <input type="checkbox"/> Een duidelijke en operationele formulering van de conclusies			
<input type="checkbox"/> 4 Zeer goed <input type="checkbox"/> 3 Goed <input type="checkbox"/> 2 Matig <input type="checkbox"/> 1 Slecht			
9 Duidelijkheid van het rapport			
<input type="checkbox"/> Duidelijke en makkelijk begrijpbare tekst <input type="checkbox"/> Het rapport als antwoord op de gestelde vragen en doelstellingen <input type="checkbox"/> Het gebruik van voorbeelden ter verduidelijking van de redering <input type="checkbox"/> Duidelijke structuur die overeenstemt met het bestek			
<input type="checkbox"/> 4 Zeer goed <input type="checkbox"/> 3 Goed <input type="checkbox"/> 2 Matig <input type="checkbox"/> 1 Slecht			
<table border="0" style="width: 100%;"> <tr> <td style="width: 50%;"> Aantal keer zeer goed (0-9) ____ x4 = ____ Aantal keer goed (0-9) ____ x3 = ____ Aantal keer matig (0-9) ____ x2 = ____ Aantal keer slecht (0-9) ____ x1 = ____ Totaal = ____ </td> <td style="width: 50%;"> <input type="checkbox"/> 7 (78%) tot 9 : Zeer goed <input type="checkbox"/> 5 (56%) tot 6 : Goed <input type="checkbox"/> 2 (22%) tot 4 : Matig <input type="checkbox"/> 0 (0%) tot 2 : Slecht ____ (Totale score) + 9 = ____ x 100 = ____ </td> </tr> </table>		Aantal keer zeer goed (0-9) ____ x4 = ____ Aantal keer goed (0-9) ____ x3 = ____ Aantal keer matig (0-9) ____ x2 = ____ Aantal keer slecht (0-9) ____ x1 = ____ Totaal = ____	<input type="checkbox"/> 7 (78%) tot 9 : Zeer goed <input type="checkbox"/> 5 (56%) tot 6 : Goed <input type="checkbox"/> 2 (22%) tot 4 : Matig <input type="checkbox"/> 0 (0%) tot 2 : Slecht ____ (Totale score) + 9 = ____ x 100 = ____
Aantal keer zeer goed (0-9) ____ x4 = ____ Aantal keer goed (0-9) ____ x3 = ____ Aantal keer matig (0-9) ____ x2 = ____ Aantal keer slecht (0-9) ____ x1 = ____ Totaal = ____	<input type="checkbox"/> 7 (78%) tot 9 : Zeer goed <input type="checkbox"/> 5 (56%) tot 6 : Goed <input type="checkbox"/> 2 (22%) tot 4 : Matig <input type="checkbox"/> 0 (0%) tot 2 : Slecht ____ (Totale score) + 9 = ____ x 100 = ____		

Tabel 18: Voorbeeld van checklist voor een meta-evaluatie

Literatuur

European Commission (1999). *The Means Collection – Evaluating socio-economic programmes*. Luxembourg: Office for Official Publications of the European Communities.

Widmer T., Landert C. and Bachmann N. (2000). *Evaluation Standards of SEVAL, the Swiss Evaluation Society*. s.l.

7 Ontwikkeling en professionalisering van evaluatiecapaciteit

7.1 Ontwikkeling van institutionele capaciteit

Beleidsbeoordelingen zijn ondertussen ingeburgerd in tal van landen binnen en buiten Europa. Toch verschillen deze landen op het vlak van de institutionalisering van evaluatie in de beleidscyclus. Landen hebben historisch verschillende routes gekozen om evaluaties te ontwikkelen en te implementeren. De verklaringen daarvoor werden al uitvoerig besproken in hoofdstuk 2.

Het concept institutie wordt in verschillende contexten aangehaald en gebruikt. De definitie van institutie in het kader van evaluaties luidt als volgt: een kader ten behoeve van verschillende actoren waardoor het gedrag en de rol van deze actoren met betrekking tot het plannen en uitvoeren van evaluaties een zeker voorspelbaar karakter krijgt. Institutionalisering duidt dus zowel op het vastleggen als vastliggen van een aantal praktijken.

Een evaluatiecultuur kan binnen een organisatie sterk ontwikkeld zijn zonder dat deze formeel bestaat of vastgelegd is intenties of wettelijke bepalingen. Auteurs bepalen verschillende institutionaliseringstypes. Het kan gaan over regulering, cognitieve processen of normatieve processen. Anderen zien institutionalisering als een voorwaarde om tot evaluatie te komen, om te vermijden dat evaluaties slechts sporadisch plaatsvinden en dat voldoende capaciteit wordt voorzien om evaluaties te kunnen uitvoeren. Volgens bepaalde auteurs vindt de institutionalisering van evaluatie plaats op het ogenblik dat men het evaluatieproces kan definiëren als een sociaal proces.

Institutionalisering bestaat uit tal van elementen. Toch kan men drie categorieën van variabelen onderscheiden: de organisatiestructuur, de evaluatiebepalingen en de evaluatiegemeenschap.

De variatie in structuur van overheidsorganisaties is enorm. De mate van institutionalisering van evaluatie in een bepaald land hangt grotendeels af van de staatsstructuur en de rol van de centrale instanties als het parlement, de regering en de administratie. In een federale staat zal de institutionalisering van evaluatie wellicht anders verlopen dan in een nationale eenheidsstaat.

De evaluatiebepalingen zijn elementen die door de wetgever in bij wet of in reglementeringen worden vastgelegd. De gedetailleerdheid verschilt in sterke mate van land tot land. Uit onderzoek blijkt dat dit element in tal van landen slechts van secundair belang is.

Het bestaan van een evaluatiegemeenschap is het derde element in de definitie van institutionalisering. Dergelijke gemeenschap vormt het centrum voor de overdracht en uitwisseling van kennis, ervaringen en

ideeën. Wetenschappelijke tijdschriften en rapporten worden door deze fora ontwikkeld en verspreid. Een indicatie van institutionalisering is de mate waarin wetenschappelijk over het onderwerp wordt gereflecteerd. Een sterk ontwikkelde gemeenschap kan zelfs haar eigen standaarden en kwaliteitsnormen ontwikkelen over de wijze waarop een evaluatieproces dient te verlopen.

Naast deze drie centrale elementen van institutionalisering kunnen nog twee andere elementen worden onderscheiden om de mate van institutionalisering te bepalen.

Het bestaan van professionele evaluatieteams is tevens een factor aan de hand waarvan de mate van institutionalisering wordt bepaald. Professionele teams of individuen zijn terug te vinden in verschillende contexten en settings: interne evaluatoren, private consultants, universitaire onderzoeksinstituten, etc. Vaak is het zeer moeilijk om het juiste aantal te bepalen. Men kan als proxy het aantal consultants gespecialiseerd in evaluatie, het aantal offertes of het gespendeerde budget aan evaluatie nemen.

Het aantal gerealiseerde evaluaties is een laatste indicator van institutionalisering. Deze indicator is het meest evident en het minst betwistbaar. Een dergelijke telling maakt het mogelijk om verschillende landen of regio's met elkaar te vergelijken. In de praktijk blijkt echter dat deze optelsom niet eenvoudig te maken is. De vlag ('evaluatie') dekt niet altijd de lading of, omgekeerd sommige studies krijgen niet het predicaat van beleidsevaluatie, maar zijn dit de facto wel.

7.2 Groeimodel voor de inbedding van een evaluatiecapaciteit op maat

Bij het ontwikkelen van een aangepaste evaluatiecapaciteit staan verschillende vragen centraal:

- Welke capaciteit is er nodig?
- Hoe kunnen we deze capaciteit opbouwen?
- Welke knelpunten zijn er?
- Hoe kunnen de obstakels weggenomen worden?

Tussen evaluatiecapaciteit en evaluatiecultuur bestaat een wisselwerking. Om een capaciteit op te bouwen zal er een begin van evaluatiecultuur moeten zijn als drijfveer en om bepaalde weerstanden te overwinnen. Omgekeerd kan de opbouw van de capaciteit en de aanwending ervan bijdragen tot de verdere ontwikkeling van een evaluatiecultuur.

Doorheen Europa is de aandacht voor evaluatie exponentieel gegroeid, parallel met of dankzij moderniseringsprocessen. Tevens hebben verscheidene ervaringen reeds aangetoond dat organisaties de aanwezige weerstand kunnen overwinnen.

De implementatie van een beleidsevaluatie en de wijze van organisatorische inbedding is niet éénvormig. Verschillende trajecten zijn mogelijk. Een gedeeltelijke verklaring hiervoor is de verschillende vertrekpositie van de organisaties. Tradities, ervaringen, professionele en technische kwaliteiten en academische tradities vertonen grote verschillen tussen landen.

De ontwikkeling van de evaluatiecapaciteit start met de bestaande situatie en de diagnose van de aanwezige noden. Het is een proces dat tijd kost en verschillende fasen doorloopt. De eerste stappen dient men te nemen in het kader van een lang termijnperspectief gedefinieerd via doelstellingen op middellange en lange termijn. De definiëring van deze doelstellingen dient vergezeld te zijn van de noodzakelijke budgetten. Een groeimodel is dus de meest aangewezen aanpak om een evaluatiecapaciteit op te bouwen.

Het model zoals hieronder gepresenteerd is een ideaaltype; een implementatiestrategie via verschillende fasen en intermediaire doelen. Een ideaalmodel is vooral nuttig om een aantal opties te presenteren en keuzes te verduidelijken. In de praktijk kiezen verschillende organisaties verschillende implementatietrajecten. Het model is zeer flexibel en aanpasbaar aan concrete omstandigheden. Het model wil zowel organisaties op weg helpen die al enige ervaring hebben met evaluaties als organisaties die voor het eerst met evaluaties worden geconfronteerd.

Fase 1: Opstart

De aanleiding of het vertrekpunt van organisaties om te starten met de opbouw van een interne evaluatiecapaciteit kan zeer verschillend zijn.

Essentieel is duidelijkheid over welk evaluatiemotief centraal staat.

Het evaluatiemotief heeft een impact op verschillende elementen:

- het type evaluatie
- mogelijke knelpunten
- de vereiste capaciteit
- de organisatie van de capaciteit

In hoofdstuk 2 werd de link tussen het evaluatiemotief en een aantal typologieën van evaluaties besproken. De boodschap daarbij was dat er vaak geen 1-op1 relatie mogelijk is tussen een bepaald evaluatiemotief en een type van evaluatie, maar er is wel een clustering mogelijk. Onderstaande tabel herneemt de onderlinge verbanden nogmaals.

Evaluatiedoel	timing	focus	intern <--> extern
Ondersteuning van beleidsplanning	ex ante	beleidsinhoud	
Verbeteren van de beleidsuitvoering en organisatorische ontwikkeling	tussentijds		
Verantwoording afleggen	ex post		
Beleidsleren	ex post	beleidsimpact	

Tabel 19: Koppeling van evaluatiedoelstellingen en evaluatietypologieën

Om de impact van het evaluatiedoel op randvoorwaarden en organisatie van een evaluatiecapaciteit te illustreren, nemen we een zogenaamd kritiek scenario: evaluatie van het verantwoordingsdoel met een verplichtend karakter. Het scenario is kritiek doordat het wellicht de grootste spanningsvelden en knelpunten oproept. Tegelijk is het een zeer herkenbaar scenario voor veel overheidsorganisaties in België en Vlaanderen. Dat veel overheidsorganisaties voor het eerste met evaluatie in aanraking komen in het kader van dit evaluatiedoel, heeft te maken met het traject dat ons land gevolgd heeft op het vlak van de verspreiding van het evaluatiedenken. Externe druk speelde hierbij een grote rol als katalysator voor de implementatie van evaluaties, hetzij vanuit buitenlandse internationale organisaties, hetzij vanuit ideeën die centraal stonden in binnenlandse processen van hervorming en modernisering.

Het kritiek karakter van dergelijke evaluaties heeft te maken met het feit dat ze vaak verankerd zijn in een formeel en verplichtend karakter. Het risico op disfuncties en knelpunten is het grootst bij evaluaties vanuit een opgelegd verantwoordingsdoel:

- beperkt eigenaarschap
- informatieasymmetrie
- ontwijkend gedrag
- implementatieproblemen

Een louter formeel en extern aangedreven evaluatiekader vormt een rem op het eigenaarschap van de uitgevoerde evaluaties. Evaluaties worden enkel geassocieerd met controle en een verplichte formaliteit.

We stelden dat voor evaluaties vanuit het verantwoordingsdoel een externe vorm van evaluatie meer aangewezen is, omwille van de sterktes van dit type evaluatie onder meer op vlak van geloofwaardigheid en onafhankelijkheid. In de context van een verplichte evaluatie ter verantwoording zijn er twee mogelijkheden: de betrokken organisatie dient een zelfevaluatie af te leveren of een derde actor - een externe evaluator - voert de evaluatie uit.

We willen de lezer er hier op wijzen dat zulke constructie op verschillende schaalgrootte kan voorkomen en varianten kan hebben. Zo kan het parlement een evaluatie vragen van het beleid van de minister en diens administratie. De administratie kan de evaluatie zelf uitvoeren. Het parlement kan ook eisen dat een externe evaluator ingeschakeld wordt. Op een kleinere schaal kan het topmanagement van een overheidsorganisatie aan één van de onderafdelingen een evaluatie vragen. Deze afdeling kan een zelfevaluatie uitvoeren. Het is ook mogelijk dat een externe actor, hetzij een andere afdeling die gespecialiseerd is in evaluatietaken binnen de organisatie, of een consultant belast wordt met de uitvoering van de evaluatie.

Ongeacht de schaal of de variant, zal echter steeds de problematiek van informatieasymmetrie spelen. De betrokken administratie of afdeling in het voorbeeld zal meer informatie hebben dan de opdrachtgever (parlement, topmanagement) en desgevallend ook de externe evaluator.

In combinatie met het voorgaande punt kan dit leiden tot ontwijkend gedrag. De actor verantwoordelijk voor het evaluandum gaat bepaalde informatie achterhouden of filteren, om op die manier de verantwoording sturen zodat een gunstig resultaat wordt behaald en eventuele sancties (bv. budgetreductie) worden ontlopen. Dergelijk extern aangestuurd en formeel systeem kampt dan ook onvermijdelijk met implementatieproblemen. De organisaties zullen ook minder aandacht besteden aan het evaluatiedesign, vermits zij voor zichzelf er geen meerwaarde in zien.

Bovenstaand scenario is realistisch. Deze handleiding heeft niet de ultieme oplossing in petto. Wel stelt dit scenario een aantal spanningsvelden en aandachtspunten zeer scherp die belangrijk zijn bij de uitbouw van een evaluatiecapaciteit. In wat volgt gaan daar verder op in. Een aantal spanningsvelden zal altijd blijven bestaan en zijn misschien nooit helemaal oplosbaar.

Wanneer men echter een duurzame evaluatiecapaciteit en evaluatiecultuur wil uitbouwen en verder ontwikkelen zijn een aantal strategische aandachtspunten van belang:

- openheid voor de verschillende evaluatiemotieven
- streven naar coördinatie van capaciteit en synergieën van evaluatiemotieven
- inbedding van evaluatie doorheen de beleidscyclus

In de opstartfase worden verschillende acties ontplooid.

- Identificeren van stakeholders
- Stakeholders betrekken bij de opstartfase
- Diagnose van de behoeften
- Introductie van een werkwijze en kader voor zelfevaluatie
- Investeren in monitoring

Het belangrijk om de stakeholders in de organisatie te identificeren en te betrekken van bij de opstartfase. Dit heeft als doel om de hun aandacht voor evaluaties te verhogen en om hun verwachtingen te managen. Met de stakeholders dient men na te gaan welke behoeften op het vlak van evaluatie er bestaan. Vervolgens dient men een kader te ontwikkelen om zelfevaluaties te kunnen uitvoeren. Dit maakt het evaluatieproces zichtbaar in de organisatie en draagt bij tot de ontwikkeling van een evaluatiecultuur. De uitwerking van kader in de eerste fase van de ontwikkeling heeft nieuwe evaluatieactiviteiten tot gevolg. Evaluaties kunnen niet zonder gegevens en informatie. Het investeren in monitoring zal noodzakelijk zijn ter ondersteuning van toekomstige evaluatieactiviteiten.

Fase 2: Evaluatiecoördinatie

De diagnose van de behoeften en de uitwerking van een kader en ondersteunende informatie-instrumenten, in een eerste stap in de capaciteitsopbouw.

Een tweede stap is het voorzien in ondersteunende instrumenten en een structuur in functie van een gecoördineerde evaluatiecapaciteit.

De kwaliteit en coördinatie van de evaluaties kan verhogen met de ontwikkeling van handleidingen en de voorbereiding van instrumenten ten behoeve van de evaluatoren. De output van dit proces kan zeer divers zijn, gaande van de publicatie van handboeken, handleidingen, een set van sleutelindicatoren, de ontwikkeling van software en dergelijke. Het grote voordeel van deze manier van werken is de mogelijkheid om een kwaliteitsbewaking in te voeren. Dit is dan ook de reden waarom de voorbereiding van kwaliteitsstandaarden te beschouwen valt als een onderdeel van de coördinatie. Dit zorgt er tevens voor dat een te grote differentiatie aan evaluatiebenaderingen en terminologie wordt vermeden.

Daar richtlijnen en handleidingen niet voldoende zijn om een evaluatieproces tot een goed einde te brengen dient men de nodige aandacht te besteden aan de creatie van een professioneel personeelskader belast met de evaluatieactiviteiten. De belangrijkste bijdrage is de expertise en de bijstand die men kan verlenen aan de verantwoordelijken voor de ontwikkeling en implementatie van evaluatiestudies. De uitbouw van een goed getrainde en functionerende stap is de grootste kritieke succesfactor in de uitbouw van de noodzakelijke evaluatiecapaciteit.

De coördinatie van evaluatiecapaciteit kan ook structureel ondersteund worden. Afhankelijk van de grootte van de organisatie, de soort evaluaties die men uitvoert en opvolgt en specifieke omstandigheden zal men een andere werkwijze volgen. De organisatorische setting kan onder meer volgende constructies aannemen:

- Een centrale eenheid die rechtstreeks instaat voor de coördinatie van de evaluatieactiviteiten over de verschillende departementen of diensten heen.
- Onafhankelijke eenheden of cellen binnen de departementen of diensten.
- Een netwerk met een centrale coördinerende eenheid indirect gelinkt aan eenheden of cellen binnen departementen of eenheden.
- Gedecentraliseerde netwerken en eenheden binnen departementen waarbij het personeel verbonden is aan verschillende beleidsdirectoraten en samenkomen in een ad hoc project- of teamstructuur voor coördinatiedoeleinden.

De eerste twee figuren zijn de polen van een continuüm. De tweede constructie biedt geen aangepaste structuur voor de coördinatie van evaluatiecapaciteit.

Een voorbeeld van de eerste vorm is de Franse 'Conseil national de l'évaluation'. De centrale administraties in Nederland en het Verenigd Koninkrijk werken met eenheden die deel uitmaken van de centrale departementen.

In veel gevallen zal in eerste instantie een centrale eenheid de aansturing van de capaciteitsopbouw op zich nemen. Belangrijke aandachtspunten voor het afbakenen van de rol van deze coördinerende eenheid zijn de vereiste professionele kennis en de accumulatie van evaluatie-expertise.

De kennis waarover een centrale eenheid dient te beschikken is tweërlei: enerzijds voldoende kennis over het te evalueren beleid; anderzijds voldoende professionele kennis en kunde met betrekking tot evaluatiemanagement en -technieken. Tussen beide soorten kennis dient een evenwicht te bestaan, zeker wanneer de centrale eenheid de coördinatie van alle evaluatieactiviteiten op zich dient te nemen. Die aspiratie wordt moeilijker realiseerbaar des te groter de span of control wordt. In grotere organisaties en organisaties met sterke specialisaties binnen afdelingen zal men dan ook sneller moeten evolueren naar een structuur waarbij de coördinatie van de centrale eenheid zich beperkt tot ondersteuning op het vlak van methodische kennis en management van evaluatieprocessen. De eigenlijke planning en uitvoering of opvolging van evaluaties moet dan gebeuren op decentraal niveau.

De coördinerende centrale eenheid evolueert dan naar een rol van coaching. De ontwikkeling van de noodzakelijke expertise en vaardigheden kan op twee manieren worden bewerkstelligd. Een eerste manier is het lanceren en aanbieden van een uitgebreid trainingsprogramma, vanuit een klassieke benadering. Het is evident dat mensen in de eerste plaats dienen te beschikken over een grondige basiskennis om zich de algemene denkkaders en terminologie eigen te maken. De tweede benadering legt de nadruk op de praktijkervaring 'learning by doing'. Dit is wellicht de snelste en meest effectieve manier om mensen op te leiden. De centrale eenheid wordt betrokken bij de uittekening en uitvoering of opvolging van een evaluatie door een decentrale eenheid. Het grote voordeel van deze laatste benadering is de

onmiddellijke betrokkenheid van de participanten bij het evaluatiegebeuren en de eerstelijns ondersteuning wanneer men met moeilijkheden wordt geconfronteerd.

Het spreekt voor zich dat het personeel van de coördinerende eenheid zelf eerst de nodige kennis en vaardigheden moet bezitten om deze coachende rol op te nemen. Tevens dient men aandacht te besteden aan de vervanging van ervaren leden van de centrale eenheid. Voldoende aandacht voor nieuwe medewerkers is een must.

Een belangrijk effect van deze graduele verbreding en verdieping van de evaluatiecapaciteit is dat ook een evaluatiecultuur ingang kan vinden doorheen de organisatie. Door de evaluatiecapaciteit te versterken in functie van de diverse 'lokale' noden, kan de evaluatiepraktijk evolueren in functie verschillende evaluatiemotieven. De meerwaarde van evaluatie ter ondersteuning van beleidsplanning, het management of beleidsleren kan 'ontdekt' worden, naast het verantwoordingsmotief. Het management kan op deze wijze evaluatie aanwenden als een instrument voor een lerende organisatie en ter verbetering van beleidsprocessen. Samenwerking tussen evaluatoren en beleidsmakers is daarbij belangrijk. Door de verspreiding van de evaluatiecapaciteit over de lijnorganisatie creëert men evaluatie eenheden die zeer nauw aansluiten bij de verschillende managementniveaus.

De verbreding van de evaluatiepraktijk leidt tot een mix van evaluatietypes gebaseerd op het tijdsperspectief: ex ante, tussentijdse en ex post evaluaties, om tegemoet te komen aan de diverse evaluatiemotieven. Verschillende evaluatietypes vereisen echter een andere aanpak en deels andere soort informatie. Het komt er dan op aan om te kijken tot op welk niveau de interne evaluatiecapaciteit aan de gedifferentieerde behoeften kan beantwoorden. In veel gevallen zal men een combinatie van interne en externe capaciteit moeten aanspreken, in het bijzonder voor de uitvoering van de evaluaties.

Bekeken op grote schaal, bijvoorbeeld op het niveau van de regeringsbrede administratie, zullen binnen verschillende departementen dezelfde uitdagingen, moeilijkheden en vragen terugkeren rond capaciteitsopbouw. Het is dan van belang om een netwerkstructuur van evaluatoren op te zetten om ervaringen en ideeën uit te wisselen. Buiten de grenzen van de overheid kan dit gebeuren op bredere fora zoals een evaluatieplatform of associatie waar ook de academische wereld, private evaluatie-uitvoerders en middenveldorganisaties ervaringen en ideeën kunnen uitwisselen.

Fase 3: Inbedding van een evaluatiesysteem in de beleidscyclus

In een derde fase van dit groeimodel is de capaciteitsopbouw dermate gevorderd dat het de drager wordt van een evaluatiesysteem waarbij evaluatieprocessen op elkaar zijn afgestemd en volledig ingebed zijn in het beleidsproces. Twee belangrijke elementen komen in deze fase aan bod:

- De versterking van de interne links tussen de verschillende elementen van het systeem
- De afstemming van evaluatie met planning en monitoring

In het kader van een evidence-based beleid dienen de verschillende evaluaties op elkaar afgestemd te worden. Dit vergt een verregaande horizontale coördinatie tussen de verschillende evaluatiecellen of functies op het niveau van de gehele organisatie, bijvoorbeeld tussen de diensten van de gemeentelijke administratie, de beleidsdomeinen van de Vlaamse overheid of de federale overheidsdiensten.

Evaluatie dient ook verder afgestemd worden met de beleidsplanning en monitoring. Monitoring dient in het bijzonder uitgewerkt te worden ten behoeve van beleidsevaluaties en evaluatiemomenten dienen structureel ingebed te worden in de beleidscyclus.

Bronnen

- Ackoff, R. L. (1989). "From Data to Wisdom", *Journal of Applied Systems Analysis*, Vol. 16, pp. 3-9.
- Afdeling Beleidsanalyse, Nederlands Ministerie van Financiën (1989). *Heeft beleid effect? Enkele benaderingen voor de evaluatie van bestaand beleid*. 's Gravenhage: SDU.
- Alkin M.C. (ed.). (2004). *Evaluation roots : tracing theorists' views and influences*. Thousand Oaks, CA: Sage.
- Balthasar A. (2006). "The effects of the institutional design on the utilization of evaluation: Evidenced using qualitative comparative analysis (QCA)" in: *The International Journal of Theory, Research and Practice*, 12 (3).
- Bressers J.Th.A. en Hoogerwerf A. (1991). *Beleidsevaluatie*. Alphen aan den Rijn: Samsom H.D. Tjeenk Willink.
- Chelimsky E. and Shadish W.R. (eds.) (1997). *Evaluation for the 21st Century – A Handbook*. Thousand Oaks CA: Sage.
- Chen H.-T. (2005). *Practical Program Evaluation*, Sage: Thousand Oaks.
- Conings V., Sterck M., Van Dooren W. en Bouckaert G. (2005). *Beleidsnota's versus Beter Bestuurlijk Beleid. Een toetsing van de beleidsnota's aan het Comptabiliteitsdecreet en het Kaderdecreet Bestuurlijk Beleid*. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen, 102 p.
- De Peuter B. en Bouckaert G. (2004). *Een monitoringsysteem voor lokale bestuurskracht : naar de modellering van een Vlaams design*. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen, 111 p.
- Donaldson S.I. (2003). "Theory-Driven Evaluation in the New Millennium". In: Donaldson S.I. and Scriven M. *Evaluation Social Programs and Problems: Visions for the new millennium*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- Donaldson S.I. and Lipsey M.W. (2006). "Roles for theory in contemporary evaluation practice: Developing practical knowledge". In: Shaw I., Greene J. and Mark M. (Eds.) *Handbook of Evaluation*. Thousand Oaks (Ca): Sage.
- European Commission (1999). *The Means Collection – Evaluating socio-economic programmes*. Luxembourg: Office for Official Publications of the European Communities.

- Fetterman D.M. (1997). *Empowerment Evaluation : knowledge and tools for self-assessment and accountability*. Beverly Hills, CA: Sage.
- Fetterman D.M. Wandersman A. and Millet R.A. (2005). *Empowerment Evaluation Principles in Practice*. New York: Guilford.
- Furubo J.-E., Rist R.C. and Sandahl R. (eds.) (2002). *International Atlas of Evaluation*, New Jersey: Transaction Publishers.
- Jacob S. et Varone F. (2003). *Evaluer l'action publique: état des lieux et perspectives en Belgique, Série Modernisation de l'Administration*, Gent: Academia Press, 244 p.
- Mathison S. (ed.) (2005). *Encyclopedia of Evaluation*, Thousand Oaks (Ca): Sage.
- McDavid J.C. & Hawthorn L.R.L. (2005). *Program Evaluation and Performance Measurement, an Introduction to Practice*, Thousand Oaks, CA: Sage.
- Patton M.Q. (1997). *Utilization-focused Evaluation – The New Century Text*. Thousand Oaks, CA: Sage.
- Rieper O. and Toulemonde J. (eds.) (1997). *Politics and Practices of Intergovernmental Evaluation*. London: Transaction Publishers.
- Rossi P.H., Lipsey M.W. and Freeman H.E. (2004). *Evaluation – A Systematic Approach*. 7th edition. Thousand Oaks, CA: Sage.
- Scriven M. (1980). *The Logic of Evaluation*, Inverness CA: Edgepress.
- Scriven M. (1991) *Evaluation Thesaurus*. Thousand Oaks: Sage.
- Stern E. (ed.) (2005). *Evaluation Research Methods. – Volume III*. London: Sage.
- Vlaams Parlement (07-05-2004). *Decreet houdende regeling van de begrotingen, de boekhouding, de controle inzake subsidies en de controle door het Rekenhof* ('Comptabiliteitsdecreet').
- Vlaams Parlement (09-12-2005). *Decreet Provinciedecreet*.
- Vlaams Parlement (15-07-2005). *Decreet Gemeentedecreet*.
- Vlaams Parlement (18-07-2003). *Decreet Kaderdecreet Beter Bestuurlijk Beleid*.
- Vlaamse Regering (03/06/2005). *Besluit van de Vlaamse Regering met betrekking tot de organisatie van de Vlaamse administratie*.

- Weggeman M. (1998). *Kennismanagement: inrichting en besturing van kennisintensieve organisaties*. Schiedam: Lannoo, 221p.
- Wholey J.S., Hatry H.P. and Newcomer K.E. (eds.) (2004). *Handbook of Practical Program Evaluation*. San Francisco, CA: Jossey-Bass.
- Widmer T., Landert C and Bachmann N. (2000). *Evaluation Standards of SEVAL, the Swiss Evaluation Society*. s.l.
- Worthen B.R., Sanders J.R. and Fitzpatrick J.L. (2003). *Program Evaluation – Alternative Approaches and Practical Guidelines*. New York: Addison Wesley Longman.