



Spoor Beleid en Monitoring

HANDLEIDING BELEIDSEVALUATIE

DEEL 2: MONITORING VAN BELEID

Bart DE PEUTER

Joris DE SMEDT

Dr. Wouter VAN DOOREN

Prof. Dr. Geert BOUCKAERT

Rapport
D/2006/10106/014

2007

INHOUDSOPGAVE

1. Inleiding.....	7
2. Monitoring en beleidsevaluatie.....	9
2.1 Monitoring	9
2.1.1 <i>Inleiding: terminologie en definities</i>	9
2.1.1.1 Monitoring: een terminologische keuze	9
2.1.1.2 Definities	11
2.1.2 <i>Motieven voor monitoring in het kader van beheer en beleid</i>	13
2.1.3 <i>Alternatief of component voor evalueren?</i>	17
2.2 Monitoring en beleidsevaluatie	18
2.2.1 <i>Een tandem</i>	18
2.2.2 <i>Verschillen</i>	22
2.2.3 <i>Meten en evalueren van efficiëntie en effectiviteit</i>	27
2.2.3.1 Efficiëntie	27
2.2.3.2 Effectiviteit	30
2.2.4 <i>Uitdagingen</i>	36
3. Monitoring: een algemeen stappenplan.....	43
3.1 Het vertrekpunt: uitgeklaarde motieven en duidelijke beleidsdoelstellingen	43
3.2 Het bepalen van de scope: wat meten we wel, en wat niet?	47
3.2.1 <i>Welke uitsnede van de organisatie zal gemeten worden?</i>	48
3.2.2 <i>Welke uitsnede van de beleidsdoelstellingen en –context zal gemeten worden?</i>	51
3.3 Selecteren en opstellen van indicatoren: hoe meten we en wat willen we meten?	52

3.4 Ontwikkelen van normen: wat is een goed meetresultaat?	64
3.4.1 Soorten normen	64
3.4.2 Waar komen normen vandaan?	66
3.5 Meten.....	69
3.5.1 Betrouwbaarheid van het meetinstrument.....	70
3.5.2 Types van validiteit	71
3.5.2.1 Validiteit van statistische conclusies	72
3.5.2.2 Interne validiteit.....	73
3.5.2.3 Constructvaliditeit	76
3.5.2.4 Externe validiteit	78
3.5.3 Sensitiviteit van indicatoren	79
3.5.4 Meetproblemen.....	81
3.5.4.1 “Meten is niet nodig” – Pangloss stelling	81
3.5.4.2 “Meten is onmogelijk”.....	81
3.5.4.3 “Performance en publieke sector gaan niet samen”	82
3.5.4.4 Convex / concaaf.....	82
3.5.4.5 Output of effect stijgen abnormaal door het meten	82
3.5.4.6 Output of effect dalen abnormaal door het meten.....	83
3.5.4.7 Meer registraties door het meten	83
3.5.4.8 Geen duidelijk onderscheid tussen output en effect	83
3.5.4.9 Inflatie van indicatoren	84
3.5.4.10 Korte termijnfocus.....	85
3.5.4.11 Het waanbeeld.....	85
3.5.4.12 De focus wordt verlegd door het meten	85

3.6 Rapporteren	86
4. Invoering, beheer en gebruik van een monitoringsysteem: enkele kritieke succesfactoren	97
4.1 Het implementatieproces	97
4.2 Het beheer	98
4.3 Gebruik	100
5. Uitbouwsценario voor monitoring: van een beheers- naar een beleidsperspectief	101
5.1 Van output naar effect.....	101
5.2 Monitoren van effecten	102
5.2.1 Moeilijker te meten?	102
5.2.2 Gedifferentieerde effecten	103
5.2.3 De rol van omgevingsfactoren.....	104
5.2.4 Effectenketting.....	104
5.2.5 Wat met neveneffecten?.....	105
5.2.6 Kwalitatieve informatie als complement	106
6. Besluit	107
Bronnen.....	109

Lijst van figuren en tabellen

Figuur 1: De beleids- en beheerscyclus als open systeemmodel	20
Figuur 2: Spreidingdiagram van inputoutput combinaties	28
Figuur 3: Strategieën voor efficiëntieverbetering	28
Figuur 4: Effectniveaus, effectverandering en beleidseffect	34
Figuur 5: De beleids- en beheerscyclus als open systeemmodel	40
Figuur 6: Het conceptueel raamwerk van beleidstheorie	40
Figuur 7: de Pareto analyse	49
Figuur 8: De relatie van een indicator tot het meetobject.....	53
Figuur 9: De betrokkenheid van actoren bij de indicatorenontwikkeling in drie documenten – data uit de agentschappen van het MVG	55
Figuur 10: Overschatting van de afdekkinggraad van een indicator	59
Figuur 11: Onvoldoende aandacht voor moeilijker meetbare aspecten.....	59
Figuur 12: Reductie van het meetobject tot de indicator	60
Figuur 13: Meest gebruikte methoden voor normzetting in de afdelingen van het Ministerie van de Vlaamse Gemeenschap	68
Figuur 14: Types van validiteit doorheen het evaluatieproces	72
Figuur 15: Budgettaire afdekking per beleidsaspect: meetsurplus en -deficit.....	84
Figuur 16: Voorbeeld van een basistabel.....	88
Figuur 17: Voorbeeld van een geavanceerde tabel: PROMES (RVA).....	89
Figuur 18: Voorbeeld van een taartdiagram.....	90
Figuur 19: Voorbeeld van een staafdiagram	91
Figuur 20: Voorbeeld van een stapeldiagram	92
Figuur 21: Voorbeeld van een cartogram.....	93

Figuur 22: Voorbeeld van een lijngrafiek.....	94
Tabel 1: Monitoring en evaluatie: verschillen in focus en gebruik (Government Accountability Office, VS) .	22
Tabel 2: Verschillen tussen evaluatie en monitoring	23
Tabel 3: Fundamenten voor een lerende organisatie.....	37
Tabel 4: Beleidsdoelstellingen getoetst aan het SMART-principe	45
Tabel 5: Readiness checklist: is de organisatie klaar voor monitoring?.....	47
Tabel 6: Vragen voor de afbakening van het meetsysteem	48
Tabel 7: Jobkenmerken, criteria en mogelijke waarden	50
Tabel 8: Technische fiche voor een indicator	61
Tabel 9: Methoden voor normbepaling: voor- en nadelen en voorbeelden uit de Vlaamse praktijk	67
Tabel 10: Elementen voor een technische indicatorfiche.....	99
Tabel 11: Effecttypes op basis van belangrijkheid, intentie en appreciatie.....	105

1. Inleiding

In dit tweede deel van de handleiding gaan we uitgebreid in op het monitoren van beleid. 'Meten is weten'. Monitoring vormt dan ook een potentieel belangrijke informatiebron voor beleidsevaluatie.

In hoofdstuk 2 staan we stil bij definities van monitoring en de motieven om beleid via indicatoren op te volgen. Vervolgens bekijken we de onderlinge relatie tussen monitoring en evaluatie, hun gelijkenissen en verschilpunten. We gaan dieper in op het meten en evalueren van twee belangrijke criteria: de efficiëntie en effectiviteit van het beleid.

In hoofdstuk 3 presenteren we een algemeen stappenplan om een monitoringsysteem te ontwikkelen en te introduceren. Wat is het vertrekpunt? Hoe bepalen we de inhoud van het meetsysteem? Hoe ontwikkelen en selecteren we goede indicatoren? Welke normen kunnen we hanteren om de meetwaarden op de indicatoren te beoordelen? Vervolgens gaan we in op het meten zelf. Wat zijn de kenmerken van een goede meting? Welke problemen kunnen er zich voordoen bij het meten, waar moeten we op letten bij de interpretatie van metingen?

Hoofdstuk 4 lijst een aantal kritieke slaagfactoren op met betrekking tot de invoering, het beheer en het gebruik van een monitoringsysteem.

In hoofdstuk 5 schetsen we een scenario voor de uitbouw van bestaande meetsystemen. Vaak werden deze in de praktijk ontwikkeld vanuit een managementperspectief. De vraag die hier centraal staat is: Hoe kan de scope verbreed worden naar een beleidsperspectief? Welke mogelijkheden en beperkingen brengt dit met zich mee?

2. Monitoring en beleidsevaluatie

2.1 Monitoring

In deze paragraaf beantwoorden we enkele basisvragen inzake monitoring. Wat is en betekent monitoring in een overheidscontext? Wat zijn in deze context de motieven voor monitoring? Welke evoluties en trends zijn er waarneembaar in de praktijk? En hoe verhoudt monitoring zich ten opzichte van beleidsevaluatie? Het antwoord op deze laatste vraag wordt verder uitgediept in een volgende paragraaf.

2.1.1 Inleiding: terminologie en definities

Wat is en betekent monitoring in een overheidscontext? Het antwoord op deze vraag ligt in de ruime waaier aan definities en omschrijvingen die men in de literatuur aantreft. Daaruit zullen we de voornaamste kenmerken van monitoring afleiden. Eerst staan we even stil bij de keuze voor de term 'monitoring' doorheen deze handleiding.

2.1.1.1 Monitoring: een terminologische keuze

Zowel in de Engelstalige als Nederlandstalige literatuur treffen we verschillende, op elkaar gelijkende benamingen en termen aan: *performance measurement*, *monitoring*, *performance monitoring*, *prestatiemeting*. Verschillende auteurs gebruiken deze termen door elkaar. Heel wat auteurs houden consequent eenzelfde benaming aan; uit gewoonte of omdat ze bij bepaalde van deze woorden een semantische nuance of connotatie onderkennen die ze willen beklemtonen of net vermijden.

Enerzijds verwijzen al deze termen naar een gemeenschappelijke kern van een praktijk, samen te vatten met het werkwoord 'meten'. In dat opzicht zijn de vele termen in de beleids- en bestuurskundige literatuur elkaars synoniem. Anderzijds zijn we van oordeel dat bepaalde van deze woorden inderdaad semantische verschillen meedragen. In het Nederlands geldt dit bijvoorbeeld voor de termen 'prestatiemeting' en 'monitoring'.

De oorsprong van het verschil in semantiek is voor een groot deel terug te brengen naar twee grote aandachtsgebieden in de bestuurskundige literatuur: management en beleid. Een screening van deze literatuur leert dat de term prestatiemeting sterk is gelieerd aan een managementcontext, terwijl de term monitoring overwegend gehanteerd wordt in een beleidscontext.

Doorheen deze handleiding hanteren wij consequent de term monitoring. Deze terminologische keuze hangt samen met enkele andere fundamentele vertrekpunten achter deze handleiding: een integrale benadering van de beleids- en beheerscyclus van de overheid, de alternatieve

visie dat evaluatie doorheen de ganse beleidscyclus en dus niet enkel in de laatste fase een meerwaarde kan bieden, en het standpunt dat evaluatie en monitoring complementair zijn om het beleid te informeren en te ondersteunen.

Het evalueren van beleid staat centraal in deze handleiding. Zoals we al uiteengezet hebben in deel 1, kan men uit de motieven om beleid te evalueren afleiden dat evaluatie in de verschillende fasen van het klassieke schema van de beleidscyclus een meerwaarde kan leveren: ex ante bij de planning, tussentijds bij de uitvoering en ex post bij de beoordeling van het beleid.

De beleidscyclus hangt samen met de beheerscyclus. De beheerscyclus wordt in het bijzonder doorlopen tijdens de uitvoeringsfase van het beleid: de overheid wendt middelen aan in processen die in een bepaalde output resulteren. Deze cyclus valt de facto binnen de organisationele grenzen van de overheid en vormt de focus van het management. In deze optiek wordt de managementcontext en bijhorende beheerscyclus binnen het bredere kader van de beleidscontext en -cyclus gepositioneerd. Beleid zonder management kan niet slagen, management zonder beleidskader is niet functioneel.

Gegeven deze evaluatiecontext, is voor het meten in het kader van beleidsevaluatie de term 'monitoring' adequater dan 'prestatiemeting'. Volgende, samenhangende kenmerken van de term monitoring liggen aan de basis van deze woordkeuze.

(1) Semantische neutraliteit

Monitoring is een neutralere term dan prestatiemeting. Dit heeft te maken met de herkomst van deze laatste term: het is een vertaling van de Engelse uitdrukking 'performance measurement', dat als praktijk in navolging van de privé-sector ingang vond bij de overheid. Analoog met de praktijk in de privé-sector werd ook binnen de overheid prestatiemeting aan een managementperspectief verbonden. De klemtoon werd gelegd op het meten van input, processen en output; vanuit analytisch oogpunt inderdaad de componenten van de beheerscyclus. Daarenboven heeft men met de Nederlandse termen voor de componenten uit de beheerscyclus ongewild enige verwarring gecreëerd: het Engelstalige schema van "input - process - output" werd immers vertaald naar "middelen - activiteiten - prestaties". Men kan hieruit verkeerdelijk besluiten dat prestatiemeting beperkt is tot het meten van output. Anderzijds werd de associatie met een managementcontext versterkt.

Een belangrijk verschilpunt tussen de publieke en private sector is dat de overheid bepaalde effecten in de maatschappelijke omgeving, d.w.z. buiten de organisatiegrens, wil bereiken of beïnvloeden. Recentelijk zwaaide de pendel van de aandacht bij de overheid meer in de richting van het meten van effecten en omgevingsfactoren. In bepaalde gevallen te ver zelfs waarbij de componenten van de beheerscyclus niet meer opgevolgd werden.

Dat laatste levert echter evenzeer een te eenzijdig beeld op als wanneer men effecten niet meet.

In elk geval bleef de term prestatiemeting sterk geassocieerd met een managementcontext. De verklaring hiervoor is dat naast de inhoudelijke focus ook het verantwoordingsmotief voor prestatiemeting werd getransponeerd van de private sector naar de overheid. Daarmee heeft de term prestatiemeting voor velen ook de connotatie van verantwoording en afrekening gekregen. Dit is niet het geval voor de term monitoring, die bijgevolg adequater is om te hanteren in een bredere beleidscontext.

(2) Brede afdekkingsgraad

Doordat de term monitoring semantisch neutraal is, kan deze een breder spectrum van elementen afdekken: zowel de componenten van de beheerscyclus als complementaire aspecten vanuit een beleidsperspectief. Anderzijds kan evaluatie doorheen de beleidscyclus een ondersteunende rol spelen, vanuit meer motieven dan enkel verantwoording. Deze twee factoren leiden tot de keuze voor monitoring in de context van beleidsevaluatie.

(3) Consistentie met de bestuurskundige vakliteratuur

Een laatste argument dat in de terminologische keuze meespeelt, is dat we met de term monitoring ons aansluiten bij het prevalerende jargon in de hedendaagse evaluatieliteratuur. De meeste vooraanstaande auteurs in het theorieveld van beleidsevaluatie hanteren de term monitoring. Ook internationale organisaties zoals de Europese Unie en de Wereldbank zitten op deze lijn.

2.1.1.2 Definities

Wat is monitoring in het kader van beleid en beheer? In de vakliteratuur rond evaluatie en monitoring treffen we een breed scala aan omschrijvingen en definities aan. Vaak wordt in deze omschrijvingen geduid op het bestaan van verschillende soorten monitoring. Een greep uit het rijke aanbod.

“Performance monitoring involves using indicators to follow the implementation process and outcomes of a program across time. Program monitoring is helpful as a provider of the kind of basic information to which nearly all programs wish to have access: a set of vital statistics concerning the program implementation and outcomes. Process monitoring is the periodic collection of implementation information. Outcome monitoring is the periodic collection of information on outcomes of a program.” (Chen, 2005: 54, 181, 182 and 184).

“Performance monitoring can be viewed as periodically measuring progress toward explicit short-, intermediate, and long-term results. It can also provide feedback on the progress made (or not made) to decision-makers.” (Kusek & Rist in Mathison, 2005: 302).

“Monitoring is a continuous function that uses the systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds.” (OECD 2002, aangehaald in Kusek & Rist, 2004: 12).

“Program process monitoring is the systematic and continual documentation of key aspects of program performance that assesses whether a program is operating as intended or according to some appropriate standard, whereas outcome monitoring is the continual measurement of intended outcomes of the program, usually of the social conditions it is intended to improve. Organizational functions monitoring focuses on whether the program is performing well in managing its efforts and using its resources to accomplish its essential tasks.” (Rossi, Lipsey & Freeman, 2004: 171 en 191)

“Performance measures are quantitative indicators of various aspects of the performance of public or nonprofit programs, agencies or other entities that can be observed on a regular basis. Most often they focus on programs or service delivery, in which case they tend to relate most closely with evaluation. However, they are also used to track the performance of agencies or organizations themselves, perhaps focusing on a number of programs as well as other aspects of organizational performance, such as employee development and control over administrative overhead”. (Poister in: Wholey, Hatry & Newcomer, 2004: 98)

“Monitoring requires the routine measurement and reporting of important indicators. (...) Outcome monitoring is outcome-focused or results-oriented; it is built into the routines of data reporting within program operations; it provides frequent and public feedback on performance; other forms of monitoring measure important facets of program operations, especially when a new program or a demonstration project is being

implemented. This may be on inputs, on process or on outputs." (Affholter in: Wholey, Hatry & Newcomer, 1994: 97).

Uit bovenstaande definities komen enkele hoofdkenmerken van monitoring naar voor, enerzijds met betrekking tot wat het is en anderzijds met betrekking op welke manier het gebeurt.

Monitoring is het verzamelen van informatie in de context van het beleid en het beheer, over geselecteerde aspecten of factoren. Uit de bewoordingen blijkt dat monitoring beschrijvende informatie oplevert. Het stelt ons in staat om de werkelijkheid op te volgen en veranderingen in kaart te brengen. Monitoring kan de functie hebben van een 'early warning system'.

Monitoring als proces gebeurt op een systematische manier en is doorlopend aan de hand van periodieke metingen of registraties. De informatie is dus ingebed in een recurrente stroom van gegevens. Door het permanente karakter kan dan ook een systeem van monitoring worden opgezet. Een monitoringsysteem is een geheel van procedures en kanalen om gegevens te verzamelen, op te slaan, te verwerken en te rapporteren. Dergelijk systeem omvat dus één of meerdere databestanden en/of databanken waarin gegevens worden bewaard die verzameld werden doorheen de tijd. De elementaire deeltjes van dergelijk meetsysteem zijn indicatoren. Een indicator geeft een bepaald verschijnsel of aspect weer. Vaak is het een afgeleide of vertaling of benadering daarvan. Bijvoorbeeld, het aantal verkeersongevallen geeft aan hoe (on)veilig het verkeer is. Het nitraatgehalte zegt iets over de waterkwaliteit. Een indicator is een maatstaf. In hoofdstuk 3 gaan we verder in op de ontwikkeling van indicatoren.

Uit verschillende van bovenstaande definities kunnen we ook afleiden dat monitoring gekoppeld wordt aan doelstellingen, die betrekking hebben op processen en/of resultaten. Sommige bronnen (bv. Wereldbank) wijzen op een historische shift in de focus van monitoring: van processen naar resultaten. Hoewel verschillende auteurs het eens zijn dat dergelijke inhoudelijke klemtonen van monitoring bestaan, wordt een universele verandering in focus niet door allen beaamd; vaak is de focus organisatie- of sectorgebonden.

2.1.2 Motieven voor monitoring in het kader van beheer en beleid

Het meten van processen en resultaten in de publieke sector kreeg de laatste decennia steeds meer aandacht. Dit hangt samen met veranderde verwachtingen ten aanzien van de rol en verantwoordelijkheid van de overheid, en van beleidsmakers en managers in organisaties die tot de publieke en non-profit sector behoren.

Die verwachtingen hangen op hun beurt samen met een aantal brede maatschappelijke trends of bronnen van waaruit een zekere druk werd gecreëerd waardoor 'meten' belangrijker werd. Voorbeelden van deze trends zijn: een doorgezet democratiseringsproces waardoor de burger mondiger werd en diens verwachtingen vis-à-vis de overheid meer ging expliciteren in een tegensprekelijk debat met de beleidsverantwoordelijken, technologische modernisering (bv.

in telecommunicatie waardoor de wereld steeds meer een 'dorp' werd), de globalisering van economie en handel, en een toegenomen complexiteit en verwevenheid van beleidsissues en proliferatie van organisaties om deze beleidsissues aan te pakken.

Onder impuls van deze trends leven we vandaag in een kennissamenleving; uiteenlopende maatschappelijke onderwerpen en issues waarmee het openbaar bestuur wordt geconfronteerd, alsook ervaringen en organisaties, zijn in toenemende mate kennisintensief geworden. Het kennisaspect als zodanig is niet nieuw, wel is de maatschappelijke rol en het belang ervan danig veranderd. Daarmee werd ook openbaar bestuur meer en meer kennisintensief. Tevens steeg de aandacht voor een aangepaste capaciteitsopbouw om te kunnen inspelen op de veranderde uitdagingen en verwachtingen.

Het meten in de overheid is echter geen recent ontstane praktijk; de wortels ervan liggen (veel) verder terug in de tijd. De facto is de oorsprong van de aandacht en praktijk te traceren bij de lokale besturen in de Verenigde Staten, bij de start van de 20^{ste} eeuw. In de jaren '60 en '70 werd het een aandachtspunt bij grootschalige hervormingspakketten in de overheidssector in de VS. Een voorbeeld daarvan was de ontwikkeling van het systeem van planning, programmering en budgettering (PPBS), waarmee men de koppeling van beleidsresultaten met de begroting van financiële middelen beoogde. Hoewel niet elke moderniseringsbeweging altijd de verwachte resultaten opleverde, vaak door implementatieproblemen, overleefde het idee van meten doorheen deze hervormingen. In de jaren '90 werd het terug een centraal aandachtspunt in brede hervormingsbewegingen onder de vlag van het 'New Public Management'. In deel 1 wezen we al op het belang van deze moderniseringstrends voor de verspreiding van een evaluatiecultuur in tal van Europese landen.

De permanente of terugkerende aandacht over een langere periode voor het meten van processen en resultaten in de overheid als een haalbare en nuttige praktijk, heeft te maken met de wijze waarop deze overheid voorgesteld wordt. In de loop der jaren werden immers enkele metaforen ontwikkeld die het idee van meten in een overheidscontext ondersteund en gestaafd hebben. We lichten kort de meest gehanteerde metaforen even toe.

Metafoor van de machine

Deze metafoor is ontstaan als deel van de intuïtieve en visuele basis voor monitoring. Een set indicatoren voor een organisatie wordt dan als het equivalent beschouwd van wat een instrumentenpaneel of dashboard voor een machine is. Accurate indicatoren verschaffen de gebruiker in een oogwenk informatie over de werking van de overheid naar analogie met de meetinstrumenten in een vliegtuig of wagen. Het Franse "Tableau de Bord" en de "Balanced Score Card" zijn gebaseerd op deze metafoor.

Wanneer een indicatorenset het dashboard is, dan is de overheidsdienst een machine die bestaat uit complexe maar begrijpbare systemen en subsystemen die onderling verbonden zijn en die men kan monitoren. Op basis van de informatie kunnen managers en andere stakeholders de organisatie sturen.

Metafoor van het bedrijf

De overheid als een bedrijf dat gerund moet worden: deze metafoor begeleidde verschillende pogingen om managementpraktijken uit de private sector te transponeren naar de publieke sector. De klemtoon daarbij ligt op het belang van: duidelijke doelstellingen, programma's die gepland en beheerd worden om deze doelstellingen te bereiken, efficiëntie (met inbegrip van positieve effecten die men verbindt aan concurrentie, privatisering en uitbesteding), en, uiteindelijk, aandacht voor de bottom-line: winst of verlies. Indicatoren zijn vanuit dat perspectief een kernonderdeel van een managementfilosofie, die resultaten beklemtoont en resultaatgericht management stimuleert.

Metafoor van het open systeem

Deze metafoor is verwant met een biologische metafoor die organisaties vergelijkt met organismen. Deze laatste staan in interactie met hun omgeving als open systemen. Ze hebben structuren die bepaalde functies vervullen, die op hun beurt bijdragen tot een doel van homeostase of zelfregulering: de eigenschap om in een evenwichtige toestand te blijven in relatie tot fluctuaties in de omgeving. Om zich te handhaven werken biologische organismen met parameters, bv. temperatuur. Veranderingen boven of onder een normale marge duiden erop dat het organisme zich in een ongunstige toestand bevindt, die niet te lang mag aanhouden op het gevaar van (blijvende) schade.

De biologische metafoor wordt zelden letterlijk op organisaties toegepast, maar beïnvloedde wel het denken over monitoring. Men veronderstelt dat men een goed begrip heeft van de oorzaak en gevolg verbanden in programma's van beleid en beheer en de rol daarin van organisaties, zodat indicatoren kunnen aangeven hoe goed de organisatie het stelt. Accurate maatstaven zullen dan een krachtig instrument zijn om complexe organisaties te monitoren en op hun werking te beoordelen. Gemeenschappelijk met de biologische metafoor is dat men de overheid ziet als een open systeem dat verbonden is met de maatschappelijke omgeving.

Het voordeel van dergelijke metaforen is dat ze kunnen bijdragen aan een verrijkt inzicht in situaties en aan het naar voor brengen van een eigen methodiek voor het omgaan met bepaalde aandachtspunten en problemen. Alternatieve perspectieven kunnen het begrijpen versterken en acties suggereren die voorheen onmogelijk werden beschouwd of ongekend waren. Voor de relevantie en correctheid van een metafoor is het essentieel dat de kernelementen of sleutelmechanismen kunnen worden overgedragen op de organisatie of het beleid waarvoor indicatoren worden opgesteld.

Welke zijn nu de concrete aanleidingen of motieven voor monitoring? Dit is meteen een vraag naar de verschillende gebruikscontexten. Het gaat om een trits van motieven die eveneens motieven zijn om beleid te evalueren.

a) Ondersteuning van plan- en beslisprocessen

Monitoring kan een ondersteunde functie vervullen ten aanzien plan- en beslisprocessen. In het kader van beleidsplanning waarin strategische en operationele doelstellingen worden bepaald, kan monitoring specifieke informatie aanbrengen, die wordt verzameld over de problematiek die men wil aanpakken en/of over de doelgroep waarop men zich wil richten. Deze informatie kan mee beslissingen en keuzes onderbouwen met betrekking tot prioriteiten, doelstellingen, alternatieve maatregelen en beleidsinstrumenten.

In de beleids- en beheerscyclus hangen beslissingen over beleidsmaatregelen en financiële middelenallocatie idealiter nauw met elkaar samen. Ook in het kader van financiële planning op korte, middellange of lange termijn kan monitoring een ondersteunende rol spelen. Financiële indicatoren kunnen in combinatie met niet-financiële informatie het maken van keuzes in de besteding en aanwending van middelen onderbouwen en begeleiden.

b) Opvolging en verbetering van uitvoeringsprocessen en organisaties

De uitvoeringsfase van de beleidscyclus is het sterkste verbonden met de beheerscyclus, waaronder de financiële cyclus (begroten, boekhouden, audit) en contractcycli (beheersovereenkomsten) geplaatst kunnen worden. Monitoring kan dan ook zowel betrekking hebben op het beleid als in een managementcontext georganiseerd worden.

Monitoring via managementinformatiesystemen (MIS) stelt de managementverantwoordelijken in staat om de werking van de organisatie op te volgen, knelpunten te detecteren en gericht bij te sturen. Op deze wijze kan de efficiëntie van processen verbeterd worden.

Eenzijds kan een MIS betrekking hebben op interne managementaspecten, zoals personeel en infrastructuur. De allocatie van financiële middelen, de besteding en de verifiëring daarvan kan aan de hand van financiële indicatoren opgevolgd worden. Dit verschaft informatie over de capaciteit van de organisatie en hoe deze aangewend wordt.

Anderzijds kan een MIS een externe focus hebben: de opdracht van de organisatie. De taakstelling is vervat in de operationele doelstellingen. Deze kaderen in een breder beleidskader. De operationele doelstellingen ressorteren onder strategische doelstellingen die verwijzen naar de effecten die organisatie wil bewerkstelligen.

Een overheidsorganisatie zal vaak meerdere managementinformatiesystemen opzetten, waarbij zowel het interne als externe perspectief aan bod komt. Een resultaatgericht management streeft ernaar om de middelen van de organisatie optimaal te alloceren aan die processen die men het meest functioneel acht voor het behalen van de gewenste resultaten, die in de operationele en strategische doelstellingen vervat liggen.

c) Verantwoording

Informatie uit monitoring kan ook aangewend worden ten behoeve van sturing- en verantwoordingsrelaties. In een managementcontext zal de verantwoording gerelateerd zijn aan de (besteding van) middelen, (het verloop van) de processen en (de realisatie en kwaliteit van) de output van de organisatie. Sturing- en verantwoordingsrelaties kunnen zowel betrekking hebben op individuen in de organisatie (bv. mandaatcontract) als op een organisatie als geheel (bv. beheersovereenkomst tussen departement en agentschap).

Verantwoording kan echter ook betrekking hebben op de effecten van beleid: in welke mate werden de effecten gerealiseerd door (de uitvoering van) het beleid? Monitoring van effecten en omgevingsvariabelen kunnen dergelijke verantwoording ondersteunen. Monitoring van effecten en omgevingsindicatoren kan gedeeltelijk rechtstreeks gekoppeld worden aan de werking van de overheidsdiensten. Zo kan bijvoorbeeld tijdens het proces van dienstverlening informatie verzameld worden over de doelgroep van het beleid.

Monitoring kan dus zowel in het kader van management als beleid opgezet worden vanuit een verantwoordingsmotief.

2.1.3 Alternatief of component voor evalueren?

Uit de voorgaande paragrafen blijkt de functies en motieven met betrekking tot monitoring voor een deel parallel lopen of gemeenschappelijk zijn met deze van beleidsevaluatie. De vraag stelt zich dan ook of monitoring te beschouwen is als een alternatief voor evaluatie, dan wel een complementaire methode is met het oog op de aangehaalde motieven. Deze handleiding en in

het bijzonder deel 2 stoelt op het uitgangspunt dat monitoring en evalueren inderdaad complementaire methoden zijn om beleid en beheer te ondersteunen en te informeren. Vanuit het perspectief van de evaluator vormt monitoring een potentieel belangrijke informatiebron. In de volgende paragraaf gaan we dieper in op het antwoord op deze vraagstelling.

2.2 Monitoring en beleidsevaluatie

2.2.1 Een tandem

Zowel monitoring als beleidsevaluatie kregen het laatste decennium versterkte aandacht bij de overheid, onder impuls van eerder vermelde veranderingen en trends. Er werden heel wat meetsystemen ontwikkeld die betrekking hebben op de werking en de resultaten van overheidsorganisaties. Dat gebeurde met vallen en opstaan. Aan het meten werden heel wat voordelen en verwachtingen verbonden. Het creëren van transparantie, het bevorderen van resultaatsgericht werken zijn maar een paar voorbeelden. Maar wellicht nog meer werd er ook gewezen op de beperkingen van meten in een overheidscontext. Onder meer Mintzberg, Deming, Handy en Perrin stellen dat het reduceren van een publiek optreden tot een aantal kwantificeerbare indicatoren in vele gevallen weinig zin heeft. Mintzberg spreekt in dit kader van 'de meetmythe'. Het analyseren van overheidsoptreden kan vaak niet (alleen) aan de hand van harde, kwantificeerbare metingen. Negeren dat bepaalde aspecten moeilijk meetbaar zijn is misleidend. De auteurs stellen daartegenover dat kwalitatieve werkwijzen in dergelijke gevallen meer aangewezen zijn.

De meerwaarde van het meten in een overheidscontext is dan ook herhaaldelijk het voorwerp van debat tussen voor- en tegenstanders van kwantitatieve en kwalitatieve benaderingen van het overheidsoptreden. Dat debat is een universeel debat in de sociale wetenschappen en drong dan ook door in het theorie- en praktijkveld van beleidsevaluatie. De afgelopen jaren groeide bijgevolg de aandacht voor de vraag hoe monitoring en beleidsevaluatie zich onderling verhouden. Het mag de lezer niet verwonderen dat – tegen de achtergrond van de geschetste debatten- er zeer uiteenlopende ideeën terug te vinden over het samengaan van beleidsevaluatie en monitoring.

Samengevat zijn de twee extreme stellingen:

- "beleidsevaluatie heeft monitoring niet nodig"
- "monitoring volstaat voor het evalueren van beleid"

Sommige auteurs beschouwen monitoring enkel en alleen als een instrument voor het management dat niet nuttig is voor het evalueren van beleid. Ze stellen dat evaluatie, in tegenstelling tot monitoring, een grote kennis van verschillende methoden en technieken vergt. Een tweede argument heeft betrekking op de vaststelling dat managers vaak een sleutelrol vervullen in de ontwikkeling en implementatie van een monitoringsysteem, opdat dit systeem deel uitmaakt van de organisatiestructuur en management. Met name in het kader van externe beleidsevaluaties zien deze auteurs daarin een knelpunt.

Andere auteurs hechten een zeer groot belang aan monitoring en gaan er hierbij van uit dat kwantitatieve gegevens volstaan om ook het beleid te evalueren.

Een grote groep van auteurs (bv. Chelimsky and Shadish: 1997; McDavid and Hawthorn: 2006; Rossi & Freeman, 2004; Wholey, Hatry and Newcomer: 2004), neemt echter een tussenpositie in. In hun visie maakt monitoring dan systematisch deel uit van een evaluatieproces. In deze handleiding volgen we de stelling dat beide instrumenten, evaluatie en monitoring complementair functioneren. Deze integrale benadering zet zich ook algemeen door in de theorie en praktijk. Overheden besteden steeds meer aandacht aan de effectiviteit van beleid in het kader van verantwoording en beleidsleren.

Ondanks hun grote complementariteit blijven monitoring en evaluatie twee aparte processen met elk hun eigen setting, problemen en uitdagingen. Beide processen vertrekken vanuit dezelfde motieven, maar met andere aspiraties en vraagstellingen. In een volgende paragraaf gaan we verder in op de verschillen. Eerst bekijken we de complementariteit aan de hand van een metafoer van de tandem. Evaluatie en monitoring kunnen inderdaad gezien worden als een tandem. In deze tandem zal het evaluatieproces functioneren als de stuurinrichting. Beleidsevaluatie stuurt het hele proces; het bepaalt de informatiebehoefte. Aan de hand van het evaluatiedesign en de hierin gestelde evaluatievragen bepaalt het evaluatieproces welke informatie verzameld moet worden.

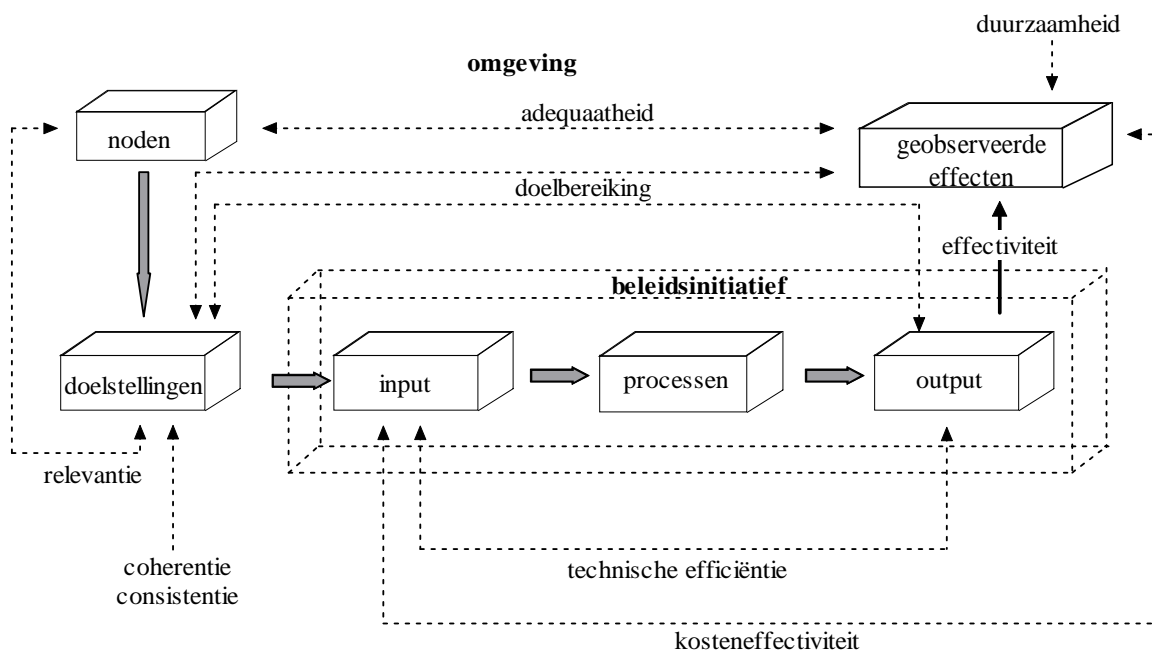
De aanwezigheid van een monitoringsysteem kan in belangrijke mate de aandrijving van het evaluatieproces versterken als informatiebron. In dat opzicht kan monitoring een belangrijke bijdrage leveren aan het welslagen van evaluaties.

Monitoring kan evenwel niet voor elk type van evaluatie een even sterke bijdrage leveren. Met name voor ex ante evaluaties zal monitoring een beperkte rol kunnen spelen. Mogelijk worden

er systematisch gegevens verzameld over de doelgroep van het beleid. Echter informatie over beleidsalternatieven zal zelden aangereikt kunnen worden via monitoring.

Mogelijk kan het evaluatieteam wel beroep doen op gegevens verzameld in eerdere beleidsprocessen. In tussentijdse evaluaties vervult monitoring een belangrijke functie. Tussentijdse evaluaties hebben baat bij de aanwezigheid van een goed uitgebouwd monitoringsysteem. Zeker in het kader van nieuw opgestarte beleidsinitiatieven kunnen monitoringsystemen een belangrijke rol vervullen in het leveren van feedback ter versterking en bijsturing van het implementatieproces. Voor ex post evaluaties kan monitoring een vergelijkbare functie vervullen.

De complementariteit van monitoring en evaluatie kan aangetoond worden aan de hand van de onderstaande figuur.



Figuur 1: De beleids- en beheerscyclus als open systeemmodel

Monitoring is een manier om op een duidelijke en systematische manier gegevens te verzamelen en te rapporteren. Monitoring kan informatie aanleveren over de componenten in onderstaand schema: de noden met inbegrip van kenmerken van de doelgroep, de input of middelen, processen, de output en effecten. Ook over omgevingsfactoren kunnen indicatoren informatie aanreiken.

Monitoring is functioneel voor het beantwoorden van de 'wat?-vragen'; wat gebeurt er en wat heeft zich voorgedaan? Deze vragen zijn essentieel, daar ze een eerste inzicht schenken in de

processen en resultaten die zich afspelen in een beleidsproces. Monitoring levert beschrijvende informatie op. Uit het overzicht in boekdeel 1 over gegevensverzameling in het kader van een evaluatie, is gebleken dat monitoring één van de informatiebronnen kan zijn voor een evaluatie. De evaluator zal meestal aangewezen zijn op een combinatie van primaire en secundaire bronnen. Een monitoringsysteem is een secundaire bron van informatie: het meetsysteem bestaat reeds en werd niet voor een specifieke evaluatie ontworpen.

Evaluatie is functioneel voor het beantwoorden van de 'hoe?-vragen' en de 'waarom?-vragen'; hoe en waarom gebeurt er wat of heeft er zich iets voorgedaan? Deze vragen hebben andere aspiraties: ze zoeken naar verbanden en verklaringen, en verbeterpunten. Een monitoringsysteem kan onmogelijk deze vragen beantwoorden, een evaluatie echter beschikt over geëigende technieken en aanpak om een antwoord te formuleren op deze vraag. Een beleid kan falen in het behalen van de vooropgestelde doelstellingen omdat de achterliggende beleidslogica problematisch is. Dit falen kan echter een gevolg zijn van zeer uiteenlopende oorzaken; onduidelijke doelstellingen of het niet identificeren van andere mogelijke effecten, verkeerd management en implementatieprocessen, onder –of overfinanciering, het juiste beleid in de verkeerde setting omwille van de unieke omstandigheden, te vroeg meten van effecten, meten op het verkeerde ogenblik of op basis van verkeerde assumpties. De correcte verklaring is een antwoord op een hoe?- of waarom?-vraag.

Een voorbeeld zal het onderscheid tussen beide processen verduidelijken. Beeld je in, je hebt net een nieuwe wagen gekocht. Zowel de verkoper als de handleiding stellen dat je wagen 100 kilometer kan rijden per 8 liter diesel. Momenteel heb je de wagen zes maanden en je hebt minutieus het verbruik en de kilometerstand bijgehouden. Nu stel je vast dat je wagen gemiddeld 10 liter per 100 kilometer verbruikt. Wat kan je doen? Je gaat terug naar de garage met de vraag of een mecanicien de wagen kan nakijken om te achterhalen wat er fout loopt.

In dit voorbeeld verzamelt men via monitoring beschrijvende informatie over het verbruik en de kilometerstand (de 'wat?-vraag'). Monitoring maakt de vergelijking tussen het vooropgestelde en het eigenlijke verbruik mogelijk. Evaluatie heeft betrekking op de diagnose van de mecanicien, die concreet zal nagaan wat de oorzaak is van het meerverbruik (de waarom?-vraag). Het voorbeeld maakt duidelijk dat monitoring een bijdrage levert aan de uitvoering van evaluatieprocessen. Zonder het bijhouden van het verbruik en de kilometerstand beschikken we niet over data die aantonen dat in dit voorbeeld de wagen niet functioneert zoals het hoort. De verzamelde data vormen een vertrekpunt voor de mecanicien om de verklaring voor het meerverbruik te achterhalen. Ook op maatregel- of beleidsniveau levert monitoring data, eventueel aangevuld met bijkomende data, die het evaluatieteam in staat stelt de evaluatie op een correcte manier uit te voeren.

2.2.2 Verschillen

Ondanks de complementariteit tussen beide processen, zijn er een aantal belangrijke verschilpunten tussen evaluatie en monitoring te onderkennen. Het Government Accountability Office in de VS benadrukt de verschillen in focus en gebruik.

Een verschillende focus

<ul style="list-style-type: none"> • Monitoring - Monitoring richt zich op doelbereiking uitgedrukt in meetbare indicatoren 	<ul style="list-style-type: none"> • Evaluatie - Evaluatie onderzoekt een bredere waaier aan informatie. Afhankelijk van de focus onderzoeken evaluaties bepaalde onderdelen van het beleidsproces, factoren die zich bevinden in de context van het beleid, of proberen ze bepaalde verbanden aan te geven en de causaliteit te verklaren tussen input, processen, output en effecten. - Evaluaties kunnen de netto impact van een beleid achterhalen; wat als het beleid er niet was geweest? Ze kunnen verschillen aanduiden tussen mogelijke opties.
---	---

Een verschillend gebruik

<ul style="list-style-type: none"> • Monitoring - Monitoring ondersteunt de bronnenallocatie en andere beleidsbeslissingen ter verbetering van de dienstverlening en de effectiviteit. - Monitoring is door z'n permanent karakter bruikbaar voor het management en als instrument om de accountability van het overheidsoptreden te verhogen. 	<ul style="list-style-type: none"> • Evaluatie - Evaluatie ondersteunt de bronnenallocatie en andere beleidsbeslissingen ter verbetering van de dienstverlening en de effectiviteit. - Evaluatie heeft een meer dieptebenadering in een bredere context. Dit laat een algemeen assessment toe dat een bepaald beleid of maatregel functioneert of niet.
---	--

Tabel 1: Monitoring en evaluatie: verschillen in focus en gebruik (Government Accountability Office, VS)

Andere belangrijke verschilpunten werden aangegeven door McDavid en Hawthorn (2006) en hebben betrekking op het tijdsperspectief, de scope, de ontwikkelingscontext, de omgang met causaliteit, de bronnen, verantwoordelijken en flexibiliteit in doelstellingen. De onderstaande tabel vat de verschilpunten samen. Ze worden erna verder toegelicht.

Evaluatie	Monitoring
1. Ad hoc	Permanent
2. Onderwerpspecifiek	Ontwikkeld en opgebouwd vanuit meer algemene overwegingen. Eenmaal geïmplementeerd, geschikt voor breder gamma van vragen en kwesties waarop het ontwerp gebaseerd is.
3. Maatstaven grotendeels ontwikkeld in functie van de evaluatie	Maatstaven worden ontwikkeld en data verzameld in routineprocessen
4. Attributie van geobserveerde effecten is meestal sleutelvraag	Attributie van geobserveerde effectveranderingen aan het beleid wordt zonder meer verondersteld.
5. Gerichte bronnen vereist, o.m. monitoring	Bronnen zijn vaak verankerd in organisatiestructuur en managementprocessen
6. De evaluatoren zijn vaak niet de managers	Managers spelen vaak belangrijke rol bij ontwikkeling van het monitoringsysteem en de rapportering
7. Opzet van de evaluatie is vooraf gekend / onderhandeld.	Gebruik kan sneller wijzigen door veranderde informatienoden

Tabel 2: Verschillen tussen evaluatie en monitoring

1. Het tijdsperspectief: ad hoc vs. permanent

Een evaluatie is gekenmerkt door een afgelijnd tijdsperspectief. Een evaluatie heeft een startpunt, dat ingegeven kan zijn vanuit een nood aan informatie of vanuit een verplichting tot het leveren van beleidsinformatie aan bepaalde stakeholders. De initiële opdracht bevat reeds een duidelijke indicatie over het startpunt van het evaluatieproces. Daarnaast bevat het tevens informatie over de te verzamelen informatie, over mogelijke bevindingen, conclusies en aanbevelingen. Ingeval de evaluatie wordt uitbesteed, vermeldt het bestek tevens een indicatie over het einde van het evaluatieonderzoek, vaak met een exacte opleveringsdatum.

In tegenstelling tot een evaluatieproces, is het monitoringproces initieel opgevat als een permanent proces, dat doorlopend informatie verzamelt en in functie van rapportering op regelmatige tijdstippen. Monitoring zal, éénmaal geïmplementeerd, deel uitmaken van de informatie-infrastructuur van de organisatie. Momenteel zijn verschillende technologieën voorhanden om een database up to date te houden en permanent te upgraden.

2. De inhoudelijke scope: sterk afgelijnd vs. breed

Evaluaties zijn meestal ontworpen om specifieke vragen van stakeholders te beantwoorden in een vooraf bepaald tijdsperspectief. De begunstigden van de evaluatie zijn, in de meeste gevallen, gekend van bij de start van het evaluatieproces. De onderzoeksopdracht maakt melding van de belangrijkste evaluatievragen en het te evalueren beleid. Tevens wordt een

budget voorzien voor de evaluatieopdracht. Vaak besteedt men evaluaties uit. De opvolging van het project ligt meestal in handen van een stuurgroep. Sommige organisaties beschikken over een eigen evaluatiecapaciteit, voor ondersteuning en interne uitvoering van evaluaties. Zowel voor interne als externe evaluaties dient men steeds een moment te voorzien voor onderhandeling over de condities en de inhoud van het evaluatieproces.

Het voorgaande staat in contrast met monitoringsystemen dat een meer permanent karakter heeft zowel wat de verzameling als de verspreiding van informatie betreft. De ontwikkeling van een monitoringsysteem vereist het akkoord van het management over de verschillende elementen en processen die het systeem dient te bevatten. Na de beslissing over wat er gemeten moet worden, zal de organisatie maatregelen en processen opstarten om de verzameling, de analyse en de rapportering van de data mogelijk te maken. Éénmaal ontwikkeld blijft het systeem de vooraf bepaalde gegevens verzamelen tenzij het management het raadzaam vindt na verloop van tijd wijzigingen aan te brengen.

3. Maatstaven op basis van ad hoc keuzes vs. routine processen.

Het ad hoc karakter van evaluatie zorgt ervoor dat de informatiebehoefte ook in functie van de evaluatievragen worden gevormd. Bijgevolg dient de gegevensverzameling vaak op specifieke aspecten te focussen. Het beantwoorden van de gestelde vragen vereist traditioneel de verzameling en analyse van zowel primaire als secundaire informatie. De primaire data zullen door het evaluatieteam zelf worden verzameld in aanvulling op secundaire data.

Monitoringsystemen hanteren indicatoren die doorgaans betrekking hebben op de routinematige werking van de organisatie. Dit maakt bijvoorbeeld periodieke vergelijkingen van de actuele resultaten met de vooropgestelde doelstellingen mogelijk.

4. Het attributie- of causaliteitsaspect: in vraag gesteld vs. verondersteld

Een belangrijk verschilpunt tussen monitoring en evaluatie betreft de attributievraag, of de vraag in welke mate geobserveerde effecten daadwerkelijk toe te schrijven zijn aan de beleidsoutput. In een ex post evaluatie kan dit de centrale vraagstelling zijn: het nagaan van het causale verband tussen output en effecten. Een evaluatie toetst dan het beleid op het criterium van effectiviteit. In boekdeel 1 stelden we al dat het onderzoek naar effectiviteit één van de meest geavanceerde aspiraties is van beleidsevaluatie. Anderzijds gaat het om een zeer pertinente vraagstelling naar de raison d'être van de overheid.

Evaluatoren kunnen verschillende strategieën aanwenden om het attributievraagstuk te beantwoorden, maar de strategiekeuze zal vaak gestuurd worden door praktische en ethische

elementen (cf. deel 1). Evaluatoren kunnen vaak geen volledig sluitende verbanden en verklaringen aangeven. Evaluatieconclusies dienen dan ook beschouwd te worden als een manier om de onzekerheid rond deze zeer belangrijke evaluatievraag te reduceren.

Monitoringsystemen zullen maar zelden de capaciteit hebben gericht op het reduceren van het attributieprobleem. In de ontwikkeling van maatstaven zal men eerder aandacht besteden aan de zogenaamde constructievaliditeit (cf. infra): meten de indicatoren wat ze moeten meten? Zijn ze een accurate vertaling van de realiteit die men in kaart wil brengen en opvolgen? Beleid en maatregelen komen tot stand in een steeds veranderende omgeving en zijn bedoeld om bepaalde veranderingen of effecten in die omgeving te bewerkstelligen. Monitoring heeft dan tot doel te bekijken en te analyseren in welke mate de vooropgestelde effecten gerealiseerd zijn. De vraag in welke mate het om beleidseffecten gaat, d.i. de attributie van effecten aan het beleid, wordt niet expliciet gesteld. Monitoring in se kan deze vraag ook niet alleen beantwoorden; het levert enkel beschrijvende informatie op, draagt geen verklaringen aan. Bij het interpreteren van informatie uit monitoring zal men in de praktijk echter vaak geneigd zijn om aan te nemen dat de geobserveerde effecten het gevolg zijn van de beleidsoutput. Met andere woorden: de attributie wordt verondersteld.

Zeker wanneer men met tijdreeksen werkt en waarbij een duidelijke verandering af te leiden valt uit de gegevens die met de uitvoering van beleid samenvalt of daarop volgt, zal men vaak zonder meer de conclusie trekken dat deze verandering toe te schrijven is aan het beleid. Dat kan een correcte interpretatie zijn, maar ze kan evenzeer fout zijn. De facto kan de onzekerheid hoog blijven. Immers, ook andere, externe factoren kunnen het gemeten effect (deels) veroorzaakt hebben.

Monitoring is dus beperkt om informatie te genereren waarmee men de vraag naar attributie of causaliteit kan beantwoorden. Op het vlak van output stelt zich doorgaans geen probleem. De vraag of een bepaalde output al dan niet het gevolg is van de inzet van middelen door de overheid is meestal met grote zekerheid correct te beantwoorden. De output is immers rechtstreeks gekoppeld aan de inzet van middelen en de processen binnen de werking van de overheid. Het aanleggen van een weg, het afleveren van een identiteitskaart of sociale begeleiding voor werkzoekenden: het zijn voorbeelden van producten en dienstverlening die men zonder twijfel kan toeschrijven aan de overheid. Op het vlak van effecten is de attributie vraag heel wat gecompliceerder. Effecten doen zich voor buiten de organisatiegrens en de overheid heeft er niet in dezelfde mate vat op als op de output die ze aflevert. De mate waarin men attributie van effecten aan de output kan veronderstellen hangt samen met de

routinegraad van het beleid. In een beleid met een hoge routine, zoals een aanbod aan sociale woningen creëren, is er een hoge graad van waarschijnlijkheid dat dit aanbod zal leiden tot een verkorting van de wachttijden. Een beleidsinitiatief met een lagere graad van routine zoals het aanbieden van training in de zoektocht naar werk, is de koppeling van effecten heel wat moeilijker. Mensen terug aan werk helpen is een complexere opgave dan het bouwen van woningen. Samengevat, kan men stellen dat het attributieprobleem groter zal zijn naarmate de ambiguïteit van het takenpakket voor het uitvoeren van het beleid groter is. Des te moeilijker zal het zijn om louter op basis van data uit monitoring conclusies te trekken over de effectiviteit van het beleid.

5. Bronnen voor gegevensverzameling: ad hoc en/of ingebed in processen van dienstverlening

Uit het derde verschilpunt tussen evaluatie en monitoring volgt dat evaluatie en monitoring ook verschillen op het vlak van de bronnen voor dataverzameling. In een evaluatie zal men vaak aangewezen zijn op een mix van primaire en secundaire bronnen, aangepast aan de specifieke informatiebehoeften op basis van de evaluatievragen. Het meer permanente karakter van een monitoringsysteem zal vaak leiden tot een sterke inbedding van het meetsysteem in de werking van de organisatie. Meetgegevens worden bijvoorbeeld rechtstreeks gehaald uit de verwerking van dossiers (bv. bouwvergunningen).

6. De actoren: managers ontwikkelen doorgaans het meetsysteem in een organisatie maar zijn niet altijd de evaluatoren

De begeleiding van evaluaties in overheidsorganisaties is vaak in handen van een stuurgroep, waarin de belangrijkste stakeholders vertegenwoordigd zijn (zie boek 1). In kleinere organisaties en agentschappen zal het lijnmanagement vaak een centrale plaats innemen in de stuurgroep belast met de opvolging van de evaluatie. Maar lang niet altijd zullen managers de evaluatie zelf uitvoeren. De betrokkenheid van het management is een onderwerp waarover in theorie en praktijk geen eensgezindheid bestaat. Voorstanders beschouwen het als een belangrijke kritieke succesfactor voor het welslagen van het evaluatieproces. Tegenstanders echter zijn voornamelijk bekommerd over de onafhankelijkheid van het evaluatieteam. Veel zal dan ook afhangen van het evaluatiemotief.

Wat de ontwikkeling van meetsystemen betreft, beschouwt men algemeen de directe betrokkenheid van het management als vanzelfsprekend en zelfs als een noodzakelijke voorwaarde.

7. Opzet en gebruik: vooraf vastgelegd vs. flexibel aanpasbaar

De inhoud en scope van een evaluatie wordt steeds vooraf bepaald als eerste stap in het evaluatieproces. Alle verdere stappen krijgen vorm door beslissingen genomen in deze eerste stap.

Monitoringsystemen verzamelen permanent gegevens over de werking van een organisatie en haar omgeving. Men moet een meetsysteem kunnen aanpassen aan de veranderende verwachtingen en informatienoden. Als gevolg hiervan is een monitoringsysteem op regelmatige basis onderhevig aan wijzigingen in focus en opzet. Indicatoren kunnen ontwikkeld worden om bepaalde output en effecten te meten die in de toekomst aan belangrijkheid zullen afnemen. Nieuwe aandachtspunten van het management of andere belangrijke stakeholders kunnen vragen om een periodieke bijsturing van het meetsysteem. Een evaluatie, eens inhoudelijk afgebakend, kan minder snel nog gewijzigd worden. Andere evaluatievragen of – criteria kunnen namelijk een heel andere aanpak op het vlak van dataverzameling, -analyse en beoordeling vergen.

2.2.3 Meten en evalueren van efficiëntie en effectiviteit

Uit voorgaande paragrafen kunnen we concluderen dat monitoring en evaluatie complementair kunnen zijn als twee verschillende, ondersteunende functies om beslissingen in het kader van beheer en beleid te informeren.

Twee belangrijke evaluatiecriteria zijn de efficiëntie en effectiviteit. Het gaat respectievelijk om het verband tussen input en output en tussen output en effect. In deze paragraaf staan we even stil bij deze begrippen en het meten en evalueren ervan.

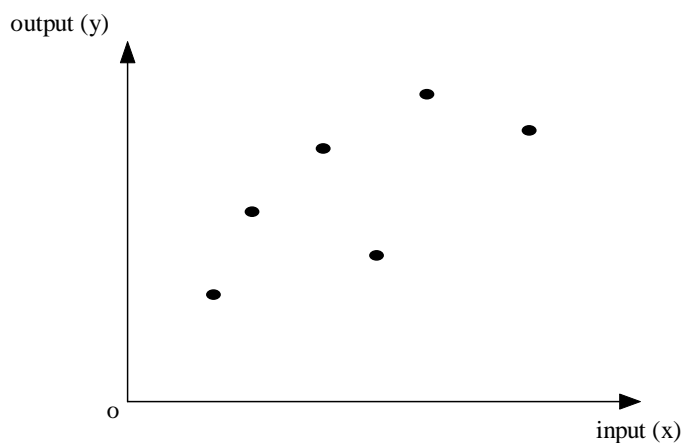
2.2.3.1 Efficiëntie

Het concept efficiëntie in een overheidscontext verwijst naar de relatie tussen de ingezette middelen (input) en de afgeleverde producten of dienstverlening (output). Bijvoorbeeld de tijd (input) die een administratie nodig heeft om de constructie van een brug (output) te plannen, uit te besteden en in gebruik te nemen.

Het evalueren op efficiëntie gebeurt doorgaans vanuit het motief om het uitvoeringsproces van het beleid te verbeteren. Dit evaluatiecriterium is dus vooral van tel vanuit een managementperspectief.

De keerzijde van efficiëntie is productiviteit. Efficiëntie wijst op de hoeveelheid input die nodig per eenheid output, terwijl productiviteit een maatstaf is voor de hoeveelheid output die men kan realiseren per eenheid input.

De relatie tussen input en output kan grafisch als volgt worden weergegeven.

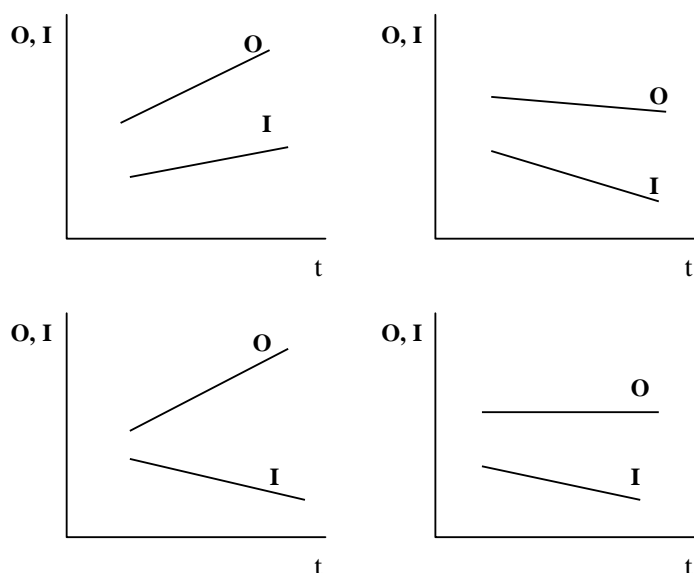


Figuur 2: Spreidingdiagram van inputoutput combinaties

De punten in het diagram stellen combinaties van een hoeveelheid input en output (x, y) voor.

Efficiëntie is een variabele, d.w.z. de efficiëntie van een proces of organisatie kan variëren doorheen de tijd; de efficiëntie kan toenemen of afnemen.

Volgende figuren stellen telkens een efficiëntieverbetering voor die op een andere wijze bereikt wordt. De output kan sneller stijgen of langzamer afnemen dan input. De output kan toenemen terwijl de input daalt. Of de output kan constant blijven, terwijl de vereiste input afneemt.



Figuur 3: Strategieën voor efficiëntieverbetering

De eerste grafiek op de bovenste rij reflecteert bijvoorbeeld de situatie waarbij meer middelen worden ingezet om politiecontroles uit te voeren, terwijl men het aantal politiecontroles meer dan evenredig opvoert. In de tweede grafiek op de bovenste rij daalt het aantal controles. Verhoudingsgewijs dalen ook de ingezette middelen (bv. aantal agenten uitgedrukt in VTE).

De eerste grafiek op de onderste rij geeft de situatie weer waarbij de middeleninzet afneemt maar waarbij het aantal controles stijgt; de efficiëntiewinst is hier het meest uitgesproken. De tweede grafiek op de onderste rij geeft aan dat het aantal controles stabiel blijft, maar dat men minder middelen eraan besteedt.

Om efficiëntie te meten dient men dus steeds twee componenten te meten: input en output. Het verband tussen input en output is dus de productiefunctie en wordt grafisch voorgesteld door een curve. Die curve is de grens tussen technisch optimale of efficiënte combinaties van input en output (punten op de grens) en niet-efficiënte combinaties (punten onder de grens). Deze functie kan met 2 soorten technieken bepaald worden: parametrische en niet-parametrische technieken.

Welke techniek is preferabel? Belangrijk voor het antwoord op deze vraag is het specifieke karakter van de overheidscontext: het exacte verband tussen input en output is niet zomaar a-priori gekend in een beleidscontext. Met andere woorden: we kennen niet op voorhand de parameters die de functie mee bepalen. Om het verband te achterhalen, hanteren we daarom een niet-parametrische techniek: deze laten meer vrijheid en vertrekken van de data uit monitoring zelf. Het verschil tussen de niet-parametrische technieken onderling zit in de formele assumpties, bv. het al dan niet veronderstellen van schaalvoordelen.

Een voorbeeld van een niet-parametrische techniek is de *Free Disposal Hull* techniek (FDH). De techniek gaat op zoek naar beste praktijken (benchmarks) met betrekking tot een combinatie van input en output. Het gaat dus om een relatieve en geen absolute beoordeling van efficiëntie.

De Free Disposal Hull techniek vertrekt van de basisveronderstelling: "Voor elke observatie met een bepaalde combinatie input-output, neemt men aan dat het ook mogelijk is om met dezelfde input *minder* output te produceren, of om dezelfde output te produceren met *meer* input". De basisassumptie is dus deze van de vrije beschikbaarheid (free disposal) van input en output.

Voor meer informatie over deze techniek en een uitgewerkt voorbeeld verwijzen we de lezer naar deel 3 van de handleiding.

2.2.3.2 Effectiviteit

Het concept effectiviteit verwijst naar de oorzaak en gevolg relatie tussen output en effect. Heeft het uitgevoerde beleid daadwerkelijk het beoogde effect gehad? In welke mate zijn er gewenste of ongewenste neveneffecten opgetreden?

Zoals uiteengezet in voorgaande paragraaf dekken monitoring en evaluatie een verschillende lading. In het kader van onderzoek naar de effectiviteit van beleid wordt dit onderscheid meteen duidelijk. Monitoringsystemen zullen op basis van indicatoren de beleidsoutput en effecten kunnen meten en registreren. Evaluatie gaat nog een stap verder en tracht na te gaan of de geobserveerde effecten toe te schrijven zijn aan de output. Het beleid dient dan afgezonderd te worden van alternatieve verklaringsgronden voor de waargenomen effecten of verandering.

Het effectiviteitsvraagstuk in een beleidscontext vraagt om het aantonen van een causale relatie tussen output en effect. In het dagelijkse leven leggen we vaak eenvoudige causale verbanden. Je kan vaststellen dat het botsen van een andere auto tegen de jouwe de oorzaak is van de schade aan je wagen. De vele studie-uren liggen aan de basis van de leerresultaten. Te weinig beweging en ongezonde voeding liggen aan de basis van overgewicht. Ook trachten we meer complexe verbanden te achterhalen. De stress op het werk kan men aanduiden als de oorzaak van een slaapgebrek. Doordat men te weinig slaapt, is men overdag nogal prikkelbaar wat opnieuw leidt tot stress. Men kan echter in een vicieuze cirkel terecht komen waarin oorzaak en gevolg niet zo goed te onderscheiden zijn. In het voorbeeld kan stress zowel oorzaak als gevolg zijn.

In boekdeel 1 werd uitgebreider stilgestaan bij de voorwaarden van causaliteit. Shadish, Cook en Campbell (2002) definiëren als volgt: een causale relatie bestaat als een oorzaak het effect vooraf gaat, als de oorzaak gerelateerd is aan het effect en indien geen verklaring kan gevonden worden voor het effect anders dan de oorzaak. De laatste voorwaarde uit de definitie is meteen de moeilijkste om aan te tonen. Het aantonen van beleidseffectiviteit impliceert dat men het beleid kan afzonderen van alternatieve verklaringsgronden voor de geobserveerde effecten. In de mate waarin de effecten daadwerkelijk het gevolg zijn van het beleid, kunnen we spreken van beleidseffectiviteit en van "beleidseffecten".

In deel 1 van de handleiding gaan we uitvoerig in op mogelijke strategieën om beleidseffectiviteit aan te tonen. We hernemen hier slechts in hoofdlijnen de belangrijkste opties en aandachtspunten. Voor een uitgebreide bespreking verwijzen we de lezer naar boekdeel 1.

Het causaliteitsvraagstuk is in de eerste plaats een vraagstuk over zekerheid en onzekerheid. Om causaliteitsvraagstukken op te lossen is het belangrijk een analyse te maken van

belangrijkste tegenargumenten. Afhankelijk van de omstandigheden kan het voldoende zijn om alleen de belangrijkste tegenargumenten te elimineren, in andere gevallen kan de evaluator gedwongen zijn om praktisch elke mogelijke verklaring uit te schakelen. Hieronder presenteren we een overzicht van mogelijke strategieën die toelaten andere oorzaken of variabelen te verklaren. Ze zijn gerangschikt van vrij eenvoudige tot meer geavanceerde opties. Soms is het mogelijk of aangewezen om ze te combineren.

Het reduceren van de onzekerheid kan plaatsvinden aan de hand van de onderstaande 'strategieën mix'. Zelfs indien een hoog 'zekerheidsgehalte' dient gehaald te worden is het aan te raden niet op één enkele strategie in te zetten, maar op een uitgebalanceerde mix van technieken met verschillende sterkten en zwakten. Een evaluatieteam beschikt over verschillende mogelijkheden om causaliteit aan te tonen, van zeer eenvoudige technieken tot meer complexe methoden.

Davidson (2004) zet een aantal aandachtspunten op een rij.

- Vraag het aan stakeholders

Stakeholders zijn van dichtbij betrokken bij het beleidsproces. Zo kan men bijvoorbeeld de doelgroep van het beleid bevragen naar hun ervaringen en gedragsverandering. In welke mate heeft de doelgroep een verandering ervaren? Vervolgens kan men nagaan waarom personen en bepaald gedrag vertonen en wat zij als oorzaak van hun gedragsverandering zien. Naast de oorzaak krijgt het evaluatieteam op deze wijze ook inzicht in mogelijke beïnvloedende contextvariabelen die bijgedragen hebben aan de gemeten effecten. Het spreekt voor zich dat in heel wat situaties de feedback van individuele leden uit de doelgroep niet voldoende is om de eindconclusies over beleidseffectiviteit te trekken.

- Stemt de beleidsinhoud overeen met de beoogde effecten?

Een beleidsinitiatief heeft een bepaalde focus: het wil bepaalde hefboomen beïnvloeden om de beoogde verandering te realiseren. Men kan nagaan of deze hefboomen daadwerkelijk een rol hebben gespeeld in de effectveranderingen die men heeft opgetekend. Zo kan een opleiding voor werklozen bepaalde vaardigheden en strategieën aanleren voor het voeren van een degelijk sollicitatiegesprek. Men gaat er dan vanuit dat werklozen die werk vonden na deze training, hierin geslaagd zijn mede door gebruik te maken van de aangeleerde technieken. Wanneer echter blijkt dat het merendeel van de werklozen de aangeleerde kennis niet aansprak tijdens een sollicitatiegesprek, kan men ernstige vragen stellen bij de meerwaarde van het trainingsprogramma. Dit geldt ook wanneer blijkt dat het merendeel

van de werklozen, dat wel gebruik gemaakt heeft van de aangeleerde technieken, na verloop van tijd nog steeds geen werk vond. In beide gevallen blijkt de beleidsinhoud niet te stroken met de noden van de doelgroep.

- Treden effecten op volgens het verwachte patroon?

De geobserveerde effecten kunnen ofwel overeenstemmen, ofwel tegenover gesteld zijn aan de verwachte effecten. Als ze overeenstemmen, versterkt dit het vertrouwen in de causale relatie tussen het beleid en de effecten. Indien beiden verschillen, bestaat een reële kans dat andere factoren een belangrijke rol spelen. Dit verzwakt de assumptie dat het beleid effectief is. Deze algemene check is het meest zinvol voor beleidsinitiatieven waarvan men de verwachte effecten duidelijk kan benoemen. Indien slechts weinig geweten is van mogelijke verbanden tussen beleid en mogelijke effecten, kan men zoeken naar gelijkaardige, reeds gekende situaties en nagaan of er analoge patronen zijn in effecten.

- Timing van effecten

Belangrijk voor het inschatten beleidseffectiviteit is de kennis over de timing waarmee eventuele beleidseffecten in principe waarneembaar moeten zijn. Meten is weten, maar men dient te weten wanneer men best metingen uitvoert. Een sensibiliseringscampagne bijvoorbeeld, zal vrij snel tijdens en na de campagne een mogelijke invloed hebben op het gedrag van de doelgroep. Dit effect zal geleidelijk aan uitdoven. Om het effect van de jaarlijkse BOB-campagne na te gaan, zal men best niet wachten tot in de zomerperiode om een verandering in het gedrag van chauffeurs te onderzoeken. Men weet immers uit ervaring en monitoring dat het effect van dergelijke campagne tijdelijk is en geleidelijk uitdooft. Daarom wordt ze ook elk jaar herhaald in de eindejaarsperiode.

- Staat de inspanning van het beleid ('dosis') in een logische verhouding tot de effecten ('respons')?

Wanneer beleid daadwerkelijk tot effecten leidt, kan men meestal stellen dat de mate waarin er een effectverandering observeerbaar is in verhouding staat met de inspanning van het beleid. Vanuit dit perspectief kan men ook de effectiviteit van het beleid beoordelen in situaties wanneer het beleid niet of niet helemaal werd uitgevoerd zoals gepland. In de praktijk komen dergelijke situaties vaak voor. Bijvoorbeeld: een hogesnelheidslijn waarvan het traject maar ten dele wordt gerealiseerd, een steunmaatregelenpakket gericht op startende bedrijven of cultuurverenigingen dat een lager budget krijgen toegekend dan aanvankelijk begroot of dat slechts voor een deel wordt uitgevoerd, enz. Wanneer men kan vergelijken tussen gelijkaardige beleidsinitiatieven die in verschillende mate werden

uitgevoerd, wordt het mogelijk om zicht te krijgen op de dosis - respons verhouding. Wanneer effecten in dezelfde of zelfs sterkere mate optreden wanneer het beleid onvolledig werd ingevoerd in vergelijking met een case waarbij het beleid wel volledig werd geïmplementeerd, dan is er een indicatie dat de effecten weinig of niet het gevolg zijn van het beleid.

Een belangrijke bemerking bij de analyse van deze dosis en respons relaties heeft betrekking op de vaststelling dat niet alle verbanden lineair verlopen.

Zo kan er een plafondeffect optreden: een hogere beleidsinspanning resulteert vanaf een bepaald niveau niet meer in een evenredige stijging van het verwachte effect zoals bij een lagere 'beleidsdosis'. Stel dat een opleidingsprogramma voor werkzoekende doorlopen kan worden in 4 of 6 weken. Wanneer blijkt dat herintredingsgraad op de arbeidsmarkt niet hoger ligt bij de personen die de lange versie hebben gevolgd in vergelijking met zij die de korte opleiding volgden, dan kan men besluiten dat de cursus algemeen kan teruggebracht worden naar 4 weken; een langer begeleidingstraject blijkt dan geen extra voordelen op te leveren met betrekking tot kansen om werk te vinden.

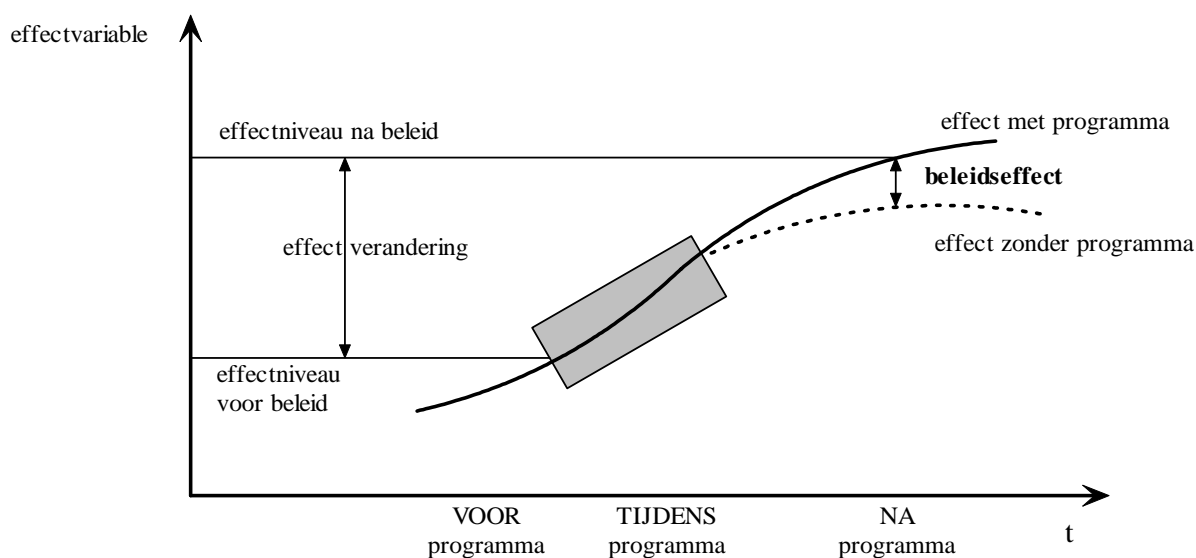
Ook het fenomeen van een 'overdosis' kan vóórkomen; een te grote beleidsinspanning of tijdsduur van het beleid leidt tot een suboptimaal (of zelfs negatief) effect. Schoolkinderen bijvoorbeeld zullen een bepaald aantal uren huiswerk aanvaarden, tot op een moment dat ze een afkeer van schoolwerk krijgen. Te zware taakopdrachten buiten de school kunnen dan leiden tot demotivatie en slechtere leerprestaties.

Dosis en respons relaties kunnen het best onderzocht worden in een beleidscontext waarin verschillende varianten van een beleidsinitiatief in uitvoering worden gebracht.

- Vergelijken tussen de 'policy-on' en de 'policy-off' situatie

Het aantonen van beleidseffectiviteit omvat dat men op een doordachte en systematische wijze het aandeel van het beleid als verklaring voor de waargenomen effecten of veranderingen probeert af te zonderen van alternatieve, externe verklaringsgronden. Het komt erop neer om de situatie met en zonder beleid met elkaar te vergelijken. Als men kennis heeft over wat er gebeurd zou zijn zonder beleid, kan men conclusies trekken over de situatie waarin het beleid wel werd uitgevoerd.

Het evaluatiekader voor het evalueren van beleidseffectiviteit wordt weergegeven in onderstaande figuur. De stippellijn geeft aan wat er gebeurd was indien het beleid niet was uitgevoerd. Het verschil in effectverandering tussen de situatie mét en de situatie zonder beleid, geeft het feitelijke "beleidseffect" aan.



Figuur 4: Effectniveaus, effectverandering en beleidseffect

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 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

De keuze van een evaluatiedesign in het kader van het evalueren van beleidseffectiviteit wordt mee gestuurd door praktische en ethische elementen en dus meestal geen vrije keuze.

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

Elk van deze designs is echter gebaseerd op het meten van effecten of veranderingen in de waarde van zogenaamde effectvariabelen (indicatoren).

In een experimenteel design werkt men bijvoorbeeld met twee metingen (O): voor en na de uitvoering (X) van het beleid in de beleidsgroep.

Beleidsgroep	O	X	O
Controlegroep	O		O

De metingen leveren waarden op de effectvariabele op. Per groep kan men het verschil in waarden (voor en na) berekenen. Vervolgens kan men de veranderingen tussen groepen vergelijken. Het saldo komt overeen met het beleidseffect.

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

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.

In een quasi-experimenteel design werkt men in bepaalde gevallen met een vergelijkingsgroep die echter kan verschillen van de beleidsgroep op relevante factoren die een alternatieve verklaringsgrond voor de waargenomen effecten kunnen zijn, naast beleid. Deze factoren dienen dan ook gemeten en opgevolgd te worden, voor zover de beleids- en vergelijkingsgroep hierop verschillen.

Beleidsgroep	O	X	O
Vergelijkingsgroep	O		O

In andere gevallen is het onmogelijk om met een controle- of vergelijkingsgroep te werken en dient men de 'beleidsgroep' met zichzelf te vergelijken in de tijd, waarbij men ook relevante externe factoren zal monitoren.

Beleidsgroep	O	X	O
--------------	---	---	---

Door het opbouwen van (al dan niet comparatieve) tijdreeksen via meerdere metingen of monitoring, kan men beter zicht krijgen op evoluties bij de effectvariabele.

Beleidsgroep	O	O	O	O	X	O	O	O	O
--------------	---	---	---	---	---	---	---	---	---

Beleidsgroep	○	○	○	○	X	○	○	○	○
Vergelijkingsgroep	○	○	○	○	○	○	○	○	○

Voor een uitgebreide bespreking van de diverse strategieën om effectiviteit vast te stellen, verwijzen we de lezer naar boekdeel 1 van de handleiding. Elke strategie heeft voor- en nadelen en zal met een bepaalde graad van zekerheid conclusies kunnen genereren met betrekking tot de mate van beleidseffectiviteit.

Het ontwerpen van stevige conclusies vereist de verzameling van informatie betreffende mogelijke tegenargumenten. Het proces dient verder gezet te worden tot dat een voldoende hoog zekerheidsniveau bereikt is.

Kan een evaluator alle onzekerheden weg nemen? In het merendeel van de gevallen niet, de maatregel, project of beleid vindt plaats in een complexe realiteit met tal van invloeden en interacties.

2.2.4 Uitdagingen

In voorgaande secties werd nader ingegaan op de stelling dat evaluatie en monitoring complementaire methoden zijn om het beleid en beheer te ondersteunen en te informeren. Evalueren en monitoring kunnen als een tandem beschouwd worden. Een optimale werking van deze tandem in de praktijk is echter niet vanzelfsprekend. In de voorgaande secties wezen we op belangrijke verschillen tussen monitoring en evaluatie. De ontwikkeling van een capaciteit voor beide functies, met hun onderlinge complementariteit tegelijk als uitgangspunt en het beoogde resultaat, gaat gepaard met enkele uitdagingen. Een integrale capaciteit voor monitoring en evaluatie kan slechts binnen een context van een lerende organisatie opgebouwd worden, wat een reeks randvoorwaarden met zich mee brengt.

De context: naar lerende organisaties

Het benutten aan het potentieel uit de complementariteit tussen monitoring en evaluatie kan slechts wanneer dit gebeurt in een overkoepelend kader van een lerende organisatie.

Senge (1999) identificeerde drie kerncompetenties waarvoor lerende organisaties een capaciteit moeten opbouwen:

1. scannen en anticiperen van verandering in de bredere omgeving om relevante variaties te detecteren
2. de mogelijkheid om gangbare normen en assumpties in vraag te stellen, te analyseren en te veranderen

3. het laten tevoorschijn komen van een aangewezen strategische koers en patroon van organisatie

Garvin (1993) suggereert een vijftal fundamentele om een lerende organisatie te creëren. In de onderstaande tabel staan deze bouwstenen samen met de bijhorende activiteiten vermeld.

<ol style="list-style-type: none"> 1. Systematisch probleemoplossing <ul style="list-style-type: none"> • Problemen aanpakken door het gebruik van een sequens van hypothesen creëren en vervolgens acties ondernemen om deze hypothesen te testen • Meer belang hechten aan gegevens dan aan veronderstellingen • Aandacht voor details: “Hoe weten we dat het waar is?” 2. Experimenteren <ul style="list-style-type: none"> • Kleine, gecontroleerde veranderingen en testen van bestaande programma's • Zoeken naar en testen van nieuwe kennis (pilot- en demonstratieprojecten) • Managers moeten zowel incentives als vaardigheden hebben om te experimenteren 3. Leren uit vroegere ervaringen <ul style="list-style-type: none"> • Systematisch registeren, tonen en beoordelen van bewijs van eerdere verwezenlijkingen • Deze informatie en de vaardigheden om ze te interpreteren en te gebruiken moeten wijdverspreid zijn binnen de organisatie 4. Leren van anderen <ul style="list-style-type: none"> • Er is niets mis met het overnemen van ideeën • Onderzoek wie “de beste” is, begrijp waarom ze dat zijn, en pas hun praktijk aan de eigen organisatie aan 5. Overdragen van kennis <ul style="list-style-type: none"> • Kennis moet snel en efficiënt verspreid worden doorheen de organisatie • Kennis wordt behandeld als een hulpbron

Tabel 3: Fundamentele voor een lerende organisatie

Doorheen de activiteiten opgesomd in de tabel kan men een sleutelrol voor evaluatie waarnemen.

Een lerende organisatie kenmerkt zich door een dubbele leercirkel: in een eerste cyclus leert zij in welke mate zij haar vooropgestelde doelstellingen haalt (beschrijvend). Echter, in een tweede cyclus leert zij waarom en hoe de vooropgestelde doelstellingen worden behaald (verklarend). De vragen die behoren tot de eerste cyclus kunnen beantwoord worden met monitoring. Voor

de vragen die ressorteren onder de tweede cyclus is evaluatie noodzakelijk. Dergelijke dubbele leercirkel vereist dat verschillende perspectieven en motieven voor meten bij elkaar worden gebracht. In het kader van de eerste cyclus zijn voornamelijk opvolging van de beleidsuitvoering en verantwoording van tel als motief. Voor de tweede leercirkel zijn de motieven van planning en beleidsleren belangrijk.

Merk op dat we een parallel kunnen trekken met de twee fundamentele vraagstellingen in een overheidscontext. De eerste leercirkel sluit aan bij de vraag “Doen we de dingen goed”?, terwijl tijdens de tweede leercirkel de vraag “Doen we de goede dingen?” centraal staat.

Het creëren van een lerende organisatie als de aangewezen context voor het valoriseren van de complementariteit van monitoring en evaluatie, omvat het scheppen van volgende randvoorwaarden.

Randvoorwaarden

- Erkenning van verschillende perspectieven en motieven voor meten

Monitoring en evaluatie hanteren deels dezelfde perspectieven en motieven voor meten (cf. supra). Om te komen tot een lerende organisatie met een capaciteitsopbouw voor beide complementaire functies ter ondersteuning van het beleid en beheer, dient die verscheidenheid in de eerste plaats erkend te worden door de verantwoordelijken voor de opbouw van deze capaciteit. In de tweede plaats zal deze erkenning doorvertaald moeten worden naar het hanteren van verschillende perspectieven en motieven. In heel wat praktijkgevallen zal het evolueren naar een lerende organisatie immers neerkomen op een uitbreiding en verdieping van een eerste leercirkel naar een tweede leercirkel. De bestaande capaciteit voor monitoring werd bij veel overheidsorganisaties geïntroduceerd vanuit een managementperspectief: het opvolgen van processen en resultaten. Dergelijk perspectief sluit nauw aan bij de motieven van implementatieopvolging en verantwoording. Daarnaast zal de ondersteuning van planning in veel gevallen ook worden erkend als motief om dergelijke informatiecapaciteit op te bouwen.

Een belangrijke uitdaging ligt echter in de erkenning van beleidsleren als evaluatiemotief door managementverantwoordelijken. Beleidsleren is van een andere aard dan verantwoording. Aan verantwoording wordt immers vaak een context en mechanisme van sanctionering gekoppeld, terwijl dit afwezig is bij beleidsleren. Vanuit het motief van beleidsleren kunnen deels dezelfde evaluatievragen worden gesteld als vanuit een verantwoordingsmotief. Echter verantwoording blijft vaak beperkt tot de organisatiegrens. Omwille van de context van afrekening en sanctionering wordt de focus meestal beperkt tot die elementen waarop de organisatie het meeste vat heeft: middelen, processen en output. Effecten zijn echter belangrijk om op te volgen vanuit meerdere motieven. Maar de wijze waarop met deze

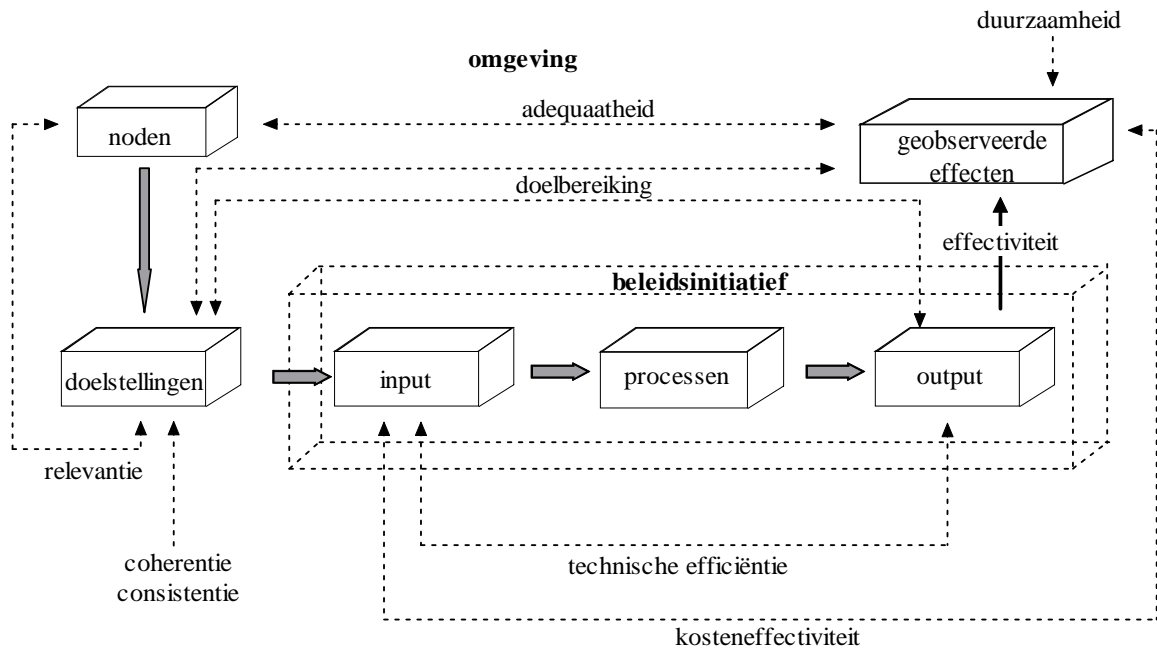
effectmetingen omgaat, geeft een volgende uitdaging aan in de context van een lerende organisatie.

- Geen onvoorwaardelijke attributie van effecten aan het beleid

Wanneer effecten ook opgevolgd worden met monitoring vanuit een managementperspectief worden deze meestal buiten beschouwing gelaten in rapporteringen in het kader van verantwoording, ofwel worden deze effecten als een vanzelfsprekend gevolg gezien van de output van de organisatie. Deze laatste link is echter een assumptie die zelden onderbouwd of nader onderzocht wordt, en waarover dus vaak geen zekerheid bestaat. In bepaalde gevallen zal deze link eenduidig en evident zijn. In veel gevallen echter is dat niet het geval. Dan bestaat het risico dat beslissingen worden genomen, bijvoorbeeld over de verder- of stopzetting van een beleidsinitiatief of aanpak gebaseerd op foutieve veronderstellingen. Het is daarom van belang dat effecten niet onvoorwaardelijk aan het beleid of de werking van de organisatie worden toegeschreven. In een evaluatie kan deze attributievraag net centraal staan, wanneer het evaluatiecriterium effectiviteit is.

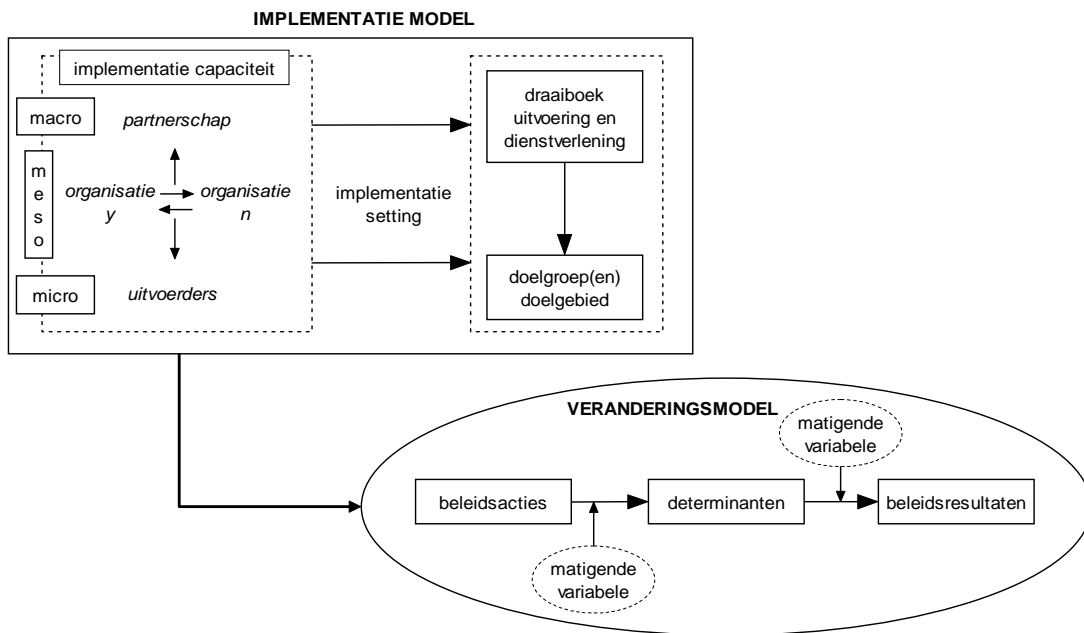
- Inhoudelijke afstemming van gegevensverzameling

In de context van een lerende organisatie is het van belang om de gegevensverzameling inhoudelijk af te stemmen, gegeven de verschillende perspectieven en motieven die aan de basis van het meten kunnen liggen. Een robuuste kenniscapaciteit zal daarom moet gefundeerd worden met een mix van indicatortypes. De verschillende soorten gegevens, zoals aangegeven in onderstaande figuur dragen bij tot een complete en veelzijdige gegevensverzameling.



Figuur 5: De beleids- en beheerscyclus als open systeemmodel

Wanneer het managementperspectief moet aangevuld worden met een beleidsperspectief, is de achterliggende beleidstheorie het aangewezen kader voor de bepaling van relevante indicatoren om in het monitoringsysteem op te nemen. Zoals uiteengezet in deel 1 kan de beleidstheorie schematisch voorgesteld worden aan de hand van een veranderingsmodel en implementatiemodel.



Figuur 6: Het conceptueel raamwerk van beleidstheorie

De componenten van deze modellen kunnen als bron voor indicatoren gehanteerd worden, waardoor verschillende perspectieven en motieven met betrekking tot monitoring en evaluatie afgedekt en ondersteund kunnen worden.

- Operationele afstemming van gegevensverzameling

Naast een inhoudelijke afstemming is ook een operationele stroomlijning van de infrastructuur voor dataverzameling vereist wanneer deze verschillende motieven moet ondersteunen in het kader van een lerende organisatie. Zo vormt het tijdsperspectief een verschilpunt tussen monitoring en evaluatie: monitoring is een permanente functie terwijl evaluaties meestal binnen vooraf afgelijnde tijdssegmenten plaats vinden. Echter, een belangrijke uitdaging ligt in het integreren van gegevens voor evaluatie in de kanalen voor dataverzameling en de bijhorende databanken. Evaluatie kan baat hebben bij gegevens verzameld via een permanent monitoringsysteem. Dergelijk voordeel geldt overigens niet enkel met betrekking tot tussentijdse en ex post evaluaties. Ex ante evaluaties kunnen beleidsvoorbereiding en -bepaling ondersteunen. In dat kader kunnen periodiek en systematisch verzamelde gegevens evenzeer een belangrijke ondersteunende informatiebron vormen, bijvoorbeeld over kenmerken van de doelgroep van het beleid en de dienstverlening. Wanneer bepaalde evaluatievragen en -criteria frequent terugkeren, dient het systeem van monitoring daarop afgestemd te worden.

De operationele afstemming houdt ook in dat technieken voor dataverzameling op elkaar worden afgestemd. Monitoring is doorgaans geassocieerd met kwantitatieve technieken, terwijl evaluatoren vaak een gemengde aanpak volgen met kwantitatieve en kwalitatieve gegevensverzameling. Een afgestemd monitoringsysteem kan een belangrijke databron vormen om evaluaties te ondersteunen. Het operationeel stroomlijnen heeft ook consequenties voor het aspect van de registratie en rapportering van gegevens: op welk niveau dient men te registreren en wat is het aangewezen format om de gegevens aan te leveren? Qua timing moeten registratie en aanlevering afgestemd worden op het evaluatieproces dat op zijn beurt in het tijdsperspectief van te nemen beslissingen moet passen.

- Verantwoordelijkheden en betrokkenheid bij de uitbouw van informatiestructuren

Een belangrijke uitdaging ligt ook in de mate waarin managementverantwoordelijken betrokken kunnen zijn bij de uitbouw van monitoringsystemen die niet enkel meetmotieven vanuit een managementperspectief dienen, maar ook functioneel zijn om beleidsevaluaties te ondersteunen. Eén van de betrachtingen van een lerende organisatie is het ontwikkelen van

een capaciteit waardoor het mogelijk wordt om gangbare normen en assumpties in vraag te stellen, te analyseren en desnoods te veranderen. Het management is echter verantwoordelijk voor de voorbereiding en uitvoering van het beleid. In dat opzicht zijn ze voortrekkers van wat ze ondernemen. De vraag stelt zich dan ook of en hoe management en beleidsevaluatie kunnen samengaan?

Vanuit een rationeel en technisch perspectief dienen managementbeslissingen gebaseerd zijn op informatie en gericht op resultaten, en dan maken evaluaties integraal deel uit van managementprocessen. Anderzijds kan men organisaties ook vanuit een politiek en cultureel perspectief benaderen, waarbij de rol van evaluator en manager eerder in contrasterende kleuren geschetst worden, met oog voor een potentieel contrast tussen organisatiebelangen en maatschappelijke belangen.

In de lijn van deze vraagstelling kan men dan ook de keuze situeren tussen externe en interne evaluatie. De voor- en nadelen van beide vormen worden in deel 1 van de handleiding toegelicht. Zowel het theorie- als praktijkveld van evaluatie zijn nog niet klaar met het antwoord. Getuige daarvan het bestaan van verschillende benaderingen en standpunten op het vlak van de betrokkenheid van stakeholders in het evaluatieproces (cf. deel 1).

Een tendens die wel uit de debatten naar voor komt, is dat de betrokkenheid van managementverantwoordelijken zowel meer haalbaar als wenselijk is in het kader van evaluaties vanuit de motieven van het ondersteunen van planning, implementatie en institutionele verbetering en beleidsleren, in vergelijking evaluaties in het kader van verantwoording aan externe stakeholders.

Ook bij evaluaties in het kader van beleidsleren kan een grote betrokkenheid van het management een meerwaarde betekenen. Anderzijds kunnen ook aan deze evaluaties meer fundamentele vragen en consequenties vasthangen, bijvoorbeeld een beslissing over het al dan niet verder zetten van een beleidsmaatregel of -programma.

Samengevat, veel hangt af van het evaluatiemotief en van de heersende evaluatiecultuur in de organisatie.

3. Monitoring: een algemeen stappenplan

3.1 Het vertrekpunt: uitgeklaarde motieven en duidelijke beleidsdoelstellingen

Voor de ontwikkeling van een capaciteit en instrument voor het monitoren van beleid zijn twee voorwaarden noodzakelijk.

Ten eerste moet het motief om te meten duidelijk zijn. Het motief bepaalt voor een groot deel het inhoudelijke profiel van de indicatorenset, net zoals de evaluatievragen en -criteria dit voor een evaluatie doen. Het motief om te meten kan aansluiten bij een frequent evaluatieschema dat men wil invoeren. In dat geval vertrekt men van gemeenschappelijke motieven om te evalueren en te meten (in functie daarvan). Bij monitoring ter ondersteuning van planning zal het zwaartepunt liggen op omgevingsindicatoren: kenmerken van de doelgroep of het doelgebied. Monitoring ter verbetering van het management en de uitvoering zal oog hebben voor efficiëntie en dus voornamelijk geconcentreerd zijn op input- en outputindicatoren. In het kader van het verantwoordingsperspectief verschuift de klemtoon naar output en effecten.

Ten tweede dient het meetsysteem aan te sluiten bij duidelijk beleidsdoelstellingen. Voornamelijk om bestaand beleid op te volgen en te beoordelen aan de hand van gegevens uit monitoring, is het hebben van goede geformuleerde doelstellingen noodzakelijk voor het selecteren van de indicatoren.

Goede doelstellingen in het kader van meten beantwoorden aan het zogenaamde SMART-profiel: ze zijn specifiek, meetbaar, afgesproken of acceptabel, realistisch en tijdsgebonden.

Specifiek

Een doelstelling moet concreet zijn. Van belang is de wat?-vraag: wat willen we bereiken? Het antwoord op deze vraag verschaft een focus aan de doelstelling; of een beoogd resultaat of gewenste evolutie. Termen die in de formulering vervat zijn, moeten eenduidig zijn en begrepen worden door alle stakeholders. Wanneer een doelstelling specifiek geformuleerd is, wordt het ook makkelijker er uitvoering aan te geven (hoe?-vraag).

Meetbaar

Een doelstelling dient in meetbare termen geformuleerd. Alleen dan kan men progressie en de mate van doelbereiking opvolgen. Een doelstelling volgens het SMART principe is dan ook normerend: ze bevat een maat voor de kwaliteit en/of kwantiteit van wat vooropgesteld wordt.

Kwantificeren draagt sterk bij tot de meetbaarheid van een doelstelling. Actiewerkwoorden zijn ook beter observeerbaar en geven ook de intentie van wat men probeert te doen beter weer.

Sommige doelstellingen zijn in de praktijk moeilijker meetbaar dan andere. Moeilijk betekent echter nog niet onmogelijk. Daar komen we ook bij de eigenheid van indicatoren: ze zijn vaak een benadering, geven een indicatie van datgene wat men opvolgt. Het op onrechtstreekse wijze meten is meestal beter dan helemaal geen meetpunt hebben. Het komt er op aan om zo accuraat mogelijke indicatoren op te stellen. De doelstellingen dienen daartoe de aanzet te vormen.

Acceptabel

De doelstelling moet ook afgesproken of minstens acceptabel zijn onder de stakeholders. Er moet voldoende draagvlak zijn om de doelstelling na te streven. Voor beleidsdoelstellingen speelt uiteraard een politieke factor mee. Operationele doelstellingen behoren meestal tot de verantwoordelijkheid van het management, maar kunnen evenzeer vanuit het primaat van de politiek aangereikt worden. Ze kunnen bijvoorbeeld opgenomen worden in een beheersovereenkomst tussen een verzelfstandigde organisatie (agentschap, gemeentebedrijf,...) en de aansturende politiek verantwoordelijke (minister, college van burgemeester en schepenen). Een participatieve aanpak bij de formulering van de doelstelling kan het draagvlak vergroten. Operationele doelstellingen op korte termijn behoren idealiter tot de autonomie van de verantwoordelijke organisatie of afdeling.

Soms wordt de 'A' in SMART ook wel uitgelegd als "aanwijsbaar", waarbij men refereert naar de noodzaak van verantwoordelijkheid en verantwoording. Nog een andere toelichting voor de 'A' is "activerend" of "actiegericht", verwijzend naar een bijhorend actieplan. Hierin worden de beleidsinstrumenten en de processen bepaald in functie van de doelstellingen. Belangrijk is echter dat een doelstelling een bepaald resultaat voorschrijft, en niet zozeer de inspanningen. Deze laatste zijn voorwerp van het actieplan.

Realistisch

Doelstellingen moeten realistisch, d.w.z. haalbaar zijn. Wat theoretisch mogelijk of politiek wenselijk is, is niet altijd haalbaar in de praktijk. Kan men de hefbomen in beweging zetten om de doelstelling te halen? Heeft men voldoende kennis, capaciteit, middelen en bevoegdheden? Dit is belangrijk, anders komt de uitvoering in het gedrang. Een realistische doelstelling moet rekening houden met de praktijk. Dat neemt niet weg dat doelstellingen ambitieus en wervend mogen zijn. Het kan dan helpen om deze op te splitsen in subdoelstellingen of tussentijdse bakens. Soms wordt de 'R' in SMART ook wel uitgelegd als "relevant". Dit slaat op het verband tussen de doelstelling en de maatschappelijke behoefte of problematiek. Relevantie is tevens een evaluatiecriterium dat zowel bij ex ante als bij ex post evaluaties een belangrijke toetssteen kan zijn.

Tijdsgebonden

Een SMART doelstelling geeft ten slotte ook een tijdshorizon aan. De mate waarin een doelstelling realistisch is, hangt immers ook af van de vooropgestelde tijdspanne waarin men ze wil bereiken. Een concreet tijds kader is ook van belang om tussentijdse en/of ex post evaluaties er op af te kunnen stemmen.

Hieronder zetten we enkele voorbeelden van doelstellingen en geven we aan in welke mate met hun formulering aan het SMART principe is voldaan.

Doelstelling		S	M	A	R	T
1.	“Op 11 december 1997 werd het Kyoto-Protocol goedgekeurd door 158 landen, waaronder de EU-lidstaten. Hiermee verbinden alle ondertekenende industrielanden er zich toe om tegen 2008-2012 hun gezamenlijke broeikasgasemissies met 5,2% te verminderen ten opzichte van het niveau van 1990. De reductiedoelstelling behelst zowel koolstofdioxide (CO ₂), methaan (CH ₄) en lachgas (N ₂ O), de 3 belangrijkste broeikasgassen, als waterstoffluorkoolwaterstoffen (HFK's), perfluorkoolwaterstoffen (PFK's) en zwavelhexafluoride (SF ₆).”	+	+	+	+	+
2.	“Inzake verkeersveiligheid wil de regering het aantal verkeersslachtoffers verminderen met minstens 33 procent tegen 2006 en 50 procent tegen 2010. Daartoe zal de regering rekening houden met de aanbevelingen van de Staten-Generaal inzake verkeersveiligheid.”	-	+	+	?	+
3.	“Eind vorig jaar waren er 4.200.000 mensen aan het werk in België. Onze ambitie is dat er na afloop van deze legislatuur 4.400.000 moeten worden, waardoor de activiteitsgraad tot boven 65 procent zal uitstijgen.”	+/-	+	+	+	+
4.	“De Stad Gent wil met een ambitieus inclusieproject, Dubbelspoor, van start gaan: waarom zouden kinderen met speciale noden niet bij hun leeftijdsgenoten in het gewoon basisonderwijs school lopen? (...) Vier stedelijke basisscholen engageren zich ertoe om vanaf september 2003 niet langer leerlingen door te verwijzen en tegelijkertijd zelf kinderen op te nemen uit het buitengewoon basisonderwijs. Voor zover het aantal kinderen met een attest voor buitengewoon onderwijs de 10% niet overschrijdt.”	+/-	+/-	+	+/-	-
5.	“In 1997 besliste de Vlaamse regering tot een gevoelige uitbreiding van de Vlaamse bossen: 10.000 hectare ecologisch verantwoorde bosuitbreiding, als onderdeel van het Ruimtelijk Structuurplan Vlaanderen; en 10.000 hectare bebossing van landbouwgronden en braakliggende terreinen, onder meer met Europese steun. De termijn van realisatie valt samen met die van het Ruimtelijk Structuurplan: 2007.”	+	+	+	?	+
6.	“De FOD's streven binnen hun personeelsbestand naar 1200 personen met een handicap (2%). De federale overheid blijkt nu dit cijfer geenszins te halen. ”	+/-	+	?	?	-

Tabel 4: Beleidsdoelstellingen getoetst aan het SMART-principe

Doelstelling 1 beantwoordt aan het SMART principe: ze is voldoende specifiek, meetbaar, realistisch en tijdsgebonden. Bovendien wordt aangegeven dat er een ruime consensus over bestaat.

Doelstelling 2 kan nog wat specifiekere geformuleerd worden: welke soort verkeersslachtoffers bedoelt men: lichtgewond en/of zwaargewond, dodelijk? Het gaat hier duidelijk ook om een politiek engagement. De vraag is in welke mate de doelstelling realistisch is, dit wil zeggen is het ambitieniveau binnen de vooropgestelde termijn een haalbare kaart?

Doelstelling 3 beantwoordt in sterke mate aan het SMART principe. De formulering geeft aan dat er 200.000 jobs netto moeten bijkomen. Mogelijk is er nog discussie over het aantal voltijdse equivalenten.

Doelstelling 4 laat wat ambiguïteit bestaan: de doelstelling geeft een drempelwaarde aan met betrekking tot een ongewenste situatie, niet voor een gewenste situatie. Met andere woorden: wanneer in de praktijk helemaal geen inspanningen worden gedaan, er geen kinderen worden opgenomen uit het bijzonder onderwijs, is paradoxaal genoeg ook voldaan de doelstelling.

Doelstelling 5 is in meetbare termen gesteld, bevat een tijdselement en is specifiek geformuleerd. Het realisme van deze doelstelling beantwoorden ligt moeilijker. Gaat het om het aanleggen van 10.000 hectare nieuw bos, of een surplus op het totaal in vergelijking met de startsituatie in 1997? Bij het beoordelen van de inspanningen zal men rekening moeten houden in welke mate er ook bestaand bos verdwenen is.

Doelstelling 6 mist ten eerste een tijds kader. Tegen wanneer wil men de doelstelling halen? Bovendien kan men zich vragen stellen bij de formulering van de doelstelling. Om de cijfers te halen zouden FOD's personen van een handicap kunnen aanwerven los van elke ander criterium in verband met competenties, ervaring of scholing in functie van de beschikbare jobs.

Met het oog op het ontwerpen van een instrumentarium voor monitoring zijn niet enkel inhoudelijke noodzakelijke voorwaarden van belang. Er is ook het organisatieperspectief, of de gereedheid van de organisatie om een monitoringcapaciteit op te bouwen. Volgende elementen zijn daarbij van belang.

- De bevoegdheden van de organisatie zijn door het beleid duidelijk afgebakend
- De organisatie heeft een strategisch plan
- De organisatie kan haar operationele planning afleiden uit het strategische plan
- Medewerkers hebben een inbreng bij het planproces
- Het management gebruikt het strategische plan om de organisatie te sturen op hoofdlijnen
- De plannen bevatten SMART doelstellingen
- Het management vraagt en ondersteunt monitoring
- Verantwoordelijke medewerkers in het kader van monitoring kregen vorming rond de opzet en uitwerking van monitoring
- Er werd een verantwoordelijke aangewezen voor de coördinatie van de invoering van het monitoringsysteem
- Er is kennis in de organisatie aanwezig om verzamelde gegevens te interpreteren in het kader van het beleid en meer bepaald vanuit de achterliggende beleidstheorie

Tabel 5: Readiness checklist: is de organisatie klaar voor monitoring?

3.2 Het bepalen van de scope: wat meten we wel, en wat niet?

Een tweede stap is het verder afbakenen van het onderwerp van de meting. Hier stelt men zich de vraag waarover het monitoringsysteem eigenlijk informatie moet aanreiken. We beschreven reeds dat monitoring in de breedte werkt en evaluatie in de diepte. Op deze manier zijn monitoring en evaluatie complementair. Nochtans kan ook een monitoringsysteem niet alles vatten. Er moeten keuzes gemaakt worden over wat men wel en wat men niet zal meten. Deze keuze moet een bewuste keuze zijn. In deze paragraaf geven we enkele richtlijnen om de afbakening van het monitoringsysteem te structureren.

De afbakening bepaalt de afdeckingsgraad van het monitoringsysteem. Om tot een concrete afbakening van de meetpraktijken te komen zijn er twee aanknopingspunten:

- de werking van de organisatie
- de beleidsdoelstellingen

Deze aspecten worden gekoppeld in het schema van de beleids- en beheerscyclus als open systeemmodel (cf. supra: figuur 5).

Welk uitsnede van de organisatie en welke uitsnede van de nagestreefde beleidsdoelstellingen dienen door het meten in kaart gebracht te worden? Hieronder hernemen we deze vragen op een meer systematische manier.

Welke uitsnede van de organisatie zal gemeten worden?
<ul style="list-style-type: none"> - Welk deel van het organogram; alle diensten of partiële opdeling? - Welke input; bv. budget, personeel, infrastructuur, ...? - Welke processen en beleidsinstrumenten? - Welke output; welke 'producten' of dienstverlening?
Welke uitsnede van de beleidsdoelstellingen en de beleidscontext zal gemeten worden?
<ul style="list-style-type: none"> - Welk deel van het beleidsdomein; welke beleidsvelden? - Welke strategische en operationele doelstellingen? - Welke neveneffecten? - Welke aspecten van de beleidssetting (doegroep, doelregio)? - Welke externe factoren?

Tabel 6: Vragen voor de afbakening van het meetsysteem

3.2.1 Welke uitsnede van de organisatie zal gemeten worden?

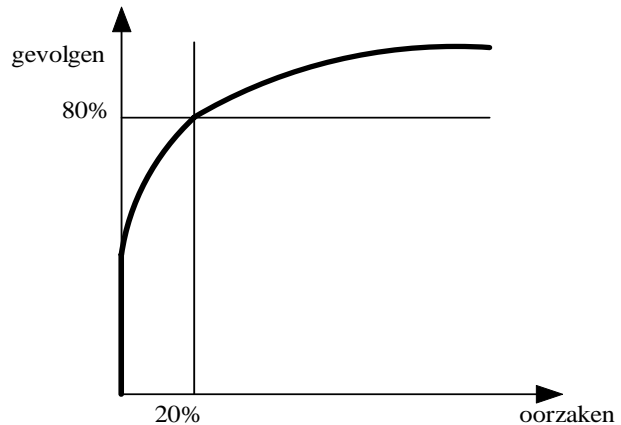
De studie van het organogram van de organisatie is een eerste praktische manier om de meetinspanningen af te bakenen. Men kan een keuze maken uit de verschillende diensten of afdelingen in de organisatie. Op die manier kunnen de meetinspanningen gericht worden op die diensten die van prioritair belang zijn vanuit het beleidsperspectief. Het gaat dan meestal om de uitvoerende diensten. Ondersteunende diensten (financiën, personeel) hebben immers geen directe impact op de nagestreefde beleidseffecten. Zo kan men bijvoorbeeld een opdeling maken tussen front- en backoffice; diensten die rechtstreeks in contact komen met de doelgroep en andere diensten.

Een tweede manier om de vraag "Wat meten we?" te beantwoorden is een inschatting van de mate van beslaglegging op de middelen of input. De gedachte hierachter is dat de middelen het relatieve gewicht weergeven dat de organisatie hecht aan een bepaalde output (product of dienstverlening). Middelen geven dan de belangrijkheid van een output aan. Onder input dient men niet alleen financiële middelen te verstaan, maar ook personeelsmiddelen, infrastructuur, enz.

Een methode om een aantal elementen uit te snijden uit het totaalpakket, is de Pareto analyse.

De Pareto analyse (Prokopenko, 1996)

De Italiaanse econoom Vilfredo Pareto merkte in de jaren '80 op dat 80 à 90% van de rijkdom van een land in de handen ligt van 10 à 20% van de bevolking. Dit is de oorsprong van de Pareto analyse die er vanuit gaat dat 80% van de gevolgen hun oorzaak vinden in 20% van de oorzaken. Dit is een empirische vaststelling. Het is niet het resultaat van een formele wetenschappelijke analyse. De regel blijkt m.a.w. vaak te gelden in de praktijk hoewel er geen rationele verklaring voor is. De 80:20 verhouding is slechts een schatting. Het kan 90:10, of 75:25, of nog een andere verhouding zijn. In het algemeen kan nochtans gesteld worden dat een beperkt aantal oorzaken de meerderheid van (als problematisch ervaren) gevolgen veroorzaakt.

**Figuur 7: de Pareto analyse**

Toegepast op de vraag “Wat te meten?” kan de Pareto analyse inhouden dat 80% van de input gebeurt in functie van 20% van de output. Zo zal een belastingdienst 80 % van haar middelen aanwenden om de typische inningdossiers af te werken, en zal de overige 20% naar een gamma nevenactiviteiten gaan zoals inspectie, beleidsadvies, kwaliteitscontrole, facilitaire diensten, stafdiensten. Men kan ervoor opteren om met zo weinig mogelijk indicatoren, een zo groot mogelijk deel van de begroting of van het personeel af te dekken. Door, toegepast op bovenstaand voorbeeld, een beperkte set indicatoren voor de typische dossiers te formuleren, kan 80% van het budget gedekt worden. Documenten als de begroting en personeelsplannen kunnen hier een nuttige leidraad vormen om de selectie te maken. Deze manier om de scope van het meetsysteem te bepalen is geschikt wanneer men wil meten en evalueren vanuit het motief van het verbeteren van het management of het afleggen van verantwoording in een politiekambtelijke context en van de uitvoerende naar wetgevende instanties.

Een graduele ontwikkeling van het monitoringsysteem betekent dat er na verloop van tijd wordt opgeschoven naar die soorten output die minder beslag leggen op de middelen. Visueel betekent dit dat men meer en meer opschuift naar rechts in de Pareto grafiek.

Een derde manier om een meetsysteem inhoudelijk af te bakenen is het selecteren van bepaalde processen of beleidsinstrumenten. Bijvoorbeeld in het kader van innovatief beleid, een nieuwe aanpak, kunnen stakeholders geïnteresseerd zijn in het opvolgen van de resultaten,

maar ook in de ervaringen van beleidsuitvoerders en/of doelgroep met de toepassing ervan. Deze invalshoek sluit ook aan bij de werking van de organisatie die een taakstelling heeft in het kader van de beleidsuitvoering. In het kader van subsidies zullen aanvraagdossiers administratief moeten verwerkt worden, voor handhaving zullen controles of inspecties worden uitgevoerd, adviezen opgesteld worden enz. Een gebiedsgerichte werking vereist samenwerking tussen verschillende actoren en instanties (overleg, uitwisseling van informatie,...). Het jaarverslag van een organisatie vormt vaak een goede bron voor het in kaart brengen van de processen en beleidsinstrumenten waarbij die organisatie betrokken is.

Een vierde ankerpunt binnen de organisatie om een meetsysteem af te bakenen vormt haar output. Het gaat om diensten en producten die de organisatie aflevert. Voorbeelden zijn: het aantal controles van ziekenhuizen, het aantal behandelde subsidiedossiers, het aantal afgeleverde identiteitskaarten, beleidsadviezen, sociale woningen, enzovoort. De output is logischerwijs sterk gekoppeld aan de processen binnen de organisatie. Output is echter vaak makkelijker te meten dan processen. Maar ook de meetbaarheid van output kan nog verschillen naargelang de concrete beleidscontext.

Men kan er dan ook voor kiezen om de output te meten die het meest eenvoudig te meten is. Daarvoor moeten men kijken naar de grijpbaarheid van de publieke dienstverlening. De output van een groendienst bijvoorbeeld is tastbaarder dan deze van een adviserende beleidscel. Het begrip grijpbaarheid van resultaten kan zichtbaar gemaakt worden door terug te vallen op de begrippen routine en ambiguïteit van taakuitvoering. Jobkenmerken van de uitvoerders van de publieke diensten zijn dus een aanwijzing voor de grijpbaarheid van de dienstverlening zelf. Grijpbare output impliceert eerder een hoge taakroutine terwijl minder grijpbare dienstverlening eerder een hoge taakambiguïteit heeft. Voor het inschatten van taakroutine en ambiguïteit kan men zich baseren op een vijftal jobkenmerken en bijhorende criteria (Bouckaert en Balk, 1988).

Jobkenmerken	Criteria	Continuüm
Variatie	Veranderingen in werkritme Veranderingen in locatie Veranderingen in fysische handelingen	nooit – constant nooit – constant nooit – constant
Autonomie	Keuze van taakvolgorde Keuze van werkritme Keuze van middelen	geen – volledig geen – volledig geen – volledig
Interactie	Noodzaak om afwezigheden te vervangen	dringend – geen
Kennis en vaardigheden	Leertijd	kort – lang
Verantwoordelijkheid	Kans op ernstige fouten Kennis over oorzaak en gevolg relaties Zichtbaarheid van het resultaat	hoog – geen gemiddeld – afwezig direct – afwezig

Tabel 7: Jobkenmerken, criteria en mogelijke waarden

Toch is dit geen algemeen aan te raden stelregel om zo het meetsysteem af te bakenen. De meetbaarheid zegt immers nog niets over de belangrijkheid van een output in het kader van de

werking van de organisatie of in het kader van de beleidsproblematiek en de beleidsdoelstellingen. Deze laatste vormen het tweede aanknopingspunt voor het afbakenen van het meetsysteem.

3.2.2 Welke uitsnede van de beleidsdoelstellingen en –context zal gemeten worden?

Een eerste manier om monitoring vanuit het beleid af te bakenen, is vergelijkbaar met het organogram van een organisatie. Men vertrekt van de structuur van een beleidssector of beleidsdomein. Het voordeel van deze benadering is de herkenbaarheid. De opdeling in beleidsvelden –deelgebieden van een sectoraal beleid- sluit gewoonlijk aan bij een institutionele realiteit. Het nadeel is dat beleidsvraagstukken zich zelden houden aan institutionele grenzen; binnen een beleidsveld kunnen meerdere organisaties verantwoordelijkheden hebben en sommige beleidsproblemen overstijgen de grenzen van beleidsvelden en soms zelfs beleidssectoren (bv. duurzame ontwikkeling, schoolachterstand). Wanneer men hiermee geen rekening houdt, kan een monitoringsysteem lijden aan een institutioneel kokerzicht.

Een selectie binnen een beleidsdomein blijft natuurlijk ook nog vrij algemeen. Binnen een beleidsveld kan men daarom verder de strategische en operationele beleidsdoelstellingen nemen om het meetsysteem concreet in te vullen. Eerder wezen we al op de samenhang tussen strategische doelstellingen en effecten enerzijds, en tussen operationele doelstellingen en output anderzijds. Hierin ligt ook de koppeling tussen de aanknopingspunten om de scope van het monitoringsysteem te bepalen; de output van de verantwoordelijke organisaties verwijzen in principe naar de operationele beleidsdoelstellingen.

Effecten zijn doorgaans ook meetbaar maar ook effectvariabelen verschillen onderling op het vlak van meetbaarheid naargelang de concrete beleidscontext; verkeersveiligheid is bijvoorbeeld makkelijker meetbaar dan welzijn van mensen. Korte termijneffecten zijn doorgaans beter meetbaar dan lange termijneffecten of impact. Zoals eerder gesteld is meetbaarheid van doelstellingen een 'conditio sine qua non' voor het koppelen van een meetsysteem.

Vaak zal men nog een keuze moeten maken welke doelstellingen men opvolgt binnen eenzelfde meetsysteem. Een mogelijke piste is te kiezen voor de output en effecten die het meest zichtbaar zijn voor de burger. Politici willen vaak zichtbare resultaten kunnen voorleggen aan hun kiespubliek. Mensen reduceren vaak het beleid en bestuur tot wat ze zelf ervaren of zien. Percepties en het managen van percepties wordt dan een belangrijk aandachtspunt voor beleidsmakers. Dergelijke afbakeningswijze is interessant in het kader van verantwoording naar de burger, maar lijkt minder passend in het kader van het verbeteren van de beleidsuitvoering of het evalueren van beleidseffectiviteit en beleidsleren. De politieke aandacht voor bepaalde beleidsissues kan ook richtinggevend zijn om het meetsysteem inhoudelijk af te bakenen.

Doelstellingen geven de gewenste hoofdeffecten aan. Het kan echter belangrijk zijn om ook neveneffecten op te volgen die men kan voorzien, zowel ongewenste als gewenste neveneffecten. Het is vooraf vaak moeilijk om neveneffecten te formuleren. Toch kan het interessant zijn om de belangrijkste verwachte neveneffecten mee te nemen als aanzet tot een integraal beleid.

Doelstellingen hebben vaak betrekking op kenmerken of situaties van doelgroep of doelgebied. De klemtoon van een meetsysteem kan dus worden gelegd op basis van de doelgroepen of op prioritaire regio's. Deze manier van afbakenen is zeker aan te raden wanneer het om een specifiek doelgroepenbeleid gaat. Zo kan de meting in een tewerkstellingsprogramma gericht worden op allochtonen, omdat we weten dat daar de meeste problemen zijn of omdat we daar nieuwe beleidsinitiatieven hebben genomen. Het is dus perfect te verdedigen dat de beschikbare 'meet'-middelen in eerste instantie op die doelgroep worden gericht.

Ten slotte dient men ook oog te hebben voor externe factoren die naast het beleid ook effecten kunnen beïnvloeden of ressorteren. Dit perspectief is met name van belang wanneer het meetsysteem moet bijdragen tot evaluaties vanuit het motief van verantwoording en beleidsleren, rond de criteria van doelbereiking en beleidsleren.

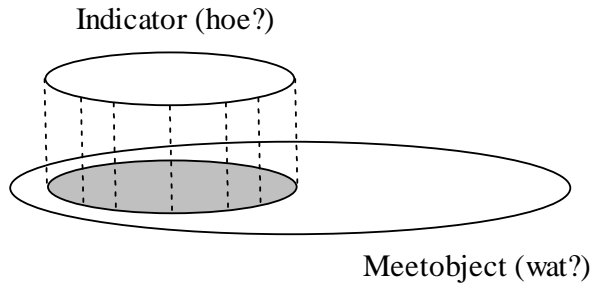
De genoemde aanknopingspunten en werkwijzen hebben elk hun voor- en nadelen. Anderzijds kan men ze vaak ook combineren. Zo kan men vertrekken van de strategische beleidsdoelstellingen (effecten) en verder selecteren op basis van het budgettaire gewicht van de operationele doelstellingen die eronder vallen.

Welk aanknopingspunt en perspectief men ook volgt, men dient te vermijden dat de niet gemeten uitsnede van de organisatie, het beleid of de beleidscontext helemaal uit het gezichtsveld verdwijnen. Anders ontstaat er een soort kokerzicht waarbij alleen nog datgene wat gemeten wordt onder de aandacht blijft en belangrijk wordt geacht.

3.3 Selecteren en opstellen van indicatoren: hoe meten we en wat willen we meten?

Wanneer de scope van het monitoringsysteem vast ligt, bestaat de volgende stap erin om de indicatoren te selecteren en op te stellen. Indicatoren zijn de basisbestanddelen van een meetsysteem. Een indicator is een kwantitatieve weergave van de kwantiteit of kwaliteit van een meetobject. Dat meetobject kan zoals eerder beschreven betrekking hebben op een component uit het schema van de beleids- en beheerscyclus van de overheid: input, processen, output, effecten of omgevingsfactoren. We kunnen dan ook overeenkomstige types van indicatoren

afleiden. In de fase van de indicatorenontwikkeling gaat men op zoek naar meetbare karakteristieken van wat in de vorige fase werd afgebakend. Het meetobject is het antwoord op de vraag wat we meten; een indicator geeft aan hoe we het meetobject meetbaar maken en meten. Dit is weergegeven in onderstaande figuur.



Figuur 8: De relatie van een indicator tot het meetobject

We benadrukken nogmaals dat een indicator enkel een indicatie geeft van de evolutie van de onderliggende realiteit. Men zal zelden de hele realiteit op een kwantitatieve wijze kunnen vatten. Hoewel dit vanzelfsprekend lijkt, zijn er verschillende praktijkvoorbeelden waar men deze fout maakt. Zo ziet men vaak de groei van het BNP als de enige maatstaf voor de welvaart van een land. Een ander voorbeeld is de IQ test, waar intelligentie gelijk gesteld wordt aan de IQ score. Een IQ score geeft een idee van het logische redeneervermogen. Dit is een belangrijk element van intelligentie. Er zijn echter nog andere vormen van intelligentie zoals het synthetische denkvermogen en de sociale intelligentie. Deze dimensies worden minder goed tot niet gemeten met IQ testen.

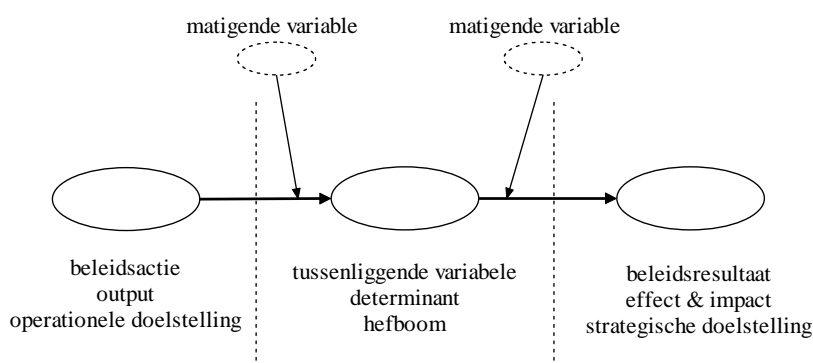
Om het meetobject in kaart te brengen zal men met andere woorden moeten terugvallen op meer dan één indicator. In veel gevallen dient men te beseffen dat zelfs een set van indicatoren nog een zekere reductie van de onderliggende realiteit of het meetobject bevat.

Wat is een goede indicator? Er bestaat geen toverformule voor het selecteren en opstellen van indicatoren. Wel kan men zich laten leiden door enerzijds een aantal 'meettechnische' criteria en anderzijds enkele aandachtspunten die betrekking hebben op het proces van selecteren en opstellen van indicatoren.

- Maak gebruik van de beleidstheorie

Beleidstheorieën vormen een belangrijke bron om beleidsevaluaties te bepalen (cf. deel 1). Hetzelfde geldt om de scope van de meetinspanningen te bepalen. Zeker voor het ontwikkelen van de individuele indicatoren vormt een beleidstheorie een belangrijk kader. Aan de hand van het logisch model kan men het implementatie- en veranderingsmodel achter een beleidsinitiatief in beeld brengen. Het

veranderingsmodel is in het bijzonder belangrijk voor het selecteren van output- en effectindicatoren, maar ook omgevingsindicatoren. Deze laatste kunnen slaan op externe factoren die eveneens effectvariabelen rechtstreeks kunnen beïnvloeden, maar ook op matigende variabelen (bv. kenmerken van doelgroepleden) die de causale relaties in het veranderingsmodel kunnen versterken of verzwakken.



Het veranderingsmodel van een complexer beleidsinitiatief zal bestaan uit meerdere beleidsacties (output), tussenliggende variabelen of determinanten en beleidsresultaten (effect en impact). Men dient dan indicatoren te ontwikkelen voor die componenten die betrokken zijn in de (politiek of financieel) meest gewichtige causale relaties (cf. supra).

- Zoek naar bestaande meetpraktijken

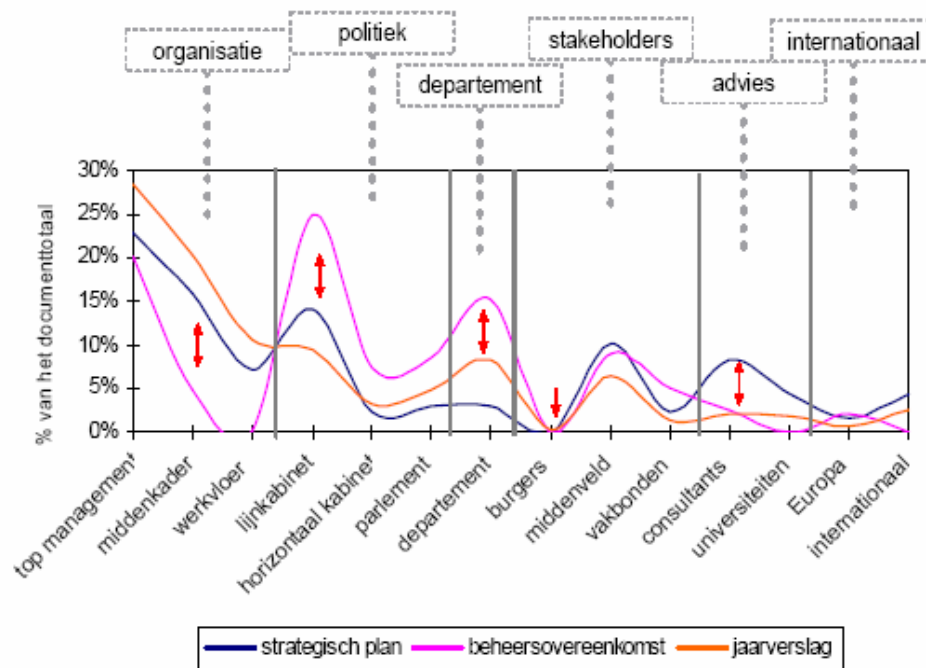
Men dient niet steeds het warm water uit te vinden in de zoektocht naar goede indicatoren. Bestaande meetpraktijken, bij vergelijkbare organisaties, overheden in binnen- of buitenland kunnen een goede inspiratiebron vormen. Deze indicatoren zullen zelden zomaar transposeerbaar zijn naar de eigen organisatie of beleidsveld. Maatwerk blijft dus geboden. Van bestaande praktijken kan men overigens ook leren op het vlak van datavergaring; gehanteerde technieken, ervaren knelpunten en beproefde oplossingen.

- Betrek de stakeholders

Betrek de stakeholders in het proces van indicatoreselectie, in de eerste plaats beleidsuitvoerders. Professionals uit de praktijk hebben een goed inzicht in het beleidsveld en de beleidsproblematiek. Zij zijn dan ook vaak het best geplaatst om tot de formulering van goede indicatoren te komen, op het vlak van relevantie, validiteit en begrijpbaarheid (cf. infra). Bovendien kunnen praktijkdeskundigen ook een beter inzicht geven in de onvermijdelijke beperkingen van een indicator. Op basis van een eerste voorstel kan er een consultatieronde worden georganiseerd. Dergelijk feedbackproces draagt bij tot een verfijning van de indicatoren alsook tot een breder draagvlak.

Overweeg om ook andere stakeholders (burgers, belangengroepen) te betrekken, wat niet altijd noodzakelijk is. Veel hangt af van de mate waarin men de indicator wil gebruiken in het kader van communicatie met de burger. In dat geval is begrijpbaarheid een belangrijk criterium (cf. infra).

De volgende figuur geeft aan hoe de betrokkenheid van verschillende actoren verschilt al naargelang de context (Van Dooren en Bouckaert, 2004). Het betreft hier surveygegevens van de Vlaamse Openbare Instellingen (N=34), over de betrokkenheid van actoren bij indicatorenontwikkeling in het kader van een drietal documenten: het strategisch plan, de beheersovereenkomst en het jaarverslag. Per document geeft de grafiek percentsgewijs de relatieve betrokkenheid van elke actor aan. Het percentage per actor resulteert uit het product van de intensiteit van betrokkenheid en het aantal maal dat een actor als 'betrokken' werd vermeld in de survey. Deze gewogen frequentie per actor werd ten slotte gedeeld door het documenttotaal (som van gewogen frequenties).



Figuur 9: De betrokkenheid van actoren bij de indicatorenontwikkeling in drie documenten – data uit de agentschappen van het MVG

- Specificiteit

De indicator heeft een 1op1-relatie met datgene wat gemeten wordt; een indicator mag slechts één aspect (aandachtspunt) van het meetobject meten. Anderzijds kan het

meetsysteem wel meerdere indicatoren bevatten om verschillende dimensies of aspecten van eenzelfde realiteit in beeld te brengen.

Zorg ervoor dat de frasering zelf van de indicator voldoende specifiek is. Aanduidingen als 'het aantal', 'het percentage van', 'de ratio van' etc. zijn een must in een indicator. Kwantificering is een noodzaak. Ook kwalitatieve inschattingen kunnen kwantitatief verwerkt worden. Een omschrijving als 'de deelnemers aan het opleidingsprogramma ervaren een meerwaarde' is niet voldoende specifiek. Beter is: 'het aantal deelnemers dat het opleidingsprogramma een score goed tot zeer goed geeft'. De mogelijke oordelen kan men bijvoorbeeld op een schaal van 0 tot en met 5 positioneren.

Stel voor elk onderdeel in de formulering van de indicator de vraag 'Wat bedoelen we nu exact?' en 'Hoe kan men dit interpreteren?'. Een kritische reflectie waarbij men zich in de plaats stelt van actoren met andere (beleids)visies, draagt bij tot het verfijnen van de indicator zodat de interpretatie geen problemen oplevert.

- Meetbaarheid

De opname van bepaalde indicatoren in de set kan impliceren dat gestart moet worden met het vaststellen en verzamelen van meetgegevens waarvoor dat voorheen nog niet gebeurde. Een indicator dient opgesteld in meetbare eenheden of hanteerbare categorieën. Dit criterium verwijst naar de puur technische haalbaarheid om de meetgegevens vast te stellen en te verzamelen.

- Afgesproken

De indicator steunt op een zo breed mogelijke consensus van alle stakeholders. Dit om polemieken te vermijden over de geschiktheid van de indicator op het moment dat men de indicator wil hanteren als databron voor evaluaties van het beleid.

- Relevantie

Dit lijkt een vanzelfsprekende voorwaarde, maar een wijziging van het meetobject kan gevolgen hebben voor de afbakening, berekening enz. van individuele indicatoren. Essentieel is ook rekening te houden met het feit dat vaak slechts de combinatie van meerdere indicatoren een relevant en evenwichtig beeld zal opleveren van datgene wat men wil meten.

- Tijdsgebondenheid

De relevantie van een indicator is geen vaststaand kenmerk maar kan variëren doorheen de tijd. Veranderingen met een impact op het meetobject (bv. wetgeving)

kunnen tot gevolg hebben dat andere indicatoren relevanter worden om hetzelfde op te volgen.

- Begrijpbaarheid

Zorg ervoor dat de indicatoren begrijpelijk zijn voor de doelgroep of gebruikers. Het betrekken van de stakeholders in de indicatorenontwikkeling draagt daar toe bij. Vooral wanneer het niet mogelijk is alle relevante stakeholders te betrekken, moet men aandacht besteden aan de begrijpbaarheid voor de doelgroep met het laagste kennisniveau over het meetobject.

- Databeschikbaarheid

Ga na of er data zijn voor een indicator. In het procesmatig scenario gaat men eerst na wat men wil meten (afbakening), hoe men wil meten (indicatoren) en zal men pas dan de databronnen overlopen. Het is niet omdat er nog geen gegevens beschikbaar zijn, dat de indicator niet goed is. Wel dient men te vermijden dat men voor het merendeel van de indicatorenset procedures voor verzameling dient te ontwerpen, tenzij het echt niet anders kan. Anderzijds verzamelt men vaak al heel wat informatie zonder dat deze gegevens echt ontsloten en gebruikt worden. Het nagaan van de status quaestionis op het vlak van dataverzameling kan de onvermoede beschikbaarheid van gegevens aan het licht brengen.

- Kosten-batenanalyse van de dataverzameling

Wanneer een indicator een nieuw te ontwerpen wijze voor dataverzameling noodzakelijk maakt, dient men de kosten en baten van de informatie die dat oplevert af te wegen. Dataverzameling impliceert immers meestal een bepaalde financiële kost en administratieve last. De meerwaarde van de extra informatie die de indicator genereert dient de meetinspanning te compenseren. De waarde van informatie kan men echter niet zo maar in geld uitdrukken; men zal dus een kwalitatieve afweging moeten maken. Het is raadzaam om dataverzameling zoveel mogelijk te stroomlijnen met bestaande informatiekkanalen en waar mogelijk in te bedden in standaardprocessen van beleidsvoering en het management van dienstverlening.

- Stabiliteit

Tegenover het criterium van tijdsgebondenheid staat de vereiste van continuïteit: dit geldt zowel voor individuele indicatoren, als voor de indicatorset als geheel; veelvuldige wijzigingen in de definitie van een indicator of een sterk verloop van indicatoren belemmert het opbouwen van tijdreeksen waardoor evoluties niet opgevolgd kunnen worden. Nochtans is permanentie één van de sleutelprincipes achter monitoring. Van

meet af aan dient men dus voldoende aandacht te besteden aan de concrete definitie van de indicator.

- Betrouwbaarheid van de meetprocedure

Een goede indicator levert ook betrouwbare gegevens op. De betrouwbaarheid is met name een kenmerk van de methode of procedure van meten. Welke garanties levert de meting op het vlak van objectiviteit en reproduceerbaarheid? Betrouwbaarheid heeft ook betrekking op het ijken van meettoestellen. De betrouwbaarheid neemt toe wanneer het aantal toevalsfouten in metingen afneemt.

De betrouwbaarheid van snelheidsmetingen door de politie.

Uit: Belgische Senaat, *Vragen en Antwoorden*, Bulletin 3-62, zitting 2005-2006.

Vraag

De correctionele rechtbank van Gent heeft recentelijk in een vonnis geoordeeld dat agenten die goed zijn opgeleid om een Multinova-toestel te bedienen bij hun vaststellingen van snelheidsovertredingen, geen marge van 6 % meer dienen in te bouwen.

Deze marge werd toegepast na een rondzendbrief van de procureurs-generaal. Daarin stond dat bij vaststellingen van snelheidsovertredingen rekening moest worden gehouden met een foutenmarge van 6 %, in het voordeel van de bestuurder.

Aan de andere kant zijn de toestellen die momenteel worden gebruikt om eventuele snelheidsovertredingen te meten, geïjkt.

1. Zal de geachte minister hieromtrent nieuwe richtlijnen uitvaardigen ?
2. Onder welke voorwaarden kan men beweren dat agenten « goed opgeleid zijn » om de Multinova-toestellen te bedienen ?
3. Bestaan er attesten voor deze opleiding ?
4. Welk standpunt neemt de geachte minister in tegenover de bewering dat de kilometeraanduiding in de wagens niet exact is ?
5. Welke impact zou de intrekking de « 6 %-regel » hebben op het ambtelijk apparaat op het vlak van administratieve verwerking ?

Antwoord

1. De ijking en homologatie van de toestellen en de sites wordt bepaald in het koninklijk besluit van 11 oktober 1997 betreffende de goedkeuring en homologatie van de automatisch werkende toestellen gebruikt om toezicht te houden op de naleving van de wet betreffende de politie over het wegverkeer en haar uitvoeringsbesluiten.

2. De opleiding van politieagenten om metingen te doen met behulp van Multinova-toestellen wordt gegeven door de dienst Metrologie van de FOD Economie (die instaan voor de ijking van de radars), tezamen met de constructeurs van de Multinova-toestellen zelf.

Bij de lokale politie zijn er momenteel zo'n 1500 mensen opgeleid om metingen met radartoestellen zoals Multinova of andere te doen.

3. De dienst Metrologie reikt aan de opgeleide politieagenten attesten met een erkenningsnummer uit.

4. Overeenkomstig de Europese normen mogen geen voertuigen op de markt gebracht worden met een foutenmarge waarbij de werkelijke snelheid hoger zou liggen dan de aangeduide snelheid. Een lichte afwijking van de snelheidsaanduiding is enkel toegestaan voor zover de snelheidsmeter een hogere snelheid aanduidt dan de werkelijke snelheid.

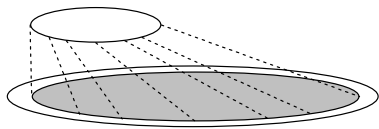
5. Indien er voor geopteerd zou worden om elke tolerantiemarge af te schaffen en derhalve een nultolerantie inzake snelheidsovertredingen in te voeren, die daarenboven streng wordt gehandhaafd, denk ik dat de werklast van het ambtelijk apparaat tijdelijk een piekmoment in de hoogte zal kennen. Wellicht zal die stijging van administratieve meerlast vrij snel terug dalen tot het normale peil, vermits de weggebruiker bij een hoge pakkans zijn rijgedrag relatief snel zal aanpassen aan het aangepaste handhavingsbeleid.

- Validiteit: meet men wat men wil meten?

Validiteit is een kenmerk dat maakt dat indicatoren daadwerkelijk een indicatie of meetbare vertaling vormen van het meetobject, zoals men het bedoelt. Dit bereikt men door aan de ene kant een goede conceptuele afbakening van het meetobject en aan de andere kant een accurate technische definiëring en concrete formulering van de gekoppelde indicatoren. Problemen met validiteit ontstaan wanneer meetgegevens op verschillende wijze geïnterpreteerd kunnen worden. Bijgevolg kunnen de conclusies over het meetobject sterk verschillen, want dan weer nefast is met het oog op de bruikbaarheid en invloed van beleidsevaluaties in het kader van beleidsbeslissingen.

Bij het ontwikkelen van indicatoren dient men op te letten voor enkele valkuilen op het vlak van validiteit:

- Overschatting van de dekkinggraad: men veronderstelt onterecht dat de indicator het hele meetobject afdekt.

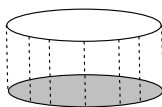


Figuur 10: Overschatting van de afdekkinggraad van een indicator

Voorbeeld: criminaliteitsstatistieken in steden als indicator van de veiligheid in steden.

Een deel van de criminaliteit ontsnapt aan de statistieken, ook verkeersveiligheid speelt een rol.

- Onvoldoende aandacht voor moeilijker meetbare aspecten van het meetobject: enkel de tastbare aspecten van het meetobject worden opgevolgd met indicatoren.

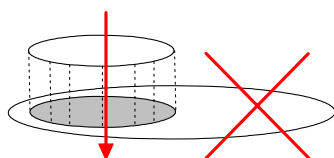


Figuur 11: Onvoldoende aandacht voor moeilijker meetbare aspecten

Voorbeeld: Citatie-indexen als indicator voor de prestaties van onderzoekers

De invloed van het werk van onderzoekers op het beleid is moeilijker te meten. De beleidsrelevantie van het werk is echter ook een belangrijk aspect, naast de outputkwantiteit (publicaties).

- Reductie van het meetobject tot de indicator: de aspecten van het meetobject die men niet opvolgt met indicatoren, krijgen geen aandacht meer bij het interpreteren van gegevens en het maken van conclusies over het meetobject.



Figuur 12: Reductie van het meetobject tot de indicator

Voorbeeld: criminaliteitsstatistieken als indicator van de onveiligheid.

Naast objectieve onveiligheid is er ook het subjectief onveiligheidsgevoel. Dit is moeilijker meetbaar, maar is wel belangrijk wanneer men praat over veiligheidsproblematiek.

- Stel een metafiche op

De operationalisering van het monitoringsysteem eindigt niet met de selectie van indicatoren. Voor elke indicator dient nog een reeks van kenmerken en eigenschappen nader bepaald te worden. Deze metagegevens zijn een essentieel aspect van het meetsysteem en leggen als het ware de regels voor het meten vast. Per indicator kan een metafiche met deze informatie opgesteld worden. Metagegevens zijn een onmisbaar 'tool' bij het beheer van het meetsysteem en hebben een belangrijke waarde voor de kwaliteit van het meetsysteem (cf. infra).

Relevante metagegevens van een indicator omvatten onder meer: *code, benaming, definitie, relevantie, type, berekeningswijze, meetniveau, meeteenheid of categorieën, bron, leverancier, meetfrequentie, aanleveringsfrequentie, tijdreeks, normering, beïnvloedende factoren, audit op de kwaliteit van meetgegevens, presentatie, opmerkingen, datum opname in set, doorgevoerde wijzigingen*. Volgende tabel bevat toelichting bij een aantal van de genoemde metagegevens.

Kenmerk indicator:	Toelichting:
oCode / nummering	Bijvoorbeeld een lettercijfercombinatie die men kan hanteren in rapportages in het kader van het gebruik van de monitor. Om de eenvoud in nummering te bewaren, worden bij veranderingen in de set van indicatoren (aanpassing, schrapping, toevoeging van indicatoren) vrijgekomen nummers best opnieuw gebruikt. Het zijn immers enkel volgnummers, zonder achterliggende gradatie in belangrijkheid, enz.
oBenaming	Naam van de indicator in een kernachtige formulering.
oDefinitie	Meer omstandige omschrijving en/of concretere afbakening van de indicator, zodat men de betekenis ervan kan begrijpen.
oRelevantie	Motivering waarom de indicator deel uitmaakt van de set, met name: - het belang van de indicator voor het gemeten aspect van het meetobject; - de veronderstelde indicatie of informatie die de indicator geeft.
oType	Classificatie van de indicator binnen de typologie van indicatoren, geënt op de beleidscyclus: input – proces – output – effect - omgeving
oBerekeningswijze en -criteria	De manier waarop de waarde van de indicator bepaald wordt, d.w.z. hier wordt aangegeven welke informatie in aanmerking komt en welke niet. Deze rubriek vormt aldus een aanvulling op de definitie van de indicator en zorgt mee voor de eenduidige afbakening en interpretatie, zowel inhoudelijk als in de tijd. Bv.: De indicator 'personeelsbestand gemeente', verwijst naar het absolute aantal of het aantal VTE personeel werkzaam bij de gemeentelijke overheid, exclusief politie-, brandweer- en onderwijspersoneel, en naar het effectief bestand op tijdstip t.
oMeetniveau	Er zijn 3 meetniveaus: - <i>nominaal of categorisch</i> → de meetgegevens of waarden op basis van de indicator geven een eigenschap van het meetobject aan (identiteitsrelatie) - <i>ordinaal</i> → de meetgegevens of de waarden op basis de indicator duiden op een rangschikking (orderrelatie) - <i>metrisch</i> → de meetgegevens of de waarden op basis van de indicator duiden op een rangschikking met vaste afstanden (intervallen). De aard van het meetniveau heeft consequenties voor de mogelijkheden voor statistische bewerkingen bij de analyse van de verzamelde meetgegevens, en bijgevolg voor de (grafische) voorstellingswijze van de meetresultaten.
oMeeteenheid of –categorieën	Bij indicatoren met een metrisch meetniveau: de maat of grootheid waarin de meetgegevens worden uitgedrukt. Bij indicatoren met een nominaal of ordinaal meetniveau kunnen hier de categorieën opgesomd worden, indien men beschikt over een vooraf gekende, exhaustieve lijst van mogelijkheden.

Tabel 8: Technische fiche voor een indicator

Kenmerk indicator:	Toelichting:
<ul style="list-style-type: none"> o Bron 	<p>De databron is tweeledig: -De actor die primair de meetgegevens, d.w.z. de waarden op de indicator, vaststelt, registreert, die de meting uitvoert. -Het gegevensbestand of document waarin de officiële meetwaarden vastgelegd worden.</p>
<ul style="list-style-type: none"> o Leverancier 	<p>De actor die instaat voor het doorgeven van de meetwaarden naar de centrale databank van het monitoringsysteem. Dit is niet noodzakelijk dezelfde actor die de meting uitvoert. Het in kaart brengen van de betrokken actoren wanneer data-aanlevering trapsgewijs verloopt, is belangrijk om administratief dubbelwerk te vermijden. Mogelijk dienen dus afspraken gemaakt met leveranciers rond gedeeld eigenaarschap van meetgegevens.</p>
<ul style="list-style-type: none"> o Meetfrequentie 	<p>Frequentie waarmee meetwaarden op de indicator opnieuw vastgesteld worden, bv. kwartaal, maandelijks, jaarlijks.</p>
<ul style="list-style-type: none"> o Aanleverings-frequentie 	<p>Frequentie waarmee de databank geüpdatet wordt met nieuwe meetgegevens. De aanleveringsfrequentie kan lager zijn en maximaal gelijk zijn aan de meetfrequentie.</p>

Vervolg tabel 8: Technische fiche voor een indicator

Samengestelde indicatoren: ratio's en indexen

Indicatoren kunnen ook bestaan uit meerdere componenten. Twee indicatoren kunnen in verhouding tot elkaar worden gesteld; in zogenaamde ratio's. Voorbeelden hiervan zijn:

Zuinigheid:	input / input
Efficiëntie:	input / output
Productiviteit	output / input
Effectiviteit	effect / output
Kosteneffectiviteit	effect / input

De kostprijs van een flitscamera is een indicator voor zuinigheid. De doorlooptijd van een vergunningsdossier bij de milieudienst zegt iets over de efficiëntie van deze dienst: de tijdsbesteding in uren of dagen is de input, het afgewerkte dossier met beslissing is de output. De productiviteit van de milieudienst kan uitgedrukt worden in het aantal dossiers dat men per maand afwerkt. De input kan worden gemeten in voltijds equivalenten (vte). Het aantal lessen aan deelnemers aan een beroepsopleiding is een outputindicator. Nuttiger is echter het aantal lessen per plaatsing van een werkzoekende in een duurzame job. Dit is een maat voor effectiviteit. De kosteneffectiviteit van het bijgestuurd verkeersveiligheidsbeleid zou men kunnen definiëren door de daling in verkeersslachtoffers in verhouding te stellen tot de budgettaire middelen voor intensievere handhaving en uitgevoerde verbeteringswerken aan het wegennet. Wat is bijvoorbeeld de kost om het aantal slachtoffers met 10 per jaar te doen dalen?

Verschillende indicatoren kunnen ook samengevoegd worden tot een index. Door de waarden van verschillende indicatoren te aggregeren, komt men tot een globaal cijfer. De waterkwaliteit van een rivierbekken kan gemeten worden aan de hand van het stikstofgehalte, het visbestand en het zuurstofgehalte in het water. Deze drie indicatoren leveren meetgegevens op die vervolgens samengevoegd kunnen worden tot een index van de waterkwaliteit. Aggregatie van gegevens kan ook gebeuren over meeteenheden heen; zo kan men op basis van de scores van de rivierbekkens een totaalscore bekomen voor de waterkwaliteit van de Vlaamse waterlopen.

Belangrijk voor het construeren van een index is het gewicht dat elke indicator krijgt die er deel van uitmaakt. In feite stelt zich hier een gelijkaardige problematiek als bij het evalueren op verschillende criteria, met behulp van multicriteria-analyse (cf. deel 3). Om te kunnen komen tot een eindoordeel of syntheseconclusie dient men elk evaluatiecriterium een bepaald gewicht toe te kennen.

Aggregeren heeft voor- en nadelen. Men kan verschillende inhoudelijke aspecten combineren en/of gegevens van verschillende meeteenheden (personen, locaties, organisaties) samenvoegen. Dit levert een omvattend beeld op van het meetobject ofwel een overzicht op een problematiek. De keerzijde is een verlies aan informatie; details gaan verloren, hetzij per inhoudelijk aspect, hetzij per meeteenheid. Het uitsplitsen van data kan dus even belangrijk zijn dat het aggregeren ervan. Belangrijk is dat het laagste niveau waarop men gegevens wil uitsplitsen bepaalt op welk niveau men de gegevens moet registreren, d.w.z. de eigenlijke metingen moet uitvoeren.

3.4 Ontwikkelen van normen: wat is een goed meetresultaat?

Een meetresultaat op zich zegt niet alles. De vastgestelde meetwaarden op indicatoren leveren beschrijvende informatie aan. Daarin ligt het kenmerkende maar tegelijk ook de beperking van monitoring. Evaluatie heeft andere aspiraties. De essentie van evalueren is het beoordelen van de meerwaarde van het evaluandum. Monitoring kan een belangrijke informatiebron zijn voor gegevensverzameling in het kader van evaluaties. De beschrijvende informatie verkregen uit het meetsysteem dient vervolgens afgetoetst te worden tegen een evaluatief referentiekader. Het referentiekader verwijst niet enkel naar de evaluatievragen en –criteria, maar ook naar normen. Wat is een goed meetresultaat en wat niet? Deze vraag kunnen we beantwoorden door het meetresultaat af te zetten tegen of te vergelijken met een norm.

We meten bijvoorbeeld het aantal hectare bos in Vlaanderen (150 000 ha). Is dit veel of weinig? Voor veel mensen is dit cijfer abstract. We kunnen het meetresultaat beter inschatten en beoordelen als we het kunnen vergelijken met een norm. In deze paragraaf gaan we verder in op normstelling. We vertrekken van volgende twee vragen:

- Welke soorten normen bestaan er?
- Waarop kan een norm gebaseerd zijn?

3.4.1 Soorten normen

Voordat we de methoden voor normzetting bespreken, moeten we eerst een aantal termen verhelderen; normen, streefcijfers en standaarden. Het gaat eerder om nuances in betekenis. In essentie gaat het telkens om hetzelfde: een ijkpunt om de gemeten waarde af te toetsen. Normen zien we als de verzamelnaam voor al deze ijkpunten. Streefcijfers hebben een minder strenge connotatie: ze vormen een richtpunt of target, een doel om naartoe te werken. Streefcijfers zijn normen die men in de toekomst wil halen. Andere normen kunnen permanent gelden. Een norm wordt een standaard wanneer de norm algemeen aanvaard is. De

zogenaamde ISO standaarden die minimale kwaliteitsvereisten voor producten en diensten aangeven, zijn een bekend voorbeeld van een standaard.

Types van normen kunnen gekoppeld worden aan het schema van de beleids- en beheerscyclus. Zo kan een norm betrekking hebben op input, processen, output of effecten, maar ook geformuleerd worden op het vlak van zuinigheid, efficiëntie, doelbereiking, effectiviteit, kosteneffectiviteit, enzovoort.

De formulering van de norm kan ook verschillen. Het kan gaan om een absolute waarde, een relatieve verandering, een grenswaarde (minimumdrempel of maximum), of nog om een marge afgebakend door grenswaarden. In het werkgelegenheidsbeleid kan men een norm stellen voor het effect van dat beleid, in termen van arbeidsplaatsen die worden gecreëerd. Stel dat er in een land momenteel 2,3 miljoen banen zijn. Een absolute norm stelt bijvoorbeeld dat men 2,5 miljoen jobs wil hebben op de arbeidsmarkt tegen een bepaald toekomstig tijdstip. Een relatieve verandering stelt bijvoorbeeld een toename met 5% voorop ten opzichte van de huidige situatie. Die 2,5 miljoen kan ook als een minimumdrempel beschouwd worden. In dit voorbeeld lijkt een maximumwaarde als norm dan weer niet relevant. Maximumwaarden zijn wel gebruikelijk in het kader van een nitraatrichtlijn, de concentratie van ozongassen in de lucht, enz.

Een norm slaat echter niet altijd op een kwantiteit. Men kan ook normen formuleren die kwalitatief van aard zijn, bijvoorbeeld met betrekking tot te volgen procedures, omgang met de doelgroep van het beleid of klanten van de dienstverlening, toegangsmodaliteiten, enz. Een beleidscel kan kwaliteitsvolle adviesverlening aan de minister als streefdoel hebben. Ook het tijdsaspect kan centraal staan in de norm wanneer het gaat om wacht- en doorlooptijden. Voor een dienst ruimtelijke ordening zijn bepaalde termijnen van belang voor de goede voortgang van vergunningsdossiers. Een norm bij een milieudienst kan zijn om 90% van de adviezen binnen een bepaald tijdsbestek te bezorgen aan de aanvragende instantie.

Normen zijn niet altijd expliciet. Om meetresultaten te kunnen beoordelen in het kader van een evaluatie is het echter belangrijk dat de norm gekend is. In het kader van beleid hebben veel normen betrekking op output en effecten. De normstelling kan dan geïntegreerd worden in de bepaling van de strategische en operationele beleidsdoelstellingen. We wezen al op het belang van duidelijke doelstellingen als vertrekpunt voor het ontwikkelen van een meetsysteem om het beleid te kunnen monitoren en evalueren.

3.4.2 Waar komen normen vandaan?

We bespreken hier een aantal methoden om normen te formuleren. De methoden hebben specifieke voor- en nadelen, en het is dan ook aan de beleidsmakers en managementverantwoordelijken om een keuze te maken. Deze keuze hangt mede af van de specifieke beleidsector en de taakstelling van de betrokken organisaties.

Normenstelling is de taak van beleidsmakers en het management. Het primaat van de politiek kan ook een eerste bron zijn voor het bepalen van de norm. Een 'politieke' norm is een norm die vooropgesteld wordt omwille van het wervende of ideologische karakter. Politiek voluntarisme ligt dan aan de basis. Voorbeelden zijn '100% bereik voor huisbezoeken in de preventieve kindercare' voor Kind & Gezin, en '0 verkeersslachtoffers' van de Zweedse overheid. Dergelijke resultaten zullen wellicht nooit bereikt worden. Om ideologische redenen moet dit echter het streefcijfer zijn. Men kan niet op zijn lauweren rusten bij een suboptimaal resultaat. De normen zijn streefcijfers die de optimale situatie weerspiegelen.

In deel 1 hebben we verschillende maatstaven besproken om beleid te evalueren op doelbereiking. Een aantal maatstaven bevat de optimale situatie als onderdeel van de vergelijkingsbasis. Merk op dat politieke en ideologische normen niet noodzakelijk onhaalbaar zijn. Ze kunnen perfect opgenomen zijn in een doelstelling die aan het SMART-principe beantwoordt (cf. supra).

Een tweede mogelijkheid is dat men vanuit beleidsoogpunt een status quo of verbetering als norm vooropstelt ten opzichte van een eerdere toestand of situatie: dit gebeurt bij het vergelijken doorheen de tijd. De norm is dan de waarde die werd gemeten op een eerder moment in een beschikbare of geplande tijdreeks van meetwaarden.

Ten derde kan een norm ook gebaseerd zijn op metingen die betrekking hebben op andere overheidsorganisaties en hun beleid; in dezelfde of een andere beleidsector, in binnen- of buitenland. In bepaalde gevallen zal men ook de private sector vergelijken. De norm kan dan het gemiddelde zijn, het topkwartiel of de best presterende.

Meer en meer is er sprake van benchmarking. Overheden worden voortdurend met elkaar vergeleken. De vergelijkingen betreffen niet enkel het milieu, de perceptie van corruptie of economische zaken. Ook bij onderwerpen als de vermindering van het aantal verkeersslachtoffers, de uitgaven voor geneesmiddelen en het werkgelegenheidsbeleid (bv. de loonmatigingsnorm) duiken internationale vergelijkingen op in de debatten. Vergelijkingen betreffen zowel het openbare bestuur zelf als de beleidsectoren waar de overheid een rol speelt. Een brede waaier aan internationale organisaties en adviesbureaus compileren diverse internationale rangschikkingen. De media maken er nieuws van en ministers stofferen hun beleidsnota's ermee.

Ten vierde zijn er wetenschappelijke standaarden. De dioxinenorm is een norm voor de indicator "dioxinegehalte" die op een wetenschappelijke manier is vastgelegd. Wetenschappelijke standaarden worden vaak als kwaliteitsnorm gekozen in het kader van marktregulering en dit in verschillende beleidssectoren.

De onderstaande tabel geeft voor de verschillende methoden en bronnen voor normen voorbeelden uit de Vlaamse praktijk. Ze somt telkens ook een aantal voor- en nadelen van gehanteerde techniek of bron toe.

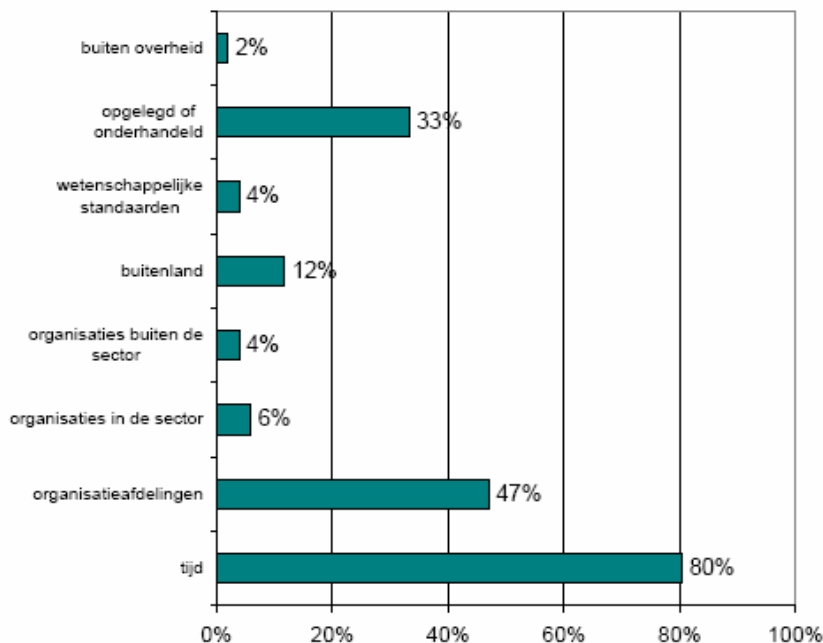
Methoden / Bron	Beoordeling van de voor- en nadelen	Voorbeeld
politieke of ideologische norm	<ul style="list-style-type: none"> - sterk gedragen, verhoogt acceptatie van het gehele meetsysteem - wervende normen, symboolnormen - vaak geïnspireerd door onderstaande bronnen (bv. internationale vergelijking) - niet altijd even realistisch, (maar niet noodzakelijk onrealistisch) 	<ul style="list-style-type: none"> - het verhogen van de werkzaamheidsgraad van 61,4% naar 66,5% tegen 2004 en 70% tegen 2010 (Lissabon doelstelling) - "De Vlaamse en federale regering hebben al laten horen dat ze het aantal verkeersslachtoffers tegen 2005 met een derde willen verminderen." www.radio1.be - "De Vlaamse overheid stelt zich tot doelstelling een aandeel van 3% groene stroom in het totale energieaanbod tegen eind 2004 en een aandeel van 5% in het globaal elektriciteitsverbruik tegen het jaar 2010." (Beleidsnota energie)
Eerdere meting	<ul style="list-style-type: none"> - geschikt om unieke beleidsinitiatieven te beoordelen - geschikt indien er geen vergelijkbare organisaties zijn - omgevingsvariabelen kunnen verstorend werken - geschikt voor vertrouwelijke informatie 	<ul style="list-style-type: none"> - aantal jongeren in bijzondere jeugdbijstand is niet toegenomen - aantal gedwongen plaatsingen / totaal aantal plaatsingen naar 40% (op basis van een tijdsreeks)
Vergelijken met andere organisaties in de sector	<ul style="list-style-type: none"> - geschikt om beleidsresultaten te vergelijken - omgevingsvariabelen spelen minder indien organisatie in quasi dezelfde omgeving opereert. 	<ul style="list-style-type: none"> - de stressindex van het personeel van de federale ambtenaren bedraagt 18,81; die van departement LIN 19,14 en die van het departement onderwijs 21,85; de ambtenaren van stad Antwerpen hebben een stressindex van 18,48 wat overeenkomt met het Belgisch gemiddelde - het ziekteverzuim bij de Post is 9,4%, bij het Ministerie van de Vlaamse Gemeenschap 4,74%
Vergelijken met andere organisaties buiten de sector	<ul style="list-style-type: none"> - geschikt om beheersresultaten te vergelijken - sectoroverschrijding kan nieuwe impulsen en ideeën opleveren - moeilijkere vergelijkbaarheid 	<ul style="list-style-type: none"> - het ziekteverzuim in de privé sector is 4,48%. In het Ministerie van de Vlaamse Gemeenschap is dit 4,74%

Tabel 9: Methoden voor normbepaling: voor- en nadelen en voorbeelden uit de Vlaamse praktijk

Methode / Bron	Beoordeling van de voor- en nadelen	Voorbeeld
Vergelijken met buitenlandse organisaties (en organisaties uit andere regio's)	- geschikt voor monopolisten die geen interne vergelijkingspunten hebben - mogelijke leereffecten zijn groot - vergelijking moet nationale verschillen in acht nemen	- vergelijking van het Vlaamse onderwijs met OESO landen waarbij de internationale vergelijking de norm bepaalt - Uit de beleidsnota buitenlands beleid; "De minister voor ontwikkelingssamenwerking engageert zich dat de inspanningen op het terrein van sensibilisatie minstens 10% van zijn budget moeten beslaan. Dit is conform de internationale normen zoals die door de OESO worden naar voor geschoven".
Vergelijken wetenschappelijke standaarden	- goed gefundeerd - technisch, risico op technocratie	- nitraatrichtlijn: 50mg nitraat per liter - vaccinatiegraad nodig om een infectieziekte uit te roeien

Vervolg tabel 9: Methoden voor normbepaling: voor- en nadelen en voorbeelden uit de Vlaamse praktijk

Vergelijking doorheen de tijd blijkt de meest gebruikte methode te zijn (Van Dooren en Bouckaert, 2004). Het is ook de meest voor de hand liggende methode. Data zijn in de organisatie beschikbaar. Heel wat variabelen in de organisatie en de omgeving blijven constant en verstoren de interpretatie van de resultaten niet. De gegevens in onderstaande figuur geven dit aan. We denken dat echter dat ook andere methoden nuttig kunnen zijn. Waarschijnlijk worden deze methoden onvoldoende benut.



Figuur 13: Meest gebruikte methoden voor normzetting in de afdelingen van het Ministerie van de Vlaamse Gemeenschap

3.5 Meten

De volgende stap is het meten zelf: het registreren en verzamelen van de meetgegevens op de geselecteerde indicatoren. Om te meten heeft men de keuze uit verschillende instrumenten. In deel 3 van deze handleiding komen diverse technieken om te registreren en te meten aan bod. In deze paragraaf gaan we dieper in op de aspecten die van belang zijn met het oog op de kwaliteit en accuraatheid van metingen.

Gegevens over beleidseffecten hebben een relatief beperkt aantal bronnen. Deze bronnen leveren meetinformatie op wanneer gegevens worden geproduceerd aan de hand van een set bepaalde, systematische bewerkingen of procedures. Voor sommige effecten of effectvariabelen kunnen evaluaties terugvallen op meetinstrumenten die al ontworpen en geaccepteerd zijn voor dat doel in het respectieve beleidsveld.

Met betrekking tot andere effecten zijn er eveneens kant-en-klare instrumenten beschikbaar, evenwel maar bestaat er geen consensus over welke procedure of techniek het meest aangewezen is. In dat geval dient de evaluator een beredeneerde keuze te maken. Praktische hanteerbaarheid, duurtijd en kostprijs zijn uiteraard belangrijke criteria bij deze afweging. Echter, het meest fundamentele selectie criterium is de mate waarin het instrument overeenkomt met wat de evaluator wil gaan meten. Een accurate beschrijving van de te meten effecten is dan ook een noodzakelijke voorwaarde. Bovendien mag men zich niet zomaar laten leiden door de benaming van het instrument. Verschillende meetinstrumenten voor eenzelfde concept of fenomeen hebben vaak een verschillende inhoud en theoretische oriëntaties die hen meer of minder geschikt maken voor een concrete toepassing.

Voor heel wat effecten die van belang zijn voor evaluatoren, bestaan er geen vastgestelde meetinstrumenten noch een reeks kant-en-klare maatstaven waaruit gekozen kan worden. In dergelijke gevallen moeten meetinstrumenten en indicatoren ontworpen worden (cf. supra). Daarvoor ontbreken soms voldoende tijd en middelen. Een aantal ad hoc technieken, bijvoorbeeld het halen van specifieke gegevens uit officiële records van gekende kwaliteit, zijn evident om beschouwd te worden als aanvaardbare meetpraktijken zonder verder bewijs. Andere meettechnieken echter, zoals vragenlijsten, attitudeschalen, kennistesten en schema's voor het coderen van systematische observaties, zijn minder vanzelfsprekend.

Het opstellen van dergelijke instrumenten opdat ze meten wat ze moeten meten op een consistente manier, is niet altijd gemakkelijk. Daarom zijn er een aantal geijkte procedures om instrumenten te ontwikkelen, die een aantal technische aandachtspunten inhouden en vaak een redelijk deel van piloottesten, analyse, revisie en valideren vereisen vooraleer dergelijk nieuw ontwikkeld instrument met vertrouwen kan worden ingezet. Wanneer geen rekening wordt gehouden met deze aandachtspunten en procedures, kan het meetinstrument weliswaar een zekere waarde hebben. Maar het zal niet noodzakelijk geschikt zijn om op een accurate wijze beleidseffecten te beoordelen.

Het is daarom aangewezen dat, wanneer de ontwikkeling van een instrument niet volgens de geijkte procedures kan gebeuren, er minstens toch een controle is van de elementaire meeteigenschappen vooraleer er een bepaald belang aan wordt toegekend in het kader van een evaluatie.

Bij het meten staan volgende meetkenmerken of kwaliteitsmaatstaven centraal: de betrouwbaarheid van het meetinstrument, de validiteit of geldigheid van de meting en de sensitiviteit van de maatstaf. Deze trits van kenmerken is essentieel om te komen tot goede metingen. Correcte gegevens zijn op hun beurt cruciaal om kwaliteitsvolle evaluaties uit te voeren.

3.5.1 Betrouwbaarheid van het meetinstrument

Een eerste kritische slaagfactor in het kader van een goede meting is de betrouwbaarheid van het meetinstrument of de meettechniek. De betrouwbaarheid is de mate waarin het instrument dezelfde resultaten produceert wanneer het herhaaldelijk gebruikt wordt om hetzelfde te meten. Variatie in deze resultaten wijst op een meetfout. Een brievenweegschaal, bijvoorbeeld, is betrouwbaar in de mate dat ze dezelfde score (gewicht) aangeeft voor eenzelfde enveloppe op meerdere tijdstippen.

Geen enkel meetinstrument, classificatieschema of tellingprocedure is perfect in termen van betrouwbaarheid, maar verschillende types van meettechnieken hebben een verschillende graad van betrouwbaarheidsproblemen. Metingen van fysieke kenmerken waarvoor standaard meetinstrumenten beschikbaar zijn, zoals bijvoorbeeld hoogte en gewicht, zullen algemeen meer consistent zijn dan metingen van psychologische kenmerken, zoals bijvoorbeeld intelligentie gemeten aan de hand van een IQ-test. Deze laatste is dan weer betrouwbaarder dan metingen die beroep doen op het geheugen van personen, bijvoorbeeld om gezinsuitgaven voor consumptie na te gaan.

Voor evaluatoren ligt een voornaam bron van onbetrouwbaarheid in de aard van meetinstrumenten die gebaseerd zijn op de antwoorden van deelnemers op geschreven of mondelinge vragen die door onderzoekers worden gesteld. Verschillen in de test- of meetcontext, verschillen tussen observatoren of interviewers bij het hanteren van het instrument (bijvoorbeeld de vragenlijst) en zelfs variaties in de gemoedsgesteldheid van de respondenten vormen aanleidingen tot onbetrouwbaarheid.

Het ongewenste effect van onbetrouwbaarheid in metingen is het doen verwateren of verbergen van reëel bestaande verschillen of veranderingen. Een beleidsmaatregel die daadwerkelijk effectief is, kan door een onbetrouwbare meting van de effecten minder effectief blijken dan hij in werkelijkheid is. De meest rechtstreekse wijze waarop een evaluator de betrouwbaarheid van een potentiële effectindicator kan controleren is het minstens tweemaal uitvoeren van de meting in een context waarin het effect normaal niet verandert, d.w.z. er treedt geen verandering op in de waarde van de effectvariabele. Technisch resulteert dit in een test - hertest betrouwbaarheidsindex met een waarde tussen .00 en 1.00. In de praktijk is deze controle echter moeilijk uit te voeren doordat er een effectverandering kan optreden tussen twee tijdstippen die niet dicht op elkaar volgen. Een survey die met een interval van een maand

tweemaal wordt afgenomen bij dezelfde ontvangers van een dienstverlening kan verschillende resultaten opleveren. De reden daarvoor ligt niet noodzakelijk bij de onbetrouwbaarheid van het meetinstrument, i.c. de vragenlijst, maar kan ook te wijten zijn aan een veranderde mening bij de respondenten door toedoen van externe factoren. Wanneer de test - hertest controle niet mogelijk is, kan de betrouwbaarheid nagegaan worden door de antwoorden op gelijkaardige items na te gaan op interne consistentie. Men bekijkt dan het antwoordpatroon van de respondenten.

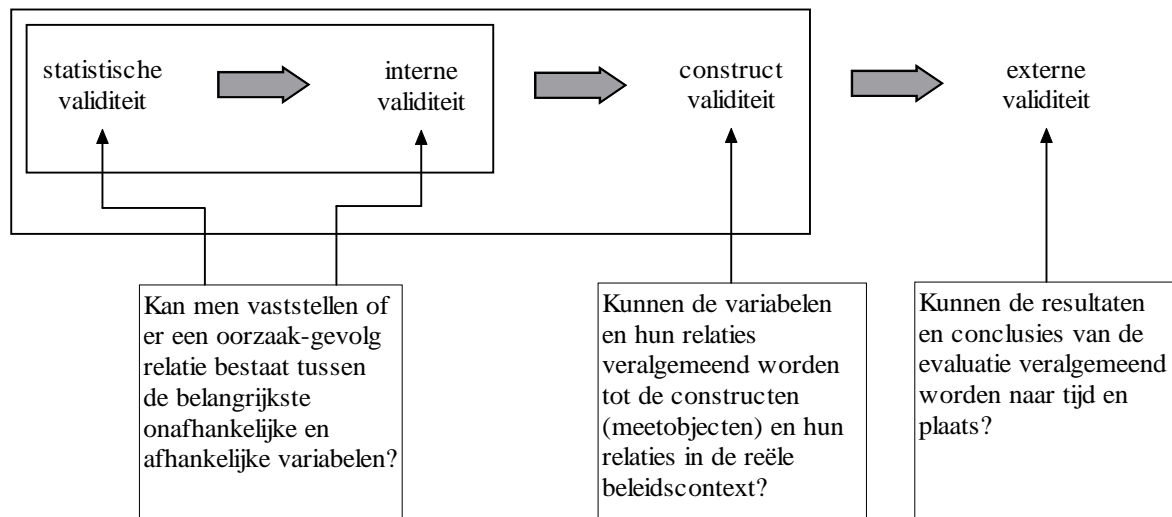
Voor een heel aantal instrumenten dat in evaluaties gebruikt wordt, zal er informatie over de betrouwbaarheid beschikbaar zijn vanuit andere toepassingen of rapporten in het kader van de initiële ontwikkeling. Anderzijds is wel voorzichtigheid geboden wanneer men aanneemt dat wanneer de betrouwbaarheid van een meetinstrument of -procedure werd aangetoond bij andere toepassingen dat dit dan ook zonder meer geldt in het kader van evaluatie.

Er bestaan geen harde regels op het vlak van accepteerbare niveaus van betrouwbaarheid. De mate waarin meetfouten de resultaten voor een effect vertekenen, hangt in grote mate af van de omvang van dat effect (bijvoorbeeld uitgedrukt in procentuele verandering). Als stelregel hanteren onderzoekers algemeen een betrouwbaarheidscoëfficiënt van .90 of meer; een marge die meetfouten beperkt houdt in relatie tot alle effecten, behalve deze met allerkleinste omvang. Voor veel meetinstrumenten die toegepast worden in een context die typerend is voor beleidsevaluatie is dit echter een relatief hoge norm.

3.5.2 Types van validiteit

Een tweede maatstaf of meetkenmerk is validiteit of geldigheid.

Er bestaan echter verschillende soorten validiteit in het kader van gehele proces van meten en evalueren. Cook en Campbell (1979) brachten vier types van validiteit in kaart en suggereerden hoe ze met elkaar in verband staan. Onderstaande figuur toont het verband en de kernvraag die bij elk type van validiteit behoort.



Figuur 14: Types van validiteit doorheen het evaluatieproces

De eerste twee types hebben betrekking op de (oorzaak - gevolg) relaties tussen de variabelen (meetbare elementen) die gemeten worden via indicatoren. In deel 1 werd gesteld dat causaliteit 3 voorwaarden omvat: een tijdsverloop waarbij de oorzaak vooraf gaat aan het gevolg, co-variatie tussen de onafhankelijke variabele (oorzaak) en afhankelijke variabele (gevolg), en het uitsluiten van alternatieve verklaringsgronden voor het vastgestelde gevolg. Statistische validiteit heeft betrekking op de tweede voorwaarde en interne validiteit op de laatste voorwaarde van causaliteit.

Construct validiteit heeft rechtstreeks betrekking op de relatie tussen een indicator en het meetobject. Dit type stelt de vraag naar: "Meten we wat we willen meten?". Met andere woorden: vormt de indicator of indicatorenset een goede vertaling of operationalisering van het meetobject? Deze problematiek kwam reeds aan bod bij de indicatorenontwikkeling (cf. supra: paragraaf 3.4).

Externe validiteit gaat over het veralgemenen van de evaluatieresultaten en –conclusies naar andere situaties. In wat volgt gaan we dieper in op elk type validiteit en de mogelijke bedreigingen.

3.5.2.1 Validiteit van statistische conclusies

Dit type van validiteit betreft het vaststellen van het bestaan en de sterkte van co-variatie tussen de oorzaak- en gevolgvariabelen. Co-variatie bestaat wanneer een variabele varieert (in waarde) als een andere variabele varieert. Co-variatie kan positief of negatief zijn. Het is één van de drie noodzakelijke maar elk apart onvoldoende voorwaarden om te kunnen spreken van een causale relatie tussen twee variabelen. De andere voorwaarden zijn enerzijds tijdsasymmetrie, d.w.z. de oorzaakvariabele gaat de gevolgvariabele vooraf, en anderzijds de afwezigheid van plausibele alternatieve hypothesen, d.w.z. andere factoren die mogelijk de co-variatie tussen de twee variabelen kunnen verklaren.

Validiteit van statistische conclusies valt binnen het kader van statistische analyse. In het bestek van dit deel van de handleiding gaan we niet in op de details van statistische methoden. We verwijzen naar deel 1 en deel 3 voor meer informatie over de aanwending van statistische technieken in het kader van evaluatie. Het is evenwel van belang te weten dat statistische validiteit het volgende behelst:

- Verzekeren dat steekproefprocedures adequaat zijn, wanneer ze gebruikt worden voor dataverzameling.
- Verzekeren dat gepaste statistische testen worden toegepast en dat de resultaten correct geïnterpreteerd worden.
- Verzekeren dat de meetinstrumenten die worden gebruikt voldoende betrouwbaar zijn.

Het laatste element beklemtoont een algemene veronderstelling waarop de theoretische modellen achter statistische testen steunen, namelijk dat de gevolgde meetprocedures betrouwbaar zijn.

3.5.2.2 Interne validiteit

Interne validiteit betreft een andere voorwaarde van causaliteit: het uitsluiten van alternatieve hypothesen. In de literatuur werden bedreigingen voor interne validiteit opgelijst. Hieronder geven de we belangrijkste soorten van bedreigingen aan. Elke soort omvat een groep van mogelijke problemen. In een concrete evaluatie- en meetcontext kunnen geen, één of meerdere factoren een probleem vormen voor een soort bedreiging.

1. Historisch pad

Externe gebeurtenissen kunnen zich voordoen tijdens de implementatie van het beleid. Stel bijvoorbeeld dat een beleidsinitiatief wordt genomen om de ongevallen met gewonden op autowegen te wijten aan alcoholmisbruik te reduceren. Later in datzelfde jaar wordt een wet van kracht die het dragen van de veiligheidsgordel verplicht maakt. Vermits die wet erop gericht is om het aantal letselongevallen te doen dalen, is het quasi onmogelijk om de effecten van de twee maatregelen te ontrafelen.

Een natuurramp kan de relaties tussen verschillende bevolkingsgroepen in een stad of regio nauwer aanhalen. Wanneer dit gebeurt tijdens een evaluatie van het integratiebeleid, kan dit leiden tot een overschatting van de effectiviteit van maatregelen die werden genomen.

2. Lange termijntrends

In een tijdsperiode waarin de verkeersdrukke stijgt en de snelheid daalt, kan een maatregel om het aantal verkeersslachtoffers te verminderen effectief lijken door een vertekening vanuit een bestaande dalende trend van ongevallen en slachtoffers. Maatregelen in de landbouwsector om de teelt van bepaalde gewassen te verhogen, kunnen ineffectief lijken wanneer de schatting van het beleidseffect

gebaseerd is op metingen die werden beïnvloed door extreem ongunstige weersomstandigheden gedurende de uitvoering van het beleid. De effecten van een programma gericht op de wedertewerkstelling van ex-gedetineerden kunnen vertroebeld worden door een algemene neerwaartse spiraal op de arbeidsmarkt door een economische terugval.

3. Rijping

Ingeval de analyse-eenheden personen zijn, kunnen ook factoren spelen die verbonden zijn met het algemene proces van maturiteit en leeftijd. Wanneer personen die deel uitmaken van de beleidsdoelgroep ouder worden, kan hun gedrag veranderen op een manier dat op een beleidseffect lijkt. Het reële effect van een campagne om jongvolwassenen aan te zetten tot sportbeoefening kan gemaskeerd worden door het gegeven dat de interesse voor sport in het algemeen daalt wanneer deze personen de arbeidsmarkt betreden. Of stel dat er een preventieprogramma wordt opgezet in een gemeente ter voorkoming van jeugdcriminaliteit. De implementatie gebeurt binnen een langere periode waarin het lokale bevolkingaantal snel stijgt. De bevolking wordt rijper als een cohorte. Kinderen die worden geboren in de gemeente doorlopen het rijpingsproces ook in een cohorte. Wanneer het preventieprogramma wordt ontwikkeld wanneer de gemiddelde leeftijd van de jeugd 12 jaar is, is het mogelijk dat tegen het tijdstip wanneer de gemiddelde leeftijd 16 jaar bereikt, het probleem verwaterd is zelfs zonder het programma.

4. Testprocedure

Wanneer bij een pretest - posttest formule telkens dezelfde test wordt gehanteerd, kan de posttest hogere scores opleveren alleen al door leereffecten met betrekking tot de testprocedure. Deelnemers aan een opleiding bijvoorbeeld die een kennistest ondergaan kunnen op de posttest beter scoren, niet door doordat hun kennis toenam door de opleiding, maar simpelweg omdat ze vertrouwd zijn geraakt met de test toen deze werd afgenomen voor de start van de vorming.

5. Instrument

Op het moment dat het beleid wordt ingevoerd, verandert ook de wijze waarop kernvariabelen worden gemeten. Stel een beleidsmaatregel wordt gelanceerd om inbraken te doen dalen, terwijl hetzelfde moment het registratiesysteem van de politie wordt geautomatiseerd, afwijkende definities van criminaliteit worden uitgeklaard, en er een grotere inspanning wordt geleverd om alle gerapporteerde feiten in de databank op te nemen. Het netto-effect is dat het aantal gerapporteerde criminele feiten stijgt.

6. Statistische regressie

Extreme scores op een eerste meting tenderen om een grotere meetfout te bevatten. Dus wanneer personen worden geselecteerd als doelgroep van een beleidsmaatregel omdat zij hoog of laag

scoorden op de test, zullen hun scores terugvallen naar de gemiddelde score van alle potentiële doelgroep personen op de posttest. Personen die men selecteerde voor het volgen van een vormingscursus in arbeidsvaardigheden op basis van hun lage zelfbeeldscore, zullen op de posttest hoger scoren op het vlak van zelfbeeld. In deel 1 wordt ook ingegaan op het fenomeen van statistische regressie in het kader van quasi-experimentele evaluatiedesigns.

7. Selectie

Personen of analyse-eenheden die deel uitmaken van de doelgroep van het beleid, verschillen van deze toegewezen aan een controlegroep op relevante kenmerken. Voor een uitgebreide bespreking van de zogenaamde selectiebias, verwijzen we eveneens naar deel 1.

8. Drop out

Tijdens de implementatie van het beleid vallen analyse-eenheden (personen, locaties, organisaties) weg uit de doelgroep. Personen die voor het einde van een psychologische behandeling afhaken, hadden mogelijk bij voorbaat de minste slaagkansen. Een vergelijking van de resultaten uit een voor- en nameting kan dan ook een vertekend beeld opleveren waardoor men verkeerde conclusies trekt over de doeltreffendheid van de behandeling.

Zo kan een kruispunt dat in een controlegroep zit voor een evaluatie van het effect van roodlichtcamera's tijdens de evaluatie afgesloten worden wegens werken. Men zal er dan geen nameting van het aantal ongevallen kunnen doen. Organisaties die aanvankelijk meewerken kunnen op een later moment weigeren om nog informatie te verschaffen.

Zolang men effectgegevens van een nameting kan recupereren voor analyse-eenheden, blijft de validiteit van het evaluatiedesign intact. Problematisch wordt het natuurlijk wanneer bij de beleidsgroep het beleid maar gedeeltelijk wordt uitgevoerd. De beoordeling van het beleidseffect zal dan minder scherp zijn; bij uitvoering van het volledige beleid was het beleidseffect mogelijk sterker geweest dan wat men nu vaststelt.

9. Ambigue oorzaak en gevolg relatie

In bepaalde situaties kan het onduidelijk zijn of een kernvariabele het effect veroorzaakt, of omgekeerd. Een voorbeeld uit een personeelsbeleid: men kan veronderstellen dat een hogere arbeidsproductiviteit wordt bereikt door het verbeteren van het moreel. Stel dat uit de monitoringgegevens blijkt dat beide aspecten verbeteren. Echter, een verbeterd moreel is misschien het gevolg van de hogere productiviteit. Of er bestaat de facto een relatie van wederzijdse beïnvloeding waarbij een betere mentale veerkracht leidt tot meer productiviteit en deze laatste versterkt op zijn beurt het moreel, wat weer een gunstig effect heeft op de productiviteit, enzovoort.

10. Selectiegebaseerde interacties

De selectie van de doelgroep van het beleid interfereert met andere soorten bedreigingen van validiteit, zodat er een gecombineerd effect ontstaat op relevante variabelen.

Een programma gericht op het verbeteren van leesvaardigheden binnen een schooldistrict wordt zo georganiseerd dat de klassen met het programma gesitueerd zijn in wijken met hogere inkomens, terwijl de controlegroepen gesitueerd zijn in wijken met lagere inkomens. De testresultaten van de voor- en nameting in de twee groepen zullen vertekend zijn, niet enkel door een bias in de selectie (verschillend gezinsinkomen), maar ook door het sociologisch feit dat kinderen uit gezinnen met een hoger inkomen gemiddeld vroeger schoolrijp zijn.

Elk van de bovenstaande tien soorten gevaren voor interne validiteit vragen om mogelijke oplossingen voor de specifieke problemen, hoewel de uitklaring van ambigue causale relaties moeilijk zal blijven. Met betrekking tot het historisch pad, bijvoorbeeld, kan men op bepaalde contextuele gebeurtenissen die kunnen samenvallen met de invoering van het beleid anticiperen. Ook door het voorzien van een controlegroep (cf. deel 1) kunnen historische factoren beide groepen beïnvloeden, waardoor incrementele of netto-effecten van het beleid afgezonderd en gemeten kunnen worden. Een terugkerend knelpunt bij het toepassen van dergelijke oplossingen is de mate van middelen en controle over het ontwerp en de uitvoering van het beleid die men heeft, om de situatie dermate te organiseren dat men een probleemvrije meet- en evaluatiecontext kan creëren.

3.5.2.3 Constructvaliditeit

Constructvaliditeit slaat op het verband tussen enerzijds de variabelen en hun onderlinge relaties die worden gemeten in het kader van monitoring en beleidsevaluatie, en anderzijds, de meetobjecten of constructen en hun relaties in de beleidsrealiteit.

Metten komt neer op het vertalen van het meetobject naar waarneembare elementen. Op dat 'niveau' van waarneembaarheid werkt men met variabelen en indicatoren; niet met de constructen zelf. Problemen met constructvaliditeit kunnen voortkomen uit de wijze waarop het beleid werd geoperationaliseerd voor monitoring en evaluatie; de manieren waarop output, determinanten, of effecten uit het veranderingsmodel worden vertaald naar indicatoren.

Shadish, Cook en Campbell (2002) wijzen erop dat problemen met constructvaliditeit alsook problemen van interne validiteit betrekking hebben op factoren die onze interpretatie van resultaten van een evaluatie in de war sturen.

Een viertal bedreigingen voor interne validiteit heeft te maken met de context voor beleidsimplementatie. Deze problemen stellen zich niet bij afwezigheid van het beleid. Ze tasten de mogelijkheid aan om te veralgemenen van de geobserveerde resultaten naar de concepten in de beleidstheorie. Daarom halen we ze hier aan als gevaren voor constructvaliditeit. Deze gevaren zijn vooral van tel in een beleidscontext waarbij een maatregel of programma wordt uitgevoerd gericht op een welbepaalde doelgroep. Bijgevolg zijn er andere personen buiten de doelgroep, die in het kader van een beleidsevaluatie een controlegroep kunnen vormen.

1. Diffusie van beleid

Personen uit de doelgroep kunnen soms communiceren over hun ervaringen met het beleid met personen uit de controlegroep. Bij trainingsprogramma's bijvoorbeeld kunnen op deze wijze vaardigheden of kennis informeel doorgegeven worden naar personen die niet tot de doelgroep behoren.

2. Compenserende gelijkstelling van beleid

Het beleid wordt uitgebreid naar personen die niet deel uitmaken van de initiële doelgroep, met het oog op het vermijden van kritiek, of om hen niet te onthouden van bepaalde voordelen, bijvoorbeeld in het kader van een fiscale maatregel.

3. Compenserende rivaliteit

Personen uit de controlegroep zijn het subject van een "betere" versie van het bestaande beleid, wat ervoor zorgt dat de verschillen tussen het nieuwe en voorgaande beleid praktisch worden opgeheven. Bijvoorbeeld wordt een elektronische interactieve afhandeling van dossiers aangeboden, terwijl andere dossiers nog via het loket worden afgewerkt. Loketbedienden zullen dubbel hun best doen om zich niet te laten verdringen door de nieuwe technologische methode.

4. Ontstemde ontmoediging

De controlegroep heeft de perceptie unfair behandeld te worden en reageert negatief. Personen die bijvoorbeeld niet deel uit maken van een leerexperiment in kleinere groepen, klagen bij de opleidingsorganisator en vragen om een gelijke behandeling. Wanneer de organisator weigert, dreigen de deelnemers aan de opleiding ermee om geen testen meer af te leggen.

Shadish, Cook en Campbell reiken ook volgende hints aan om problemen met constructvaliditeit te minimaliseren:

- Zorg voor duidelijke definities van de meetobjecten, constructen of concepten zodat ze adequaat gemeten kunnen worden;

- Differentieer constructen zodat ze niet overlappen wanneer meetinstrumenten en indicatoren worden ontwikkeld;
- Ontwerp goede maatstaven, d.w.z. maatstaven die geen vertekende informatie opleveren;
- Hanteer meerdere indicatoren voor de belangrijkste constructen uit de beleidstheorie.

Gevaren voor constructvaliditeit kunnen ook ontstaan door het feit dat personen weten dat ze betrokken zijn in een evaluatieproces. Verwachtingen van deelnemers kunnen hun gedrag beïnvloeden, waardoor pogingen om de vastgestelde resultaten te veralgemenen naar de concepten in de beleidstheorie vertekend worden. Een voorbeeld hoe gedrag van deelnemers kan interfereren met de meetresultaten, is het Hawthorne effect. Tijdens een experiment in een gelijknamige fabriek in de VS in de jaren '30 bleek dat de productiviteit van werknemers werd beïnvloed door het feit dat ze deel uitmaakten van een experiment. Dit ondanks de manipulatie van andere factoren waarvan men de invloed op het werkritme wilde nagaan, zoals de belichting, de snelheid van de assemblagelijijn, taakdifferentiatie, enz. Paradoxaal bleek dat de productiviteit steeg ongeacht welke ingreep er werd gedaan op andere factoren. De meetresultaten vertekenden de conclusies over de verbanden tussen de verschillende factoren die in de werkelijkheid, d.w.z. zonder experimenteersituatie - bestaan.

3.5.2.4 Externe validiteit

Uit de bovenstaande figuur blijkt dat externe validiteit verder bouwt op de andere soorten validiteit die reeds werden besproken: validiteit van statistische conclusies, interne validiteit en constructvaliditeit. Bij externe validiteit gaat het om het veralgemenen van de causale resultaten van een beleid naar andere settings, personen, beleidsvarianten, en/of tijdsaders.

Fundamenteel is de vraag of het beleidsvoorbeeld en de concrete context representatief zijn voor andere omstandigheden waarin men wenst het beleid of de evaluatiebevindingen toe te passen. Hieronder worden een vijftal gevaren voor externe validiteit aangehaald. Ze maken dat de resultaten van een evaluatie uniek zijn in een bepaald opzicht. Vier gevaren zijn interactie-effecten die problemen geven voor het veralgemenen naar andere analyse-eenheden (meestal personen), naar andere beleidsvarianten, naar andere effectvariëaties en naar andere settings.

1. Interactie tussen de causale resultaten van een beleid en de personen/deelnemers

Stel dat een try-out met een nieuwe onderwijsaanpak in de lagere school betere resultaten met meisjes oplevert dan met jongens. Betekent dit dat inspanningen om deze aanpak elders in te voeren geconfronteerd zullen worden met dezelfde beperking?

2. Interactie tussen de causale resultaten van een beleid en de variaties in uitvoering

Een beleidsinitiatief waarvan de effectiviteit is aangetoond werd, kan door anderen overgenomen worden. Het kan echter zijn dat men het initiatief in de uitvoeringswijze licht wijzigt, bijvoorbeeld om kosten te besparen. Ondermijnt dit de beleidseffectiviteit?

3. Interactie tussen de causale resultaten van een beleid en variaties in effecten

Stel dat een Vlaams tewerkstellingsprogramma erin slaagt om 55% van de deelnemers aan een baan te helpen van minstens 6 maanden. Een vergelijking met personen die niet deelnamen aan het beleid omwille van plaatsgebrek, suggereert dat de tewerkstelling steeg met 25%. Wallonië is geïnteresseerd om het beleid over te nemen, maar wil de klemtoon leggen op lange termijn tewerkstelling (minimum 2 jaar). Zou het beleid dezelfde resultaten opleveren als de definitie van het beoogde effect veranderd werd?

4. Interactie tussen de causale resultaten van een beleid en de setting

Een buurtinformatienetwerk (BIN) in een kleine gemeente kan succesvoller zijn dan wanneer deze formule in een grootstad wordt toegepast. In een stad is de anonimiteit in veel gevallen groter, en de contacten tussen buurtbewoners minder frequent. De drempel om met elkaar in te komen is groter, waardoor het moeilijker kan zijn om engagement te verkrijgen van de buurtbewoners, om bijvoorbeeld verdachte zaken te melden.

5. Contextafhankelijke tussenkomst

Een beleid gericht om criminaliteitspreventie kan sneller geïmplementeerd worden wanneer er al buurtinformatienetwerken bestaan. Een BIN fungeert dan als hefboom om het beleid succesvol in te voeren. De afwezigheid ervan in andere gemeenten kan daar ervoor zorgen dat er minder burgers bereid zijn om mee te werken, waardoor de algemene effectiviteit van het beleid lager zal liggen.

3.5.3 Sensitiviteit van indicatoren

De functie van effectindicatoren is om veranderingen of verschillen te detecteren in factoren die men met het beleid tracht te beïnvloeden. Om deze functie naar behoren te vervullen, dienen deze indicatoren gevoelig zijn voor dergelijke effecten. De sensitiviteit van een maatstaf is de mate waarin de waarden op de maatstaf veranderen wanneer er een verandering of verschil is bij datgene wat gemeten wordt. Zo is er bijvoorbeeld een verschil in sensitiviteit tussen een personenweegschaal en een weegbrug voor vrachtwagens. Deze laatste kan ook een betrouwbaar en valide meetinstrument zijn, maar is niet in staat om erg kleine verschillen in gewicht van enkele honderden grammen te detecteren en aan te geven. In het kader van een persoonlijk dieetprogramma is het echter noodzakelijk om een voldoende gevoelige weegschaal te gebruiken.

Er zijn twee wijzen waarop de soorten van effectindicatoren die frequent in evaluatie worden gebruikt, ongevoelig kunnen zijn voor veranderingen of verschillen in het belang dat een beleidsinitiatief kan betekenen. Ten eerste kan de maatstaf elementen omvatten die aan iets anders gerelateerd zijn dan aan datgene wat het beleid volgens de verwachtingen kan veranderen. Dergelijke elementen verzwakken de concentratie aan elementen die wel 'responsief' zijn en dempen de globale respons van de indicator. Een score op een algemene wiskundetest kan bijvoorbeeld gebruikt worden om de vooruitgang van de leerlingen te meten op het vlak van staartdelingen. Anderzijds kunnen de reële vorderingen in deze soort oefeningen overschaduw worden door scores op andere delen van de test. Een meer sensitieve maatstaf in dit geval is de score op een test die specifiek gaat over staartdelingen.

Ten tweede kunnen effectindicatoren ongevoelig zijn voor beleidseffecten wanneer ze werden ontwikkeld voornamelijk in het kader van diagnose en dus om individuele verschillen te detecteren. Het doel van dergelijke maatstaven is om scores zo te spreiden dat men een opdeling krijgt van analyse-eenheden die meer of minder het kenmerk hebben dat gemeten wordt. Psychologische testen, kennistesten of attitudeschalen zijn voorbeelden hiervan, met personen als analyse-eenheden. Dergelijke instrumenten geven een algemene indicatie en kunnen bijvoorbeeld de ernst van een probleem aangeven. Wanneer echter dergelijke meetinstrumenten in een voor- en nameting worden toegepast bij individuen die voor de beleidsuitvoering al zeer sterk verschilden, kunnen de grote verschillen in de nameting kleine verbeteringen bij de afzonderlijke individuen verbergen. Vanuit een meetperspectief vormen de individuele verschillen waarvoor deze maatstaven wel gevoelig zijn een verwaarloosbare 'ruis' met het oog op het vaststellen van veranderingen op groepsniveau. Effecten op individueel niveau dreigen daardoor niet naar boven te komen in de metingen.

Hoe kan men nagaan of een effectindicator voldoende gevoelig is in meettermen? Een eerste werkwijze is te zoeken naar bestaand (evaluatie)onderzoek waarin deze indicator erin slaagde om veranderingen te detecteren die van dezelfde grootteorde zijn als die de evaluator verwacht van het te evalueren beleid. De beste bewijsvoering zal uiteraard komen van evaluaties van gelijkaardige beleidsinitiatieven. Daarbij dient men ook aandacht te hebben voor de grootte van de beleids- en controle of vergelijkingsgroepen, vermits de omvang van een steekproef invloed heeft op de detecteerbaarheid van effecten.

Een andere werkwijze bestaat erin om de indicator toe te passen op groepen waarvan men de verschillen kent, of in situaties van gekende verandering, om na te gaan hoe responsief de indicator is. Zo kan een evaluator de vraag stellen of een standaard wiskundetest een goede effectindicator is in termen van sensitiviteit. Hierover kan twijfel bestaan, gezien de lessen focussen op bepaalde deelaspecten van wiskunde, terwijl de test een brede waaier van topics bestrijkt. Voordat de test wordt gebruikt in een evaluatie van het onderwijscurriculum, kan men de test afnemen bij kinderen voor- en nadat ze breuken en staartdelingen leren. Wanneer de test voldoende gevoelig blijkt om veranderingen tijdens de

leerperiode te detecteren, geeft dit enige zekerheid dat de test gevoelig is voor effecten bij het evalueren van het ganse onderwijspakket in wiskunde.

3.5.4 Meetproblemen

In tegenstelling tot wat sommige meettechnici zouden verwachten, is het introduceren van een monitoringsysteem niet per definitie neutraal; betrokken actoren zullen interpretaties maken van het nut en het doel van de metingen. Zeker wanneer metingen dienen als input in het kader van beleidsevaluaties, kan het meten zelf leiden tot een aantal problemen. Bepaalde disfuncties zullen zich sneller of sterker stellen naargelang het motief voor het meten en evalueren. Zo kan het verantwoordingsmotief de meeste meetproblemen met zich meebrengen.

Meetproblemen kunnen betrekking hebben op het meetsysteem als geheel of zich manifesteren bij afzonderlijke indicatoren, voornamelijk output- en effectindicatoren. Bouckaert (1999) heeft een reeks van meetproblemen of pathologieën opgesteld. Het gaat ten eerste om assumpties die organisaties hebben over het meten van hun performance. Ten tweede houden meetproblemen verband met de perceptie van meetwaarden met betrekking tot volume en aantal. Ten derde zijn er meetproblemen verbonden aan de inhoud, status of het aantal indicatoren.

3.5.4.1 “Meten is niet nodig” – Pangloss stelling

Een eerste probleem voor het introduceren van meetsystemen is de stelling dat er geen nood is aan meten. Men veronderstelt dat er geen probleem is, derhalve kan het niet in kaart worden gebracht. Net zoals dr. Pangloss uit de komedie ‘Candide’ van Voltaire stelt dat de huidige wereld de beste is van alle mogelijke, kunnen sommige stakeholders van het beleid zeggen men het optimum heeft bereikt. Dergelijke uitspraak kan slaan op de beheersing of wegwerking van wat vroeger een probleem was, de efficiëntie van de beleidsuitvoering, de bereiking van doelstellingen, de effectiviteit van het lopende beleid, enz. Metingen, zo gaat de redenering verder, zijn dan ook niet nodig.

3.5.4.2 “Meten is onmogelijk”

Ook wanneer men het eens is over het nut en de noodzaak van het meten, kan men stoten op de overtuiging dat er geen adequate indicatoren gevonden kunnen worden om relevante output of effecten te meten. De beleidsproblematiek, beleidseffecten of de beleidscontext is dermate complex of specifiek, dat men dit niet kan meten. Vaak zal men daarbij steunen op het argument dat de publieke sector geconfronteerd wordt met een veel complexere omgeving dan bijvoorbeeld de private sector, zodat men geen betrouwbare en geldige metingen kan uitvoeren.

3.5.4.3 "Performance en publieke sector gaan niet samen"

Bepaalde stakeholders kunnen een algemeen negatief gevoel hebben op de werking en het potentieel van de overheid. Deze teneur kan overslaan naar een negatieve opinie over de kwaliteit en bijgevolg de bruikbaarheid van meetsystemen in een overheidscontext. Ambtenaren kunnen bijvoorbeeld van mening zijn dat het zinloos is te spreken over efficiëntie en productiviteit in de publieke sector. Hetzij omdat dit niet de opdracht is van de overheid, hetzij omdat men meent niet te kunnen wedijveren met de private sector. Meetsystemen om dit op te volgen zijn zelf kwalitatief ook ondermaats, zo gaan sommigen verder.

De vier volgende problemen verwijzen naar foutieve percepties van meetwaarden omtrent aantal en volume.

3.5.4.4 Convex / concaaf

Naar analogie met optische vervormingen door een bolle of holle spiegel, kunnen ook meetinstrumenten hogere of lagere waarden van output of effect meten, dan er in werkelijkheid bestaat. Convexiteit wijst op onterecht hogere waarden; de grafiek toont dan een klokvorm. Een concaaf meetinstrument resulteert in een holle distributie van waarden in een grafiek. De waarden zijn lager dan ze in werkelijkheid zijn.

Citatie-indexen in de academische wereld zijn een voorbeeld van convexe meetinstrumenten: doordat een groep van onderling met elkaar vertrouwde onderzoekers elkaar citeren in hun wetenschappelijke publicaties wordt zo voor elke individuele onderzoeker een hoge (maar niet echt terechte) citatiewaarde gecreëerd. Naast de relevantie van het werk speelt immers ook de bekendheid met de auteur mee.

Manuele registratiesystemen zijn vaak concaaf. Zo kloppen bijvoorbeeld de aantallen in aanwezigheidsregisters bij het onthaal van een organisatie vaak niet met het reële aantal in het gebouw aanwezige personen, omdat binnenkomende gasten al eens vergeten in te schrijven.

3.5.4.5 Output of effect stijgen abnormaal door het meten

Wanneer een meetinstrument wordt geïntroduceerd, dan kan hetgeen gemeten wordt in volume of intensiteit toenemen tengevolge van de activiteit van het meten. In tegenstelling tot het voorgaande probleem, verandert nu door de meting de reële waarde. De meting verandert de werkelijkheid. Er ontstaan ongewenste neveneffecten of zelfs perverse effecten van datgene dat men wil bereiken. Een voorbeeld komt uit de gezondheidszorg in het Verenigd Koninkrijk. Een van de criteria waaraan Britse dokters van een radiologieafdeling door de *National Health Service* werden beoordeeld was de kost per uitgevoerde radiografie. Door het meetsysteem trachtte men de radiologen een kostenbewustzijn bij te brengen. Dit leek ook te lukken. Het perverse effect van deze meting was dat de budgetten van ziekenhuizen met radiologieafdelingen exponentieel stegen. Bedrijfseconomisch kan de eenheidskost immers dalen door het aantal eenheden te doen toenemen. Het aantal radiografieën steeg dramatisch

maar daarmee ook de totale kost. Netto betekende het gebruik van de indicator *kost per radiografie* dus een meeruitgave in plaats van de gewilde besparing.

3.5.4.6 Output of effect dalen abnormaal door het meten

Ook het tegenovergestelde effect kan als (ongewild) gevolg optreden bij de invoering van bepaalde indicatoren: een daling of zelfs het verdwijnen van het volume of de kwaliteit van output of van effecten. Het opvolgen van volumes alleen leidt bijvoorbeeld al makkelijk tot een lagere kwaliteit van de afgeleverde producten en diensten. Als men de doorrijtijden langs haltes van een openbaar vervoersnetwerk wil meten om een uurrooster aan te passen aan de gewijzigde verkeersomstandigheden, bestaat de kans dat sommige chauffeurs tijdens metingen bewust trager gaan rijden. Daardoor krijgen ze in de nieuwe uurregeling meer tijd om van halte tot halte te rijden en bijgevolg minder stress bij piekmomenten in de verkeersdrukte.

Een eenzijdige focus op het volume kan ook de aandacht voor andere criteria zoals kwaliteit of klantvriendelijkheid doen dalen. In het kader van een werklasmeting kan een administratie er alles aan doen om zoveel mogelijk dossiers binnen een gegeven termijn af te handelen. De kwaliteit van de dienstverlening kan echter hieronder lijden, waardoor ook de tevredenheid van de klanten of burgers zal dalen ondanks het feit dat ze sneller 'geholpen' zijn.

3.5.4.7 Meer registraties door het meten

Ook een herkenbaar probleem is dat door het meten van een bepaald fenomeen de frequentie ervan ogenschijnlijk toeneemt. Bijvoorbeeld, meer politie op straat resulteert in een hoger aantal vastgestelde gevallen van straatcriminaliteit. *Amnesty International* rapporteert meer gevallen van schending van mensenrechten, maar meldt tevens in hetzelfde rapport dat er meer *human right centers* (observatieposten) bestaan. Dit betekent niet noodzakelijk dat de straatcriminaliteit of het aantal schendingen in werkelijkheid stijgt. Door meer metingen te organiseren of verfijndere meetprocedures tot te passen, kunnen meer gevallen worden gedetecteerd. In dergelijk geval is het verkeerd te besluiten er een stijging is van de output of het effect; door de verscherpte aandacht wordt de bestaande toestand beter in kaart gebracht.

De resterende meetproblemen verwijzen naar meer algemene gebreken die bij meetsystemen kunnen ontstaan.

3.5.4.8 Geen duidelijk onderscheid tussen output en effect

Wanneer output en effecten niet duidelijk worden onderscheiden, maar voor een deel gemeten, ontstaat er een probleem van transparantie. Dit kan problematisch zijn voor het interpreteren van gegevens in het kader van een evaluatie. Bovendien geeft dit ruimte aan het arbitrair toekennen van gewichten aan

bepaalde indicatoren. Daardoor kan de focus van het beleidsmanagement verkeerd gelegd worden, wat op zijn beurt de beleidseffectiviteit hypothekeert.

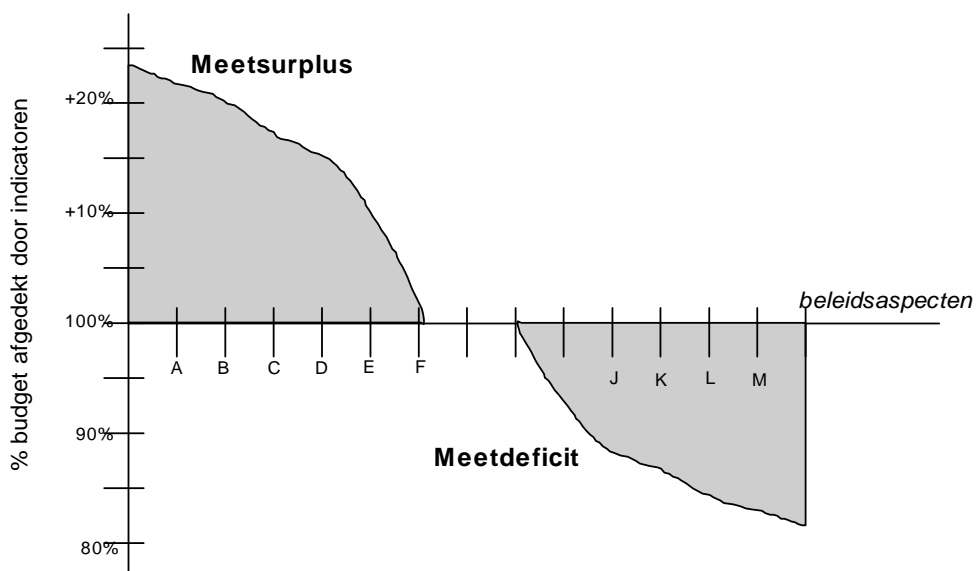
Veel organisaties opereren in een bestuurlijke keten: de output van de ene organisatie dient als input voor een andere, binnen een ruimere opdracht waarbij de taken en verantwoordelijkheden verdeeld zijn over meerdere actoren en instanties.

3.5.4.9 Inflatie van indicatoren

Alhoewel complexe maatschappelijke fenomenen waarschijnlijk een veelvoud aan indicatoren vereisen om ze accuraat te kunnen opvolgen, brengt een overvloed aan indicatoren meer na- dan voordelen op. Een meetsysteem dat alsmaar uitbreidt, wordt onbeheersbaar; de opslag en verwerking vereisen meer tijd en middelen, terwijl men het overzicht aan beschikbare informatie verliest.

Vaak is de betrachting om met het monitoringsysteem een zo groot mogelijke dekkingsgraad te verkrijgen van het op te volgen beleidsterrein. Eerder wezen we al op de Pareto optimaliteit. Men verlangt een 100% dekkingsgraad terwijl waarschijnlijk met een 80% dekking van het budget toegewezen aan het beleid, de meest belangrijke aspecten van het beleid door het meetsysteem afgedekt zijn. De kost die nodig is om de laatste 20% af te dekken zal in veel gevallen niet opwegen tegenover de verwachte baten.

Aan het inflatieprobleem is nog een ander fenomeen verbonden, dat van het meetsurplus en het meetdeficit.



Figuur 15: Budgettaire afdekking per beleidsaspect: meetsurplus en -deficit

Voor bepaalde beleidsaspecten (A-F) bestaat er een overvloed aan indicatoren (meetsurplus), terwijl andere componenten (J-M) te weinig opgevolgd worden (meetdeficit). Een meetsurplus dreigt vooral te

ontstaan bij aspecten die relatief makkelijk meetbaar zijn. Minder tastbare elementen worden niet gemeten, waardoor ze ook aan de aandacht dreigen te ontsnappen.

3.5.4.10 Korte termijnfocus

Indicatoren kunnen vooral opgesteld worden in het kader van korte termijn doelstellingen. Meestal gaat het dan om operationele doestellingen die verband houden met output en niet met maatschappelijke effecten. Twee mechanismen kunnen dit in de hand werken. Zo beslaat de begrotingscyclus in de meeste overheidsorganisaties één jaar. Dit drijft de metingen makkelijk in de richting van eenjarige cycli van informatievoorziening. Ten tweede zijn de meeste politieke actoren geneigd om op korte termijn te redeneren. Een aaneenschakeling van opeenvolgende kortetermijnvisies resulteert niet vanzelf in een zinvolle langetermijnvisie; met het gevaar op een beleid waarvan de operationele doelstellingen worden bereikt, maar dat niet effectief is doordat de strategische doelstellingen niet worden gehaald.

3.5.4.11 Het waanbeeld

Problemen met validiteit van indicatoren (cf. supra) lijden tot verkeerde percepties van wat men meet. Dit kan belangrijke implicaties hebben wanneer metingen op deze indicatoren deel uitmaken van evaluaties die op hun beurt beleidsbeslissingen beïnvloeden.

3.5.4.12 De focus wordt verlegd door het meten

Indicatoren vormen een benadering van de werkelijkheid, maar tegelijk ook een reductie daarvan. De indicatorkeuze is dan ook belangrijk. Voor de meeste output- en effectindicatoren is de koppeling met beleidsdoelstellingen essentieel. Wanneer een indicator niet in het verlengde ligt van doelstellingen, dreigt de aandacht te verschuiven naar wat de indicator wel opvolgt maar wat minder relevant is vanuit beleidsoogpunt. Voorbeelden van dit fenomeen zijn te vinden in scholen die enkel intelligente kinderen gaan toelaten om zo het aantal mislukkingen te doen dalen of in hospitalen die alleen patiënten met specifieke ziekten of met een bepaalde ziektegraad aanvaarden om zo de mortaliteit onder controle te houden. Met het aantal correct afgewerkte dossiers als indicator voor hun evaluatie, kunnen dossierbehandelaars plots selectiever te werk gaan: meer moeilijke gevallen worden geweerd en makkelijkere dossiers aangetrokken. In extreme gevallen kan men onnodig dossiers openen om ze nadien dan 'correct' te kunnen afwerken.

Een aantal van de zonet geschetste meetproblemen, zal door ervaring en verfijning van het meetsysteem opgevangen kunnen worden. Andere vormen blijvende valkuilen en vereisen dus permanent aandacht wanneer een indicatorenset ontwikkeld, aangepast of vervangen wordt.

3.6 Rapporteren

Een laatste aspect dat de aandacht verdient bij de ontwikkeling van een monitoringsysteem is het rapporteren van de verzamelde gegevens. Alle belangrijke informatie dient samengebracht en verspreid te worden, aangepast aan de specifieke kenmerken en noden van het doelpubliek. Zonder een goed uitgewerkte rapportering verliest een monitoringsysteem een belangrijk deel van zijn meerwaarde. Een duidelijke rapportering maakt het mogelijk de informatie te gebruiken als basis voor evaluaties en andere doeleinden.

In deze paragraaf gaan we aan de hand van voorbeelden dieper in op volgende vragen. Wat dient er gerapporteerd te worden? Aan wie? Hoe kan de rapportering afgestemd worden op de doelgroep? En welke vormen van rapportering zijn beschikbaar en het meest aangewezen?

Beleidsmakers en het management hebben vaak grote verwachtingen naar de resultaten van de monitoring van de verschillende beleidsprocessen en resultaten. Zoals reeds eerder gesteld is de betrokkenheid van het management bij het uitdenken van de rapporteringswijze essentieel opdat de presentatie van beleidsinformatie afgestemd is op de informatienoden van de beleidsverantwoordelijken.

Maar ook andere, externe stakeholders die ook in de fase van het ontwerpen van een evaluatie in beeld komen, kunnen vragende partij zijn voor beleidsinformatie. Voor een afstemming van vraag en aanbod van informatie dient men zicht te hebben op de informatienoden van elk van deze stakeholders, naast een overzicht van de beschikbare informatie. Vaak zullen dezelfde data op een andere manier gepresenteerd worden aan verschillende doelgroepen. Het brede publiek heeft in het algemeen slechts interesse in de algemene, makkelijk te begrijpen en duidelijk weergegeven resultaten. Het interne management echter verwacht een volledige analyse van de behaalde resultaten, tot in de kleinste details. De beleidsverantwoordelijken verkiezen een korte, en duidelijke samenvatting van de data die in een zeer korte tijdspanne kan worden doorgenomen. Op de hoogte zijn van de voorkeur en het begripsniveau van de betrokken doelgroep zal de keuze van de meest aangewezen communicatiestrategie in grote mate vereenvoudigen.

Bij de selectie van de te rapporteren data zijn drie aandachtspunten van belang. Het eerste aspect is het tijdspectief: voor welke periode(s) dienen er gegevens beschikbaar te worden gesteld? Het tweede aspect is of informatie comparatief moet zijn of niet, en in welk opzicht. Is het nuttig om data weer te geven voor verschillende doelgroepen, locaties, of organisaties? Het derde en laatste aspect is de vraag naar additionele informatie. Vaak dienen cijfers nog geïnterpreteerd te worden of vergen gegevens duiding en achtergrond.

Door aandacht te besteden aan deze drie aspecten verbetert de leesbaarheid en bruikbaarheid van gerapporteerde data. Het maakt ook een doordachte keuze van het aangewezen communicatiemedium mogelijk.

Verschillende rapporteringmogelijkheden

Data en resultaten zijn rapporteerbaar op zeer diverse wijzen, via tal van verschillende media. Hieronder presenteren we een aantal belangrijke rapporteringvormen. Tabellen kunnen we onderscheiden op basis van het niveau waarop ze uitgewerkt zijn, eenvoudig of meer geavanceerd. Wat de grafieken betreft bespreken we de meest voorkomende voorstellingen; een cirkeldiagram, staaf – of stapeldiagram, een kartogram en een lijngrafiek. We identificeren telkens de voor – en de nadelen en geven een aantal concrete tips. Hierbij is het nog belangrijk te vermelden dat afhankelijk van de aard van de informatie, tal van combinaties mogelijk zijn.

- Een basis tabel

De eerste manier van rapporteren is vrij eenvoudig. Niet elk monitoringsysteem dient te beschikken over geavanceerde rapportagetechnieken. Een goed georganiseerde en overzichtelijke tabel is vaak voldoende om een duidelijk overzicht te geven van het uitgevoerde beleid. Tabellen kunnen tevens zeer nuttig zijn voor het weergeven van informatie over verschillende tijdsperiodes heen. Hierbij heeft men de keuze uit rapporteringen per week, maand, kwartaal of per jaar.

Rapporten rechtstreeks opgesteld op basis van dit soort tabellen hebben het grote voordeel vrij duidelijk en toegankelijk te zijn. Daarenboven hebben deze rapporten een lage kost en een snelle productietijd. Éénmaal de rekenbladen ontwikkeld en ingevuld zijn met de concrete gegevens, is het rapport vrij eenvoudig op te stellen. Deze werkwijze is zeer nuttig, zeker voor rapporteringen die zeer regelmatig terugkeren. De tabellen dienen maar één maal ontwikkeld te worden, de nieuwe gegevens worden telkens gewoon ingevuld. Het is natuurlijk vanzelfsprekend dat ook het monitoringsysteem op regelmatige basis herbekeken wordt, waarna ook de tabellen aan herziening toe zijn.

Deze manier van werken, is niet in alle situaties en voor elk doelpubliek aangewezen. Eenvoudige tabellen zijn zeer geschikt voor een publiek dat zeer vertrouwd is met de materie en de manier van werken. Deze techniek levert heel wat informatie aan zodat het ook in hoofdzaak geschikt is voor een publiek dat belang hecht aan inzicht in de details. Mensen die niet vertrouwd zijn met de werking kunnen verloren lopen in dergelijke gedetailleerde informatie. Deze personen zullen meer gebaat zijn met grafische voorstellingen dan met deze overzichtstabellen.

VERDELING VAN GEZINNEN VOLGENS HET BEZIT VAN PERSONENWAGENS

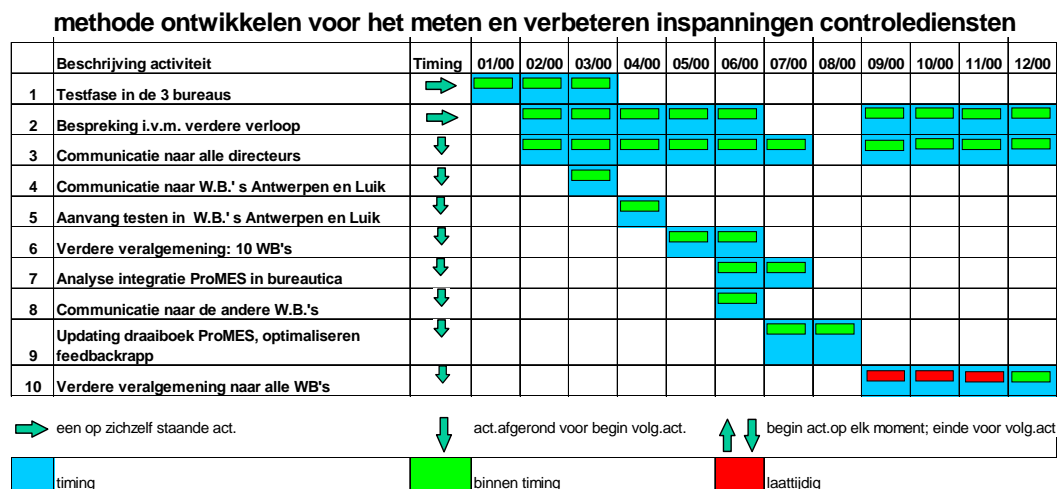
PERSWAGA	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
0	432.0382	14.3	432.0382	14.3
1	1776.143	58.7	2208.181	72.9
2	733.4658	24.2	2941.647	97.2
3	71.10032	2.3	3012.747	99.5
4	8.950949	0.3	3021.698	99.8
5	3.01397	0.1	3024.712	99.9
6	0.70908	0.0	3025.421	99.9
10	1.497048	0.0	3026.918	100.0
23	0.689395	0.0	3027.608	100.0

Figuur 16: Voorbeeld van een basistabel

Bron: Onderzoek verplaatsingsgedrag Vlaanderen – Vlaamse Gemeenschap – Departement LIN - Mobiliteitscel

- Een geavanceerde tabel

Bijkomende informatie toevoegen aan tabellen kan een grote meerwaarde hebben. Voorbeelden van deze bijkomende informatie zijn symbolen en grafieken. De data zijn weergegeven in tabellen met duidelijke en eenduidige cijfers zoals percentages en gemiddelden. Vaak zijn deze rapporten vrij te raadplegen op het internet, wat ook hun toegankelijkheid enorm verhoogt. Het gebruik van deze communicatievorm biedt tal van voordelen. Het eerste positieve element is de vaststelling dat deze data door een breed publiek kunnen geconsulteerd worden. De combinatie van symbolen en cijfers maakt dat lezers kunnen kiezen tussen enerzijds voornamelijk aandacht te besteden aan de cijfers of symbolen en, anderzijds, beiden samen te lezen en actief te gebruiken. De gebruiker kan het rapport snel overlopen, op zoek naar welbepaalde symbolen. Ook het werken met kleuren kan de leesbaarheid en overzichtelijkheid sterk vergroten.



Figuur 17: Voorbeeld van een geavanceerde tabel: PROMES (RVA)

De opmaak van dergelijke uitgebreide tabellen vereist een grotere expertise en tijdsinvestering dan deze van basistabellen. Maar ook hier komt verandering in door de steeds snellere ontwikkeling van gepaste software die deze taken aanzienlijk vereenvoudigen.

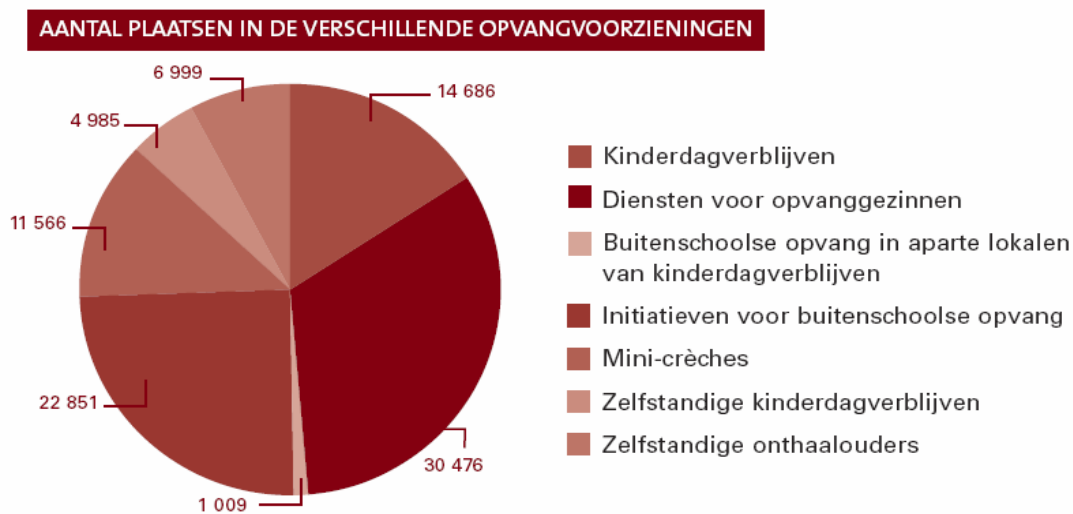
Tips voor het opstellen van tabellen

- Nummer elke tabel op een consequente wijze.
- Geef elke tabel een titel die duidelijk weergeeft wat de tabel bevat.
- Probeer in de mate van het mogelijke elke tabel zelfstandig leesbaar te maken door te werken met labels, voetnoten of een duidelijke legende.
- Tabellen spreiden over meerdere pagina's is niet wenselijk. Kan het echt niet anders, herhaal dan de rij- en kolomkoppen.

Nieuwe softwaretoepassingen maken het steeds eenvoudiger om data snel en nauwkeurig om te zetten in welogende grafische voorstellingen. Het grote voordeel van grafieken is het bewerkstelligen van een duidelijk overzicht dat door de lezer op snelle manier kan worden overlopen. Daarenboven vereisen grafieken geen diepgaande kennis van de ruwe data. Grafische voorstellingen zijn in de eerste plaats geschikt om evoluties in de tijd, trends en relaties tussen verschillende groepen in kaart te brengen en te communiceren. Het grote voordeel van grafieken heeft betrekking op hun snelle absorptie en grote verstaanbaarheid voor een zeer breed publiek. Resultaten communiceren via grafieken is zowel geschikt voor interne als voor externe stakeholders. Zowel beleidsmakers als de media zijn geïnteresseerd in trends en vergelijkingen tussen bepaalde groepen. Een laatste voordeel is het feit dat uit onderzoek blijkt dat mensen meer onthouden over bepaalde resultaten wanneer deze gecommuniceerd worden via grafieken dan via tabellen en louter cijfers.

- Het taartdiagram

Diagrammen worden veelvuldig gebruikt in zeer uiteenlopende situaties en materies. Het taartdiagram is een vrij eenvoudige grafische voorstelling. De hele taart stelt het totale aantal voor (of 100%) van een variabele. Dit geheel wordt verdeeld in een aantal parten, dit om de relaties tussen de parten enerzijds en de omvang van parten anderzijds te duiden. Onderstaande figuur bevat als voorbeeld het totale aantal plaatsen in opvangvoorzieningen. De figuur maakt duidelijk welke opvangvoorzieningen momenteel bestaan en hoe groot hun aantal is in het totale aantal plaatsen in de opvangvoorzieningen.



Figuur 18: Voorbeeld van een taartdiagram

Bron: Kind & Gezin, *Jaarverslag 2004*, p.17.

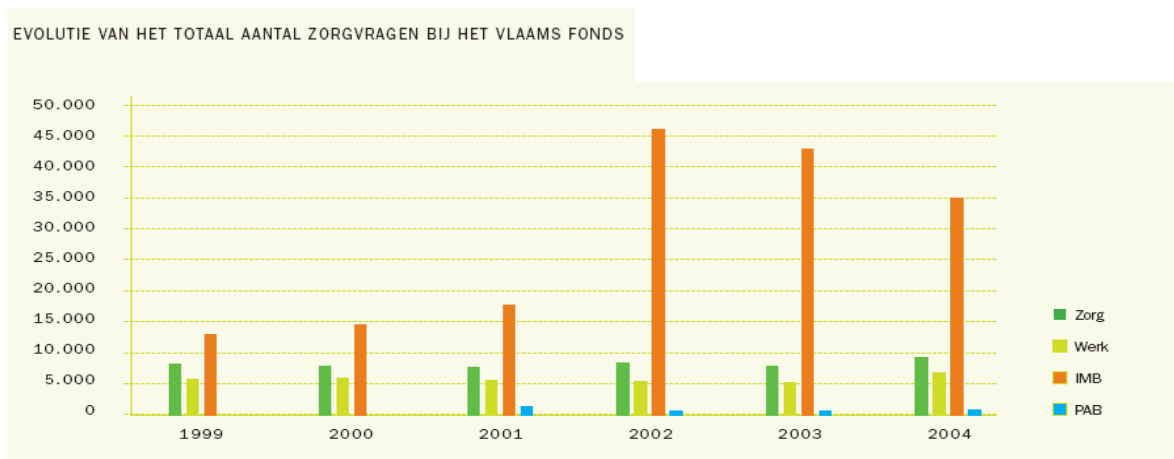
Enkele tips voor het opstellen van een taartdiagram:

- Gebruik slechts 6 of minder parten om de data te illustreren.
- Vergemakkelijk de leesbaarheid van het diagram door gebruik te maken van heldere, contrasterende kleuren. Eventueel als men slechts op één part dieper ingaat, kan men dit part met behulp van bestaande software uit de taart oplichten.
- Markeer de verschillende parten op een duidelijke en consistente wijze.
- Indien men opteert voor driedimensionale taartdiagrammen, wees u dan bewust van het feit dat parten die zich het kortst bij de lezer situeren, groter en dus belangrijker lijken.

- Het staaf – of stapeldiagram

Staafdiagrammen zijn de tweede soort die algemeen veel gebruikt wordt. Deze diagrammen zijn eveneens vrij eenvoudig te construeren en te begrijpen. Dit soort grafische voorstellingen zijn in het

bijzonder geschikt voor het weergeven van hoeveelheden geassocieerd met één variabele op verschillende tijdstippen, voor hoeveelheden van verschillende variabelen op hetzelfde ogenblik of voor de presentatie van de aantallen van onderdelen van één en dezelfde variabele. Staafdiagrammen zijn een veel toegepaste techniek om de resultaten van een monitoringsysteem te presenteren. Een staafdiagram kan slechts één bepaalde variabele of een cluster variabele bevatten. De onderstaande figuur illustreert dat een staafdiagram op een zeer overzichtelijke manier vier soorten zorgaanvragen (variabelen) kan weergeven over een tijdsperiode van vijf jaar. Door de staven verschillend in te kleuren en het gebruik ervan te duiden in een legende, wordt het voor de lezer een stuk comfortabeler om het diagram te begrijpen.

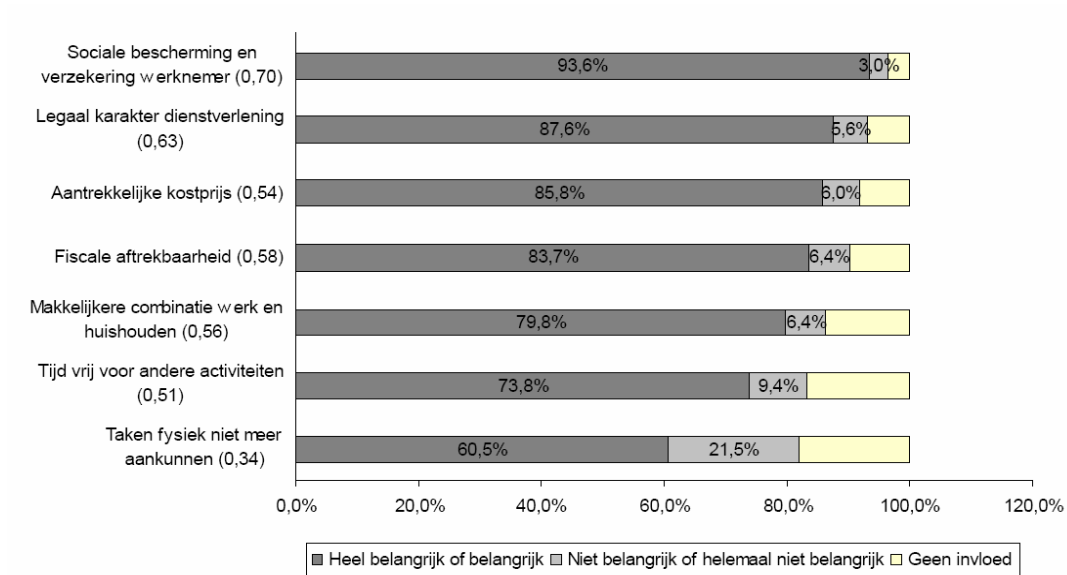


Figuur 19: Voorbeeld van een staafdiagram

Bron: Vlaams Fonds voor de Sociale Integratie van Personen met een Handicap, *Jaarverslag 2004*, p. 21

Het stapeldiagram is een staafdiagram dat verschillende variabelen bevat, het is een variant. Een stapeldiagram bevat verschillende variabelen, waarbij de balken boven elkaar worden geplaatst. Deze techniek is bijzonder geschikt voor de presentatie van cumulatieve gegevens. Het onderstaande voorbeeld maakt dit duidelijk: elke horizontale balk is een afzonderlijke variabele. De volledige balk bestaat in de onderstaande figuur uit 3 afzonderlijke balkjes (drie mogelijke antwoorden) die samen 100 procent vormen. Deze figuur toont aan dat een stapeldiagram heel wat variabelen kan bevatten die elk op hun beurt opgebouwd zijn uit een aantal onderliggende categorieën.

Motieven om beroep te doen op dienstencheques



Figuur 20: Voorbeeld van een stapeldiagram

Bron: IDEA Consult (2006). *Evaluatie van het stelsel van de dienstencheques voor buurtdiensten en – banen, eindrapport, in opdracht van de Federale overheidsdienst Werkgelegenheid, Arbeid en Sociaal Overleg.*

Enkele tips voor het opstellen van een staaf- en stapeldiagram:

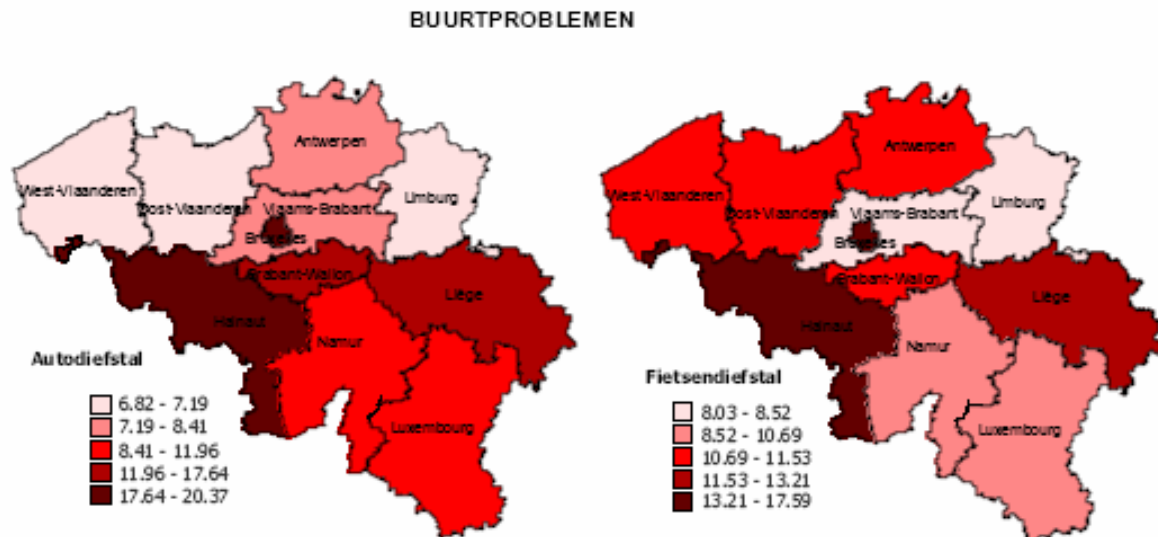
- Beperk het aantal staven, 6 wordt algemeen beschouwd als het maximum.
- Maak gebruik van kleuren om verschillen aan te duiden tussen de verschillende staven.
- Twee elementen maken het lezen en begrijpen van een staafdiagram heel wat eenvoudiger: (1) vermeld de waarde van elke staaf, ofwel boven de staven ofwel erin maar in een aangepaste kleur, (2) teken horizontale lijnen beginnend bij elk interval op de verticale as.
- Kleuren en symbolen op een doordachte en consequente manier inzetten zodat ze bijdragen aan het verhogen van de leesbaarheid van diagram.

- Cartogram

Een cartogram is een geografische kaart waarop statistische informatie is weergegeven. Deze techniek laat toe op een overzichtelijke manier vergelijkingen te maken tussen geografische gebieden. Dankzij de technologische vooruitgang biedt deze techniek tal van mogelijkheden. Een voorbeeld hiervan is het GIS (Geografisch Informatie Systeem), dat toelaat informatie te koppelen aan geografische eenheden. Het gebruik van geografische kaarten biedt tal van voordelen voor de presentatie van de resultaten aan een

breed publiek. In één oogopslag krijgt de geïnteresseerde een volledig overzicht van de resultaten. Het onderstaande voorbeeld maakt dit meteen duidelijk.

Enkele buurtproblemen naargelang de provincies in 2004. De grafiek herneemt enkel het antwoord « helemaal wel » op de vraag: Zijn de volgende feiten een probleem in uw buurt?



Figuur 21: Voorbeeld van een cartogram

Bron: *Veiligheidsmonitor 2004, Analyse van de federale enquête*, p.13

Enkele tips voor het opstellen van een cartogram:

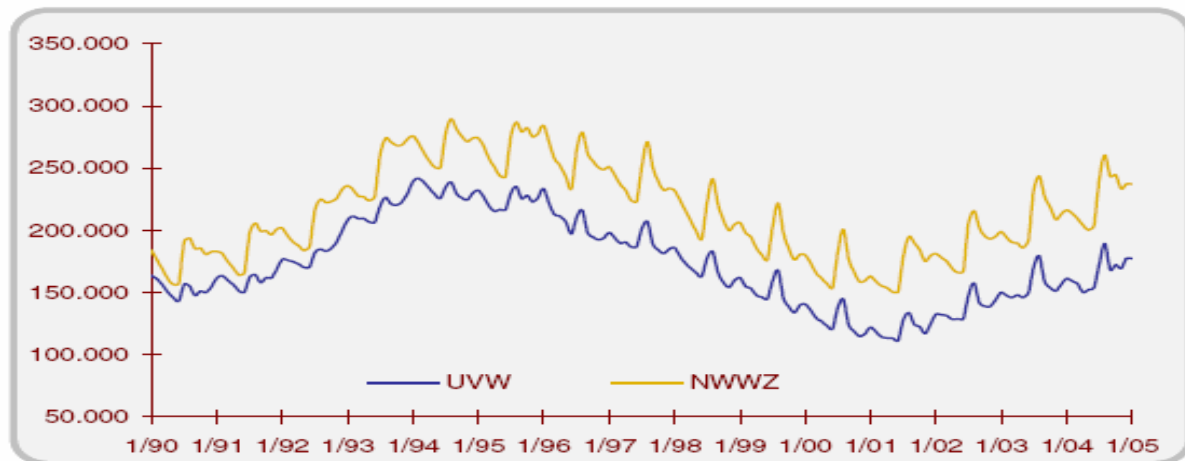
- Besteed voldoende aandacht aan de uittekening van de legende, eventueel uitgebreid tot een korte toelichting van de weergegeven informatie.
- Maak gebruik van kleuren, let op dat de lezer een duidelijk overzicht krijgt van de geografische verschillen.
- Bij het elektronisch ter beschikking stellen van deze informatie kan een kaart zeer interactief worden toegepast. Het kan de lezer mogelijk gemaakt worden om door te klikken op individuele gebieden verdere, bijkomende informatie te verkrijgen.

- Lijngrafiek

Naast diagrammen kunnen resultaten van een monitoringsysteem gepresenteerd worden aan de hand van tweedimensionale grafieken. Deze grafieken zijn bijzonder geschikt voor de presentatie van twee informatietypes, wanneer verschillende punten uitgezet moeten worden of als er zich wijzigingen voordoen tussen deze punten. Data worden uitgezet op de grafiek die één of meer lijnen vormen, dit laat

toe om zeer complexe kwantitatieve data op een visuele manier voor te stellen. De volgende figuur verduidelijkt dit. De evoluties van de uitkeringsgerechtigde volledig werklozen (UVW) en de niet-werkende werkzoekenden (NWWZ) worden grafisch voorgesteld. De maandgemiddelden vanaf 1990 tot 2005 zijn weergegeven. De grafiek laat toe op een eenvoudige en snelle manier de trends over een periode van vijftien jaar voor deze twee variabelen te begrijpen.

Evolutie UVW – NWWZ



Figuur 22: Voorbeeld van een lijngrafiek

Bron: VDAB, *De arbeidsmarkt in beeld, Maandverslag 2004*.

Enkele tips voor het uitwerken van lijngrafieken:

- Label de verschillende lijnen in plaats van gebruik te maken van een legende.
- Drie lijnen weergegeven is het maximum, indien mogelijk dient elke lijn een andere kleur te hebben, waarbij de belangrijkste lijn (variabele) de meest heldere kleur heeft.
- Indien enkel een zwartwit publicatie mogelijk is, voorziet dat de lijnen een verschillende dikte of stijl (bv. stippellijn) hebben.
- De titel van de grafiek steeds op een consequente en duidelijke manier presenteren.

Algemene tips voor het gebruik van tabellen en figuren

- *Denk na over de essentie van de boodschap en hoe de boodschap communiceren op de meest efficiënte en effectieve manier naar de verschillende doelgroepen.*

Informatie overbrengen is cruciaal voor het welslagen van een monitoringsysteem. Tabellen en figuren hebben tot doel de leesbaarheid en begrijpbaarheid te verhogen. Omwille van deze reden is het

belangrijk dat de auteurs nadenken over het gebruik van deze hulpmiddelen, alleen als ze een echte meerwaarde betekenen, worden ze opgenomen in de rapportering.

- *Breng alle mogelijke, relevante data samen.*

Alvorens over te gaan tot de ontwikkeling van tabellen en figuren, dient het team over alle mogelijke relevante data beschikken. Het is op basis hiervan dat de meest aangewezen tabellen en figuren worden ontwikkeld.

- *Maak de rapportering zo eenvoudig mogelijk.*

Hoe eenvoudiger de tabellen en figuren hoe makkelijker te begrijpen voor de lezer. Zeer ingewikkelde constructies leiden alleen maar tot onbegrip en frustratie. Men dient te vertrekken vanuit de kenmerken van het doelpubliek. Afhankelijk van hun voorkennis en interesse dient het team de meest aangewezen tabellen en figuren op te stellen.

- *Voorzie elke tabel of figuur van een hoofding.*

Tabellen en figuren in de tekst dienen voorzien te zijn van een duidelijke nummering en titel. De titel heeft als belangrijkste taak duidelijk weer te geven welke informatie de tabel of figuur bevat. Daaruit volgt dat titels zelfstandig leesbaar moeten zijn.

- *Bij het opnemen van tabellen en figuren, bespreek deze ook in de tekst.*

De tekst dient te verwijzen naar elke weergegeven figuur of tabel. Vaak zal bijkomende uitleg noodzakelijk zijn, zodat de lezer de informatie op een correct manier analyseert en begrijpt. Men kan er nooit van uitgaan dat de lezer de tabellen en grafieken begrijpt.

- *Zorg ervoor dat de tabellen en figuren toegankelijk zijn in de tekst.*

Vaak zal aan het begin van de tekst een lijst van tabellen en grafieken weergegeven zijn, zodat lezers die slechts een algemeen zicht willen verkrijgen op de behaalde resultaten snel de gevraagde informatie kunnen terugvinden. Omwille van deze reden is het tevens belangrijk om alle tabellen en grafieken van een nummer te voorzien op een consequente wijze.

- *Behoud voldoende tijd om tabellen en figuren te ontwerpen.*

Het ontwerpen van duidelijke en overzichtelijke tabellen vereist de nodige inspanning, vaak zullen verschillende ontwerpen nodig zijn alvorens de gewenste tabel bekomen wordt. Gelet op het belang van deze tabellen en figuren voor het welslagen van de communicatie, is het uittrekken van de nodige tijd en middelen een absolute noodzaak.

Besluit

De uiteindelijke output van een monitoringsysteem zal in de meeste gevallen één of andere vorm van rapportage zijn. Vandaag de dag zijn verschillende methoden en technieken voorhanden om de

resultaten te presenteren aangepast aan de noden en verwachtingen van het doelpubliek. Organisaties hebben de opdracht om hun werking en verwezenlijkingen voor te leggen aan een heel resem stakeholders, gaande van het interne management, over de beleidsverantwoordelijken tot het brede publiek. De meerwaarde van een dergelijke rapportering ligt in de snelheid waarmee de lezer bepaalde informatie kan opzoeken en begrijpen.

Om de communicatie zo goed mogelijk te laten verlopen speelt de aard van de informatie een belangrijke rol in de keuze van de meest geschikte communicatievorm. Anderzijds hebben ook de eigenschappen van het doelpubliek een onmiskenbare invloed. Het is niet ongewoon om dezelfde informatie op een andere manier te communiceren aan een verschillend publiek. Creativiteit in de keuze is belangrijk evenals het durven vernieuwen om de informatie zo goed mogelijk over te brengen. Vertrouwd zijn met het brede gamma aan communicatiemediën is een essentiële voorwaarde om te komen tot creatieve oplossingen. Het gamma breidt nog steeds uit als gevolg van de snelle technologische ontwikkelingen.

4. Invoering, beheer en gebruik van een monitoringsysteem: enkele kritieke succesfactoren

In voorgaande paragraaf werden de verschillende stappen in het proces van monitoring toegelicht. De eerste fasen, namelijk het formuleren van de opzet, het bepalen van de scope, het selecteren en ontwerpen van indicatoren en het bepalen van normen, slaan tegelijk op het design van het meetsysteem. De volgende en noodzakelijke tussenstap naar een operationeel meetsysteem is de invoering ervan. Eens in werking, is het beheer en onderhoud van het meetsysteem belangrijk. Wegens het intrinsieke belang van deze twee fasen met het oog op de relevantie van het systeem voor de gebruikers, formuleren we in deze paragraaf enkele kritieke succesfactoren.

4.1 Het implementatieproces

Net als bij de stappen van het ontwerp van een meetsysteem is ook voor een geslaagde invoering de betrokkenheid van relevante actoren belangrijk: enerzijds zij die instaan voor de werking van het meetsysteem en anderzijds zij die het gaan gebruiken, d.w.z. de stakeholders in bredere zin.

Een monitoringsysteem is opgevat als een duurzaam instrument. De invoering ervan gebeurt echter best via een *projectmatige aanpak*. Dergelijke benadering bevordert de zichtbaarheid, de oplisting van deeltaken met betrekking tot de werking van het meetsysteem, de verdeling van verantwoordelijkheden voor deze deeltaken, en een gerichte communicatie naar alle betrokkenen.

Voor de invoering is *leiderschap* nodig. In een kleine dienst kan dat één persoon zijn, in een grotere afdeling of administratie een team van personen die de coördinatie en verantwoordelijkheid op zich nemen. Zij vormen het aanspreekpunt voor andere betrokkenen tijdens de implementatie van het meetsysteem.

Communicatie naar de betrokkenen bij de werking van het meetsysteem is een andere kritieke factor voor een geslaagde invoering. Daarbij moet aandacht gaan naar de verschillende aspecten van het meetsysteem, en in het bijzonder naar het conceptueel kader en de opzet van het meetsysteem. Het moet voor alle betrokkenen duidelijk zijn wat het motief en de context is voor de dataverzameling. Dit laatste is immers nooit een doel op zichzelf, maar past in een breder proces van beleidsevaluatie. Voor de efficiënte en effectieve werking van het systeem is het belangrijk dat elke betrokkene inzicht heeft in het belang van zijn of haar deeltaak in dat bredere proces in functie van de opzet.

De invoering van het meetsysteem moet gebaseerd zijn op een *goede schatting van de tijdsinvestering* voor zij die verantwoordelijk zijn voor de operationele werking. Daarmee samen hangt de *verdeling van de verantwoordelijkheden en deeltaken*.

Het is ook van belang om de *totale kostprijs* van de werking van het meetsysteem, met name de datavergaring en -verwerking, in te calculeren. Deze kostprijs is gerelateerd aan een hele reeks factoren, onder meer:

- de inhoudelijke scope van het meetsysteem;
- het aantal indicatoren;
- de aard van de gegevens die men wenst te verzamelen;
- de technieken van dataverzameling;
- de frequentie van dataverzameling;
- de mate waarin de ruwe gegevens nog verwerkt moeten worden, bv. uitsplitsing of aggregatie, met het oog op de analyse ervan;
- het format van rapportering.

De totale kostprijs mag echter niet het enige criterium zijn bij een financiële afweging in het kader van de principiële beslissing over het ontwerp van een monitoringsysteem. Dergelijke afweging moet gebeuren aan de hand van een kosten-batenanalyse. Daarbij moet men in het achterhoofd houden dan de kosten de baten vooraf gaan in de tijd.

4.2 Het beheer

De invoering van een operationeel monitoringsysteem is geen eindpunt. Het beheer van het meetsysteem is en blijft een belangrijk aandachtspunt vanaf dan. Ook daarvoor zijn een aantal kritieke slaagfactoren op te lijsten.

Het beheer en onderhoud van een meetsysteem impliceren dat een *periodieke evaluatie van het meetsysteem* gebeurt, zowel op technische als inhoudelijk vlak.

Op technisch vlak is de efficiëntie van gegevensverzameling en -verwerking een belangrijk aandachtspunt. Het gaat daarbij om het niveau (graad van detail) van registreren, de economie van databanken, de toegankelijkheid van deze databanken, de compatibiliteit tussen verschillende databestanden en -formaten en de extractiemogelijkheden in het kader van rapportering.

Een duidelijke verdeling van de deeltaken en toewijzing van verantwoordelijkheden in het kader van de werking van het meetsysteem is een noodzakelijke voorwaarde. Een hulpmiddel daartoe is het opstellen van een technische fiche per indicator die deel uitmaakt van de het monitoringsysteem (cf. supra). Dergelijke technische steekkaart bevat volgende gegevens:

- | |
|---|
| <ul style="list-style-type: none">- Definitie- Categorisering- Formulering- Beïnvloedende factoren- Meetfrequentie- Frequentie van aanlevering- Bron- Clustering- Tijdreeks- Normering- Data opname in set- Wijzigingen aan omschrijving |
|---|

Tabel 10: Elementen voor een technische indicatorfiche

Vanzelfsprekend dient waar mogelijk de voorkeur gegeven aan een elektronische of digitale wijze van dataverzameling, verwerking en -beheer. Het meetsysteem is immers een noodzakelijke maar nog steeds een intermediaire stap in een breder proces van analyse en beoordeling (evaluatie). Tegenover een uitbreiding van de scope van het meetsysteem dient een adequate aanpassing van de verwerkingscapaciteit (personeel, tijd, knowhow, software,...) te staan.

Op inhoudelijk vlak dient de evaluatie van het meetsysteem aandacht te hebben voor de link tussen de indicatorenset en het beleid dat men wenst te evalueren. Het beoordelen van de sterkte van deze link is een cyclische opdracht: ze moet herhaald worden telkens wanneer het beleid wijzigt. Het meetsysteem moet immers afgeleid zijn van de achterliggende *beleidstheorie* en ingebed zijn in het beleidsproces. Als het beleidstheoretische kader verandert, hetzij op het vlak van maatregelen, hetzij op het vlak van doelstellingen of een combinatie van deze twee, moet de indicatorenset wellicht aanpassingen ondergaan.

Dat kan op verschillende manieren:

- het schrappen van indicatoren die niet meer relevant zijn;
- het vervangen van indicatoren door andere;
- het licht aanpassen van bestaande indicatoren (de omschrijving en/of berekening);
- het toevoegen van nieuwe indicatoren.

Ook *eerdere evaluaties* van het beleid kunnen aanleiding vormen om de indicatorenset te wijzigen, bijvoorbeeld wanneer bepaalde *neveneffecten* in beeld komen die men relevant acht om verder op te volgen omdat ze belangrijk zijn om de effectiviteit van het beleid te beoordelen in de toekomst.

Aanpassingen aan de indicatorenset kunnen ook gebeuren zonder dat het beleid verandert. Een *verschuiving in de evaluatiecriteria* kan ook gevolgen hebben voor het aandeel van bepaalde types indicatoren in de set: een shift van efficiëntie naar effectiviteit bijvoorbeeld zal impliceren dat meer effectindicatoren en omgevingsindicatoren moeten worden ontworpen.

Bestaande indicatoren kunnen ook opnieuw afgetoetst worden aan de oorspronkelijke *selectiecriteria* (cf. supra: 3.3). De ontwikkeling van nieuwe meettechnieken bijvoorbeeld kan ertoe bijdragen dat relevant geachte indicatoren nu wel opgevolgd kunnen worden aan een aanvaardbare kostprijs.

Wanneer men de indicatorenset aanpast, dient men steeds oog te hebben voor de *beheersbaarheid van de indicatorenset als geheel*: een te grote set brengt een te grote administratieve last mee om de gegevens te verwerken, verzwaart nodeloos de rapporteringen en kan de focus van evaluaties in het gedrang brengen.

Aanpassingen aan de indicatorenset zullen vaak gepaard gaan met een dilemma tussen continuïteit enerzijds en relevantie anderzijds. Continuïteit in de set maakt vergelijkingen in de tijd mogelijk wat belangrijk is om trends in kaart te brengen. Relevantie kan echter sterker doorwegen omdat het hier gaat om de finaliteit van het monitoringsysteem in functie van het beleid en de evaluatievragen.

4.3 Gebruik

De meerwaarde van gegevens verzameld via monitoring ligt in de mate dat ze bijdragen tot het formuleren van de antwoorden op de (evaluatie)vragen van de gebruikers. Een hanteren van een *gebruikersfocus* bij het opzetten van een monitoringsysteem is dan ook een kritieke succesfactor.

Naast de inhoudelijke overeenstemming met het beleidstheoretische kader, is bij een gebruikersfocus ook de stap van *rapportering* van belang. In de rapportering van de evaluatie worden gegevens, evaluatiecriteria en de resultaten van de toetsing of de beoordeling toegelicht en samengebracht. Deze rapportering dient qua diepgang, terminologie, en volume afgestemd op de aard van de gebruiker of stakeholder.

Dat dergelijke aanpak vanuit het perspectief van de gebruiker altijd 'werkt', is niet gegarandeerd: het voorhanden zijn van bruikbare informatie leidt niet vanzelf tot *daadwerkelijk gebruik* van deze informatie. Daar hangt veel af van de gebruikers zelf. Een gebruikersfocus draagt er wel toe bij.

Het gebruik van monitoring moet gezien worden als een *leerproces*. Zo zal de eerste versie van een indicator zelden meteen accuraat genoeg bevonden worden. Vandaar het nut van een periodieke evaluatie van het meetsysteem om de relevantie van de bestaande indicatoren te beoordelen en waar mogelijk te verbeteren. Ook het interpreteren van de gegevens vergt een zekere ervaring en skills. Het is van belang dat de gebruikers van de gegevens kennis hebben van de mogelijkheden maar ook de intrinsieke beperking van monitoring als een ondersteunend instrument voor evaluaties.

5. Uitbouwsценario voor monitoring: van een beheers- naar een beleidsperspectief

De laatste jaren hebben heel wat overheidsorganisaties geïnvesteerd in de ontwikkeling van meetsystemen. In de meeste gevallen gebeurde dit in de eerste plaats ten behoeve van het management. Het meetsysteem wordt opgevat als een managementinformatiesysteem (MIS). Op basis van de informatie wil het management de interne werking en de hiermee samenhangende resultaten verbeteren.

Het management staat niet los van de beleidsvoering binnen de overheid. Een goede interne werking is een voorwaarde voor een efficiënt en doeltreffend beleid. Meer nog, de interne werking van een overheidsorganisatie is geen doel op zich. Vanuit een beleidsperspectief is ook de vraag naar doelbereiking en de effectiviteit van het beleid essentieel. Tal van organisaties zijn immers betrokken bij de planning en/of uitvoering van beleid.

Wanneer men ook de beleidscontext wil opvolgen, naast het beheer, wordt de scope van meten en monitoring breder. In de praktijk zal dit vaak betekenen dat de bestaande meetsystemen uitgebreid moeten worden, van een beheersperspectief naar een meer omvattend beleidsperspectief. In wat volgt gaan we in op de belangrijkste aandachtspunten bij dit uitbouwsценario voor monitoring.

5.1 Van output naar effect

Uit onderzoek blijkt dat de meeste meetsystemen zich in de eerste plaats richten op input, processen en output, maar niet op effecten (Perrin, 1998; Mintzberg, 1996). Dergelijke prestatiemeetsystemen worden vandaag de dag vaak gekoppeld aan het afleggen van verantwoording. De ondersteuning van beheersovereenkomsten is een groeiende toepassing.

Het motief van verantwoording ligt aan de basis van het feit dat deze meetsystemen de focus hoofdzakelijk leggen op output en dat effecten vaak buiten beschouwing worden gelaten. De reden heeft te maken met de organisatiegrens. Organisaties kunnen hun output sturen en beheersen, maar hebben in vergelijking minder vat op de maatschappelijke effecten van hun prestaties (dienstverlening, producten, collectieve goederen). In het kader van verantwoordingsrelaties is er dan ook een tendens om op de output te focussen, vanuit een redenering dat overheidsorganisaties niet verantwoordelijk kunnen gesteld worden voor iets waarop ze niet volledig vat hebben.

Vanuit een beleidsperspectief is aandacht voor effecten echter vanzelfsprekend. De opdracht van de overheid ligt immers niet binnen de organisatiegrens. Het uiteindelijke doel van overheidsacties ligt in het realiseren van bepaalde effecten in de samenleving. Om deze effecten op een consistente en eenduidige

manier in kaart te brengen, leveren monitoringsystemen een belangrijke meerwaarde. Het meten van effecten maakt het ook mogelijk om de vraag te stellen naar de effectiviteit en impact van het gevoerde beleid, naast de aandacht voor doelbereiking.

Beleidsmakers verantwoordelijk stellen voor de manier waarop ze gebruik maken van feedback en evaluatieresultaten ter verbetering van het actuele beleid, is een uitbreiding van het verantwoordingsmotief. Tegelijk verwijst men daarmee naar andere motieven om te meten, naast het verantwoordingsmotief en het verbeteren van de interne werking van de organisatie in het kader van de beleidsuitvoering. Het ondersteunen van beleidsplanning en beleidsleren zijn andere belangrijke motieven in het kader van monitoring en evaluatie van beleid. Daarvoor is de opvolging van effecten onontbeerlijk.

Dat alles betekent niet dat de aspecten die de bestaande meetsystemen opvolgen niet meer van tel zijn. Maar naast de interne werking dient ook de externe dimensie aan bod te komen. Effecten vervangen de bestaande focus van veel meetsystemen dus niet om beleid op te volgen, maar zijn een aanvulling daarop. Een eenzijdige focus op effecten is evenzeer problematisch als totaal geen aandacht hebben voor effecten en enkel voor input, processen en output. Het management alleen beoordelen bijvoorbeeld op basis van de gerealiseerde effecten is geen goede werkwijze, daar tal van factoren kunnen meespelen die zich buiten het bereik van het management situeren. Met andere woorden, te veel aandacht voor verantwoording enerzijds en een te zware nadruk op effecten anderzijds kan perverse effecten genereren. Deze stelling onderschrijven tal van auteurs die managementprocessen beschreven en geanalyseerd hebben. (Perrin 1999; Mintzberg 1994, Deming 1986). Een lerende organisatie die kwaliteitsverbetering en innovatie hoog in het vaandel voert is het centrale uitgangspunt van deze auteurs.

De bottom-line in een uitbreidingscenario van een beheers- naar een beleidsperspectief voor monitoring is de aanvullende en complementaire aandacht voor effecten. In de volgende paragraaf gaan we in op enkele specifieke aspecten met betrekking tot het meten van effecten en hoe effectinformatie gebruikt kan worden.

5.2 Monitoren van effecten

5.2.1 Moeilijker te meten?

Vaak is het zo dat effecten moeilijker te meten zijn dan input, proces of output. Dit heeft te maken met een aantal factoren. Effecten zijn soms minder tastbaar. Dat maakt dat ze moeilijker te omschrijven en af

te bakenen zijn. Bijgevolg wordt het moeilijker om ze in meetbare termen uit te drukken. De grijpbaarheid van effecten varieert echter naargelang de beleidssector. Bij onder meer economisch beleid, ruimtelijke ordening, mobiliteit en sociaal beleid stelt deze problematiek zich minder dan bijvoorbeeld bij cultuurbeleid of buitenlands beleid.

De moeilijkheid van het meten heeft ook te maken met de vraag of men spreekt over effecten *tout court* of over beleidseffecten. De eerste kunnen we meten, zonder meer. De laatste verwijzen naar de effectveranderingen die daadwerkelijk aan het beleid zijn toe te schrijven. Dit impliceert kennis over oorzaak en gevolg relaties ofwel beleidseffectiviteit. Het aantonen en bepalen van beleidseffectiviteit kan niet door monitoring alleen. Het is een geavanceerde opzet van beleidsevaluatie waarvoor verschillende strategieën gevolgd kunnen worden die in deel 1 van de handleiding uitvoering uiteengezet worden.

Alleszins is het belangrijk om effecten die het beleid nastreeft zo meetbaar mogelijk te formuleren met het oog op monitoring. De belangrijkste bronnen hiervoor zijn de missie en de beleidsdoelstellingen. Vaak hebben beleidsdocumenten een algemeen karakter en is er een gebrek aan gedetailleerde beschrijvingen van de mogelijke effecten. In het kader van evaluatiedoelinden bevat een beschreven effect in het beste geval een beschrijving van de kenmerken, het gedrag of voorwaarden die de maatregel, een project of beleidprogramma wenst te wijzigen. In vele gevallen zullen de bestaande effectbeschrijvingen, alvorens deze te monitoren, verder moeten geëxpliciteerd worden. Stakeholders zullen vaak verschillend staan tegenover de beleidsdoelstellingen. Het is belangrijk te weten welke effecten als prioritair naar voor worden geschoven om deze mee te nemen in het monitoringsysteem. Verder kan reeds bestaand onderzoek een bron zijn om effecten te benoemen en te concretiseren met het oog op een follow-up.

5.2.2 Gedifferentieerde effecten

Effecten zijn niet altijd eenduidig. Vaak verschillen ze naargelang kenmerken van leden van de doelgroep (personen, locaties, organisaties). De specifieke context speelt vaak een belangrijke rol. In veel gevallen is het dan ook belangrijk om gedetailleerde effectinformatie te hebben. Zo kan een antidrugscampagne in scholen een gunstig effect hebben op het gebruik van cannabis, en tegelijk weinig of geen effect hebben op het rookgedrag en alcoholgebruik van leerlingen. Maatregelen in het verkeersbeleid kunnen dalingen van licht gewonden met zich mee brengen, terwijl het aantal zwaar gewonden stabiel blijft. Effectindicatoren dienen dus voldoende gevoelig te zijn om deze schakeringen of nuanceringen op te pikken.

Anderzijds is het een uitdaging om een totaalbeeld van effecten te verkrijgen. Beleidsissues en – problemen zijn vaak onderling verbonden. Wil men de effecten van maatregelen opvolgen, dient men oog te hebben voor deze verwevenheid om de monitoringinformatie op een zinvolle manier te kunnen gebruiken. Ook de tussenliggende variabelen of determinanten zijn belangrijk om op te volgen als men output en effect relaties wil onderzoeken.

5.2.3 De rol van omgevingsfactoren

De interpretatie van effecten is een grote uitdaging, een verantwoorde interpretatie vereist het in overweging nemen van omgevingsfactoren waarin het beleid tot stand komt, bepaalde gebeurtenissen die zich voordoen tijdens de implementatie van het beleid en karakteristieken van de doelgroep die tevens veranderen door de tijd. Neem een opleiding voor werklozen. De interpretatie van de effectindicator, het aantal mensen terug aan het werk zes maanden na de opleiding, dient verklaard te worden op basis van bijkomende informatie. Naast de karakteristieken van de doelgroep, zoals opleidingsniveau, de duur van de werkloosheid, het sociale milieu, spelen ook economische factoren, zoals de vraag naar arbeid en de economische groei. Monitoring van beleid vergt dus niet enkel de ontwikkeling van effectindicatoren, maar ook indicatoren om relevante omgevingsvariabelen op te volgen. Zo kan men veranderingen of effecten beter interpreteren en duiden. In het conceptueel raamwerk van het veranderingsmodel worden dergelijke externe factoren voor een deel meegenomen via de zogenaamde matigende variabelen: ze hebben een invloed op de relaties tussen output, tussenliggende variabelen –ook wel determinanten of hefboomen genoemd- en de uiteindelijke effecten.

5.2.4 Effectenketting

Vaak beogen maatregelen, projecten of beleidsprogramma's meer dan één effect. Deze effecten treden niet steeds gelijktijdig op maar vaak na elkaar, waarbij één effect bijdraagt tot het optreden van een ander. Deze aaneenschakeling van effecten houdt een reeks 'als – dan' relaties in. Bijvoorbeeld, als de overheid een fiscale gunstmaatregel voor het installeren van isolatiemateriaal lanceert, dan zullen mensen meer isolatie plaatsen. Als de mensen meer isolatie plaatsen zijn de huizen beter geïsoleerd. Beter geïsoleerde huizen leiden tot daling in het verbruik van fossiele energiebronnen. Als minder fossiele brandstoffen verbruikt worden, dan draagt dit bij tot de verlaging van de luchtverontreiniging. Afname van de luchtverontreiniging komt ten goede aan de volksgezondheid. Hoe verder men gaat in deze keten, hoe verder het effect verwijderd ligt van de invloedssfeer van de maatregel, het project of beleid. En des te meer omgevingsfactoren een invloed kunnen hebben op de effectwaarden. Hoe verder in de tijd, des te moeilijker het wordt om de verbanden tussen beleid, omgeving en effectveranderingen te ontrafelen.

Nochtans is het belangrijk om lange termijneffecten mee op te volgen. Om terug te grijpen naar het voorbeeld, het volstaat niet dat de overheid in haar evaluatie alleen het aantal vierkante meter geïnstalleerde isolatie beschouwt als het effect van haar maatregel. Men dient ook te kijken naar effecten die verderaf liggen en de vraag te stellen of de huizen inderdaad beter geïsoleerd zijn. Aan de andere kant mogen de effecten ook niet te veraf liggen, de vooropgestelde effecten moeten kunnen gelinkt worden aan de maatregel. Zo wordt het al problematisch om het lager verbruik van fossiele brandstoffen te koppelen aan de fiscale maatregel ten gunste van de verkoop van isolatiematerialen. Tal van andere mogelijke factoren hebben daarop ook een invloed. Het is dan ook zeer moeilijk om deze effecten te

koppelen aan de onderzochte maatregel, of anders gesteld, om het effect van de maatregel af te zonderen.

5.2.5 Wat met neveneffecten?

Dat effecten verschillen naargelang tijdstip van optreden, intensiteit en duurtijd werd reeds besproken (cf. supra). Men kan ze ook indelen op basis van hun belangrijkheid en hun appreciatie vanuit een bepaald standpunt of beleidsopzet. Zo onderscheidt men hoofd- en neveneffecten, en positieve en negatieve effecten. De twee laatste noemt ook gewenste of ongewenste effecten. Verder kan men nog opdelen tussen bedoelde en onbedoelde effecten.

Beoordeling	Belangrijkheid en intentie			
	Hoofdeffect		Neveneffect	
	Bedoeld	Onbedoeld	Bedoeld	Onbedoeld
<i>Positief / gewenst</i>				
<i>Negatief / ongewenst</i>	-		-	

Tabel 11: Effecttypes op basis van belangrijkheid, intentie en appreciatie

Uit deze indeling blijkt dat neveneffecten niet noodzakelijk een negatieve connotatie hebben. Neveneffecten kunnen zelfs bedoeld zijn vanuit het beleid. Het gaat niet om de voornaamste effecten die men met een beleidsinitiatief beoogt, maar om oversijpelingseffecten of onrechtstreekse, secundaire effecten.

Merk op dat de opdeling positief /negatief niet zomaar samen valt met bedoeld/onbedoeld. Het creëren van negatieve effecten is in principe nooit de intentie van beleidsmakers. De combinatie 'bedoelde negatieve effecten' kunnen we dan ook schrappen in de bovenstaande tabel (-). Maar het is niet denkbeeldig dat er positieve effecten vastgesteld worden die niet in de doelstellingen van het beleid vervat liggen. Ze vormen als het ware een meevaller vanuit beleidsoogpunt die niet ingecalculeerd was.

De aandacht van beleidsmakers gaat doorgaans uit naar twee types. Enerzijds de bedoelde hoofdeffecten die in de beleidsdoelstellingen vervat liggen en dus d'office als gewenst of positief worden beschouwd, en anderzijds de onbedoelde neveneffecten van negatieve aard.

Het monitoren van neveneffecten is omwille van een aantal redenen een zeer moeilijk gegeven. In de eerste plaats zijn bepaalde effecten, zeker in innovatief beleid, vaak onbekend en manifesteren ze zich na een lange tijdsperiode. In de tweede plaats zijn politici vaak enkel geïnteresseerd in de voorziene, korte termijn effecten. Zo wordt er vaak geen rekening gehouden met of onvoldoende aandacht besteed aan de identificatie en opvolging van mogelijke neveneffecten. Het monitoren van neveneffecten vraagt dan ook een bijkomende inspanning. Voorafgaand onderzoek van vergelijkbare maatregelen of beleid is dan ook zeer zinvol en kan een indicatie geven welke neveneffecten zich kunnen voordoen. Hetzelfde geldt voor informatie over de beleidssetting. Tot slot zijn stakeholders die een taak hebben in de uitvoering van beleid goed geplaatst om aan te geven waar zich mogelijke problemen kunnen voordoen.

Beleidsmakers kunnen verder gebruik maken van eerdere ervaring en inzichten om mogelijke neveneffecten te definiëren.

Anderzijds mag de aandacht voor neveneffecten ook weer niet te ver gaan. Stel dat een maatregelenpakket ter verhoging van de verkeersveiligheid vruchten afwerpt: het aantal dodelijke slachtoffers daalt aanzienlijk. Een onbedoeld negatief neveneffect hiervan is echter dat er een tekort dreigt aan orgaandonoren. Betekent dit dat men het maatregelenpakket moet herzien? Beleidsplanning die tot in het oneindige rekening probeert te houden met negatieve neveneffecten, zal in extremis resulteren in een non-beleid.

5.2.6 Kwalitatieve informatie als complement

Ten slotte wijzen we op het belang van kwalitatieve informatie als noodzakelijk complement van kwantitatieve informatie via indicatoren en monitoring. Zeker wanneer het gaat om effecten en effectveranderingen te interpreteren en te verklaren, is dergelijke informatie vaak onontbeerlijk. Indicatoren en zeker effectindicatoren hebben hun beperkingen. Ze leveren beschrijvende informatie op (cf. supra), maar geen verklaringen. De aandacht voor omgevingsfactoren en mogelijke neveneffecten is een belangrijk maar nog onvoldoende om het verband te leggen tussen beleidsoutput en effecten. Geïnformeerd beleid, met behulp van monitoring en evaluatie, dient te steunen op kwantitatieve en kwalitatieve informatie.

6. Besluit

In het slothoofdstuk beklemtonen we nogmaals het belang en de meerwaarde van monitoring van beleid in het kader van beleidsevaluatie. We nemen een drievoudige stelling in: *monitoring in de overheid is mogelijk, nuttig en noodzakelijk*.

In 1999 hanteerden Bouckaert e.a. voor een gelijkaardige stellingname over prestatiemeting een bewijsvoering 'ex absurdum'. Door de tegenargumenten te weerleggen werd de correctheid van de stelling aangetoond. Ondertussen heeft meten en het werken met indicatoren reeds een hele weg afgelegd binnen de overheid. Een negatieve bewijsvoering lijkt vandaag minder noodzakelijk, velen hebben de voordelen van meten leren kennen. Dat betekent niet dat er ook vandaag geen kanttekeningen en randvoorwaarden meer zijn waaraan aandacht dient gegeven in de praktijk van beleidsmonitoring.

Monitoring van beleid in is in de eerste plaats *mogelijk*. Dat wil niet zeggen dat het steeds even gemakkelijk is. Praktische hindernissen beletten vaak dat men de optimale informatie verkrijgt in termen van betrouwbaarheid, validiteit, tijdigheid, het gewenste niveau van detail, enz. Ook de meetbaarheid van het beleid verschilt naargelang de beleidssector of het concrete meetobject.

Monitoring is ook een leerproces: zelden zal men meteen de ideale indicatorenset kunnen samenstellen. De zoektocht naar goede indicatoren is een verhaal van graduele verbetering en verfijning. Binnen de overheid heeft men ondertussen al heel wat ervaring en expertise opgebouwd met indicatoren en meten. Beleidssectoren waarbinnen men helemaal niet met indicatoren werkt, zijn er niet meer. Een screening van beleidsnota's, beleidsbrieven, planningsdocumenten en jaarverslagen levert het bewijs.

Het meten van effecten is relatief moeilijker dan het meten van input, proceskenmerken en output. Toch werd ook al een hele weg afgelegd binnen de overheid om strategische en operationele beleidsdoelstellingen in meetbare termen te stellen. Daarin liggen in principe de voornaamste effecten vevat die men met het beleid nastreeft. Het hebben van degelijk uitgeklaarde doelstellingen hebben we ook aangewezen als een noodzakelijk vertrekpunt om monitoring stapsgewijs uit te bouwen.

Ten tweede is het monitoren van beleid *nuttig*. In deze handleiding werd het nut van monitoring in het kader van evaluaties beklemtoond. Monitoring vormt een belangrijke databron voor beleidsevaluaties. Monitoring en evaluatie zijn complementaire ondersteunende functies in het kader van een 'evidence-based policy' of geïnformeerd en onderbouwd beleid. Monitoring en evaluatie hebben een meerwaarde vanuit verschillende motieven en bijgevolg in verschillende fasen van de beleids- en beheerscyclus.

Van belang is dat monitoring één van de bronnen voor beleidsinformatie vormt, naast andere. Beleidsmakers baseren zich op verschillende soorten informatie om beslissingen te nemen doorheen de beleidscyclus. In dat perspectief dient men soms eerder te spreken van invloed dan wel van (instrumenteel) gebruik van monitoring.

Men dient wel rekening te houden met een aantal randvoorwaarden. Zo moet men steeds de kost voor dataverzameling afwegen tegen de meerwaarde die de beleidsinformatie biedt. Ook de indicatorenset moet aan de kwaliteitsstandaarden voldoen op het vlak van betrouwbaarheid, validiteit en legitimiteit. Eens men een indicatorenset heeft, zal men deze periodiek moeten herbekijken en zondig aanpassen; indicatoren worden geschrapt, vervangen, gewijzigd en toegevoegd. Daardoor blijft het meetsysteem relevant voor de gebruikers.

Ten slotte kan men stellen dat monitoring van beleid noodzakelijk is. Dit heeft te maken met een aantal belangrijke mechanismen die de drijfveer vormen voor heel wat moderniseringstrends en hervormingen binnen de overheid en dit ongeacht bestuurslagen of beleidssectoren:

- het opdrijven van interne druk binnen overheidsorganisaties;
- het opdrijven van externe druk op overheidsorganisaties;
- meer specialisatie en decentralisatie van organisaties;
- meer specialisatie en verzelfstandiging van het beleid, met mechanismen van coördinatie.

De creatie van interne en externe druk gebeurt in en door instituties, via instrumenten en procedures die afdwingbaar zijn of afgesproken worden. Dergelijke druk –volgens de algemene veronderstelling - draagt bij tot een meer doeltreffend en effectief beleid, een efficiënter beheer en kwaliteitsvolle dienstverlening.

Organisatorische opdeling leidt tot een betere stuurbaarheid en wendbaarheid van organisaties, met een afgelijnd en homogeen takenpakket. Beleidsmatige verzelfstandiging gaat uit van het idee dat ook de beleidscyclus en beleidsprocessen kunnen opgedeeld worden in deeltaken en segmenten die door verschillende types van overheidsorganisaties verzorgd kunnen worden. Specialisatie en fragmentering zijn echter geen onvoorwaardelijke hefbomen voor een efficiënt en effectief beleid. Even belangrijk zijn dan mechanismen van coördinatie zodat het samenspel van organisaties gegarandeerd kan worden in functie van de algemene beleidsdoelstellingen.

Bij elk van de vermelde mechanismen is de rol van beleidsinformatie erg belangrijk. Monitoring en evaluatie zijn dan de ondersteunende functies bij uitstek die de noodzakelijke beleidsinformatie kunnen genereren, als het kompas voor beleidsverantwoordelijken.

Bronnen

- Bouckaert G. & Auwers T. (1999). *Prestaties meten in de overheid*, Overheidsmanagement nr. 5, Brugge: Die Keure.
- Bouckaert G. en Balk W.L. (1988). *The Task Routineness study of Schenectady County Operations*. Unpublished report.
- Bouckaert G., De Peuter B. en Van Dooren W. (2003). *Metten en vergelijken van lokale bestuurlijke ontwikkeling: een monitoringsysteem voor het lokaal bestuur in Vlaanderen*. Overheidsmanagement nr. 16, Brugge: Die Keure.
- Bouckaert G., Van Dooren W. en Sterck M. (2003). *Prestaties meten in de Vlaamse overheid: Een verkennende studie*. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen.
- Bouckaert G., Van Reeth W., Auwers T., & Verhoest K. (1997). *Handboek Doelmatigheidsanalyse - prestaties begroten*. Brussel: Ministerie van de Vlaamse Gemeenschap.
- 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: Assess and Improve Program Planning, Implementation, and Effectiveness*, Thousand Oaks, CA: Sage.
- Davidson E.J. (2004). *Evaluation Methodology Basics*. Thousand Oaks: Sage.
- De Peuter B. & Bouckaert G. (2004), *Een monitoringsysteem voor lokale bestuurskracht: naar de modellering van een Vlaams design*. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen.
- Deming W.E. (1986). *Out of the crisis*. Cambridge, MA: MIT Center for Advanced Engineering Study.
- European Commission (1999). *Evaluating socio-economic programmes, MEANS collection*, Luxembourg: Office for official publications of the European Communities.
- Garvin, D. (1993). "Building a learning organization". *Harvard Business Review*, 71(4), 78-92.
- Government Accountability Office, United States (2005). *Performance Measurement and Evaluation. Definitions and Relationships*. S.I.
- Hatry P.H. (1999). *Performance management, getting results*, Urban Institute.

- Kusek J.Z. and Rist R.C. (2004). *A Handbook for Development Practitioners. Ten Steps to a Results-Based Monitoring and Evaluation System*. Washington, D.C.: The World Bank.
- Mahtison S. (ed.) (2005). *Encyclopedia of Evaluation*. Thousand Oaks: Sage.
- McDavid J.C. & Hawthorn L.R.L. (2005). *Program Evaluation and Performance Measurement, an Introduction to Practice*, Thousand Oaks, CA: Sage.
- Mintzberg H. (1994). *The rise and fall of strategic planning*. New York: Prentice Hall.
- Mintzberg H. (1996). *Structure et dynamique des organisations*. Paris: Ed. d'organisation .
- OECD (1997). *In search of results: performance management practices*. Paris: PUMA.
- Perrin B. (1998). "Effective use and misuse of performance measurement", in: *American Journal of Evaluation*, 19(3), 367-369.
- Perrin B. (1999). "Performance measurement: Does the reality match the rhetoric? A rejoinder to Bernstein and Winston" in: *American Journal of Evaluation*, 20(1), 101-111.
- Pollitt, C., & Bouckaert, G. (2004). *Public Management Reform: A Comparative Analysis – second edition*. Oxford, UK: Oxford University Press.
- Rossi P. & Freeman H.E. (2004). *Evaluation: A Systematic Approach*, Seventh Edition, Newbury Park: Sage.
- Senge P. (1999). *Het nieuwe werk van de leider : lerende organisaties bouwen*. Schoonhoven : Academic service.
- Shadish W.R., Cook T.D. and Campbell D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston (Mass.) : Houghton Mifflin.
- Van Dooren W. & Bouckaert G. (2004). *Prestatiemeting in de Vlaamse overheid: vragen voor een informatiebeleid*. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen.
- Wholey J.S., Hatry H. and Newcomer K.E. (1994). *Handbook of Practical Program Evaluation*, San Francisco, CA: Jossey-Bass Publishers.
- Wholey J.S., Hatry H. and Newcomer K.E. (2004). *Handbook of Practical Program Evaluation*, San Francisco, CA: Jossey-Bass Publishers.
- Wholey, J.S. (2003). "Improving Performance and Accountability" in: Donaldson S.I. and M. Scriven (eds.), *Evaluating Social Programs and Problems*. Mahwah, NJ: Erlbaum, pp. 43-61.