



Ministerie van Economische Zaken
en Klimaat

Durf te leren, ga door met meten

Op zoek naar kaders en methoden voor de evaluatie
van systeem- en transitiebeleid

Oktober 2022



Woord vooraf

In 2012 verscheen het eindrapport van de Expertwerkgroep Effectmeting onder leiding van Jules Theeuwes. De commissie bracht Nederlandse evaluatie-experts bij elkaar en adviseerde het ministerie van Economische Zaken en Klimaat over de vraag op welke wijze en met welke evaluatiemethoden de instrumenten van het Nederlandse innovatiebeleid het best konden worden geëvalueerd. De commissie maakte een overzicht van de beschikbare evaluatiemethoden en ging in gesprek met de dossierhouders van beleidsinstrumenten om zich te informeren over onder meer de doelstelling, doelgroep, dataverzameling en werking van beleidsinstrumenten. Op basis daarvan werd een advies gemaakt voor een evaluatieaanpak.

De essentie van hetgeen later als de Theeuwes-aanpak ingeburgerd is geraakt, is dat in beleidsevaluaties gecontroleerd wordt voor selectiviteit en exogene invloeden door econometrische methoden toe te passen waarbij gebruik wordt gemaakt van controlegroepen. Uiteindelijk gaat het er bij beleidsevaluaties om empirisch inzichtelijk te maken wat de omvang en aard van de effecten zijn van een beleidsinterventie en of die er ook zouden zijn geweest zonder overheidsbemoeienis en de inzet van belastinggeld. De Theeuwes-aanpak maakt het mogelijk om dit inzichtelijk te maken.

Tien jaar lang zijn de aanbevelingen van de Expertwerkgroep Effectmeting de leidraad geweest bij de evaluaties die bij het ministerie van Economische Zaken en Klimaat zijn uitgevoerd in het kader van het bedrijvenbeleid. De Theeuwes-aanpak heeft de evaluatiepraktijk aanzienlijk vooruit geholpen, maar de aanpak is niet voor alle beleidsinterventies toepasbaar. Vooral waar het gaat om systeem- en transitiebeleid, gericht op bijvoorbeeld verduurzaming en digitalisering, is de aanpak niet altijd even goed toepasbaar. De redenen zijn onder meer dat er niet altijd sprake is van één instrument met een helder omschreven doel dat kan worden geëvalueerd, de beschikbaarheid van data over input, output en outcome beperkt zijn en conclusies over doeltreffendheid en doelmatigheid op de wat kortere termijn vaak moeilijk te trekken zijn. Er is daarom behoefte aan andere en nieuwe evaluatiemethoden, die naast de Theeuwes-aanpak kunnen worden ingezet voor de evaluatie van systeem- en transitiebeleid.

Het ministerie van Economische Zaken en Klimaat heeft eind 2020 een nieuwe expertcommissie ingesteld om een visie te ontwikkelen op de evaluatiepraktijk en te adviseren over methodeontwikkeling voor systeem- en

transitiebeleidsevaluaties. Dit rapport biedt een overzicht van de wetenschappelijk beschikbare evaluatiemethoden. Het betreft zowel kwantitatieve als kwalitatieve methoden (of een combinatie daarvan) waarmee de effectiviteit van systeem- en transitiebeleid inzichtelijk kan worden gemaakt. Extra aandacht is besteed aan de rol van monitoring bij evaluatie omdat het vaak om langdurige beleidsinspanningen gaat, aan de legitimiteit van overheidsingrijpen en aan de hardheid en validiteit van de evaluatiemethoden. Ten slotte is een aantal concrete casussen besproken met dossierhouders om te adviseren over een evaluatieaanpak.

De commissie bestaat uit de volgende leden: Rob Aalbers (CPB), Koen Frenken (Universiteit Utrecht), Matthijs Janssen (Universiteit Utrecht), Peter van der Knaap (ministerie van Buitenlandse Zaken), Carl Koopmans (Vrije Universiteit), Valéry Pattyn (Universiteit Leiden), Gusta Renes (PBL), Reinhilde Veugelers (KU Leuven) en Bas ter Weel (Universiteit van Amsterdam). Tot maart 2021 was Bart Verspagen (Universiteit Maastricht) voorzitter van de commissie, vanaf maart 2021 is deze rol overgenomen door Bas ter Weel. Het eindrapport is geschreven door Bas ter Weel (SEO Economisch Onderzoek), Matthijs Janssen (Dialogic), Michiel Bijlsma (SEO Economisch Onderzoek) en Pieter Jan de Boer (Dialogic).

De departementale deskundigen zijn: Theo Roelandt (EZK), Henry van der Wiel (EZK), Kim Hermans (EZK), Geert Thijssen (RvO), Wouter Panneman (EZK), Henk Massink (LNV), Marcel Seip (RvO), Alexander Buitenhuis (LNV) en Thomas Niaounakis (Fin). Het secretariaat is gevoerd door Robert Schaap (RvO).

Dit rapport is een eerste aanzet voor het evalueren van systeem- en transitiebeleid. We staan met de evaluatie van dit type beleid aan het begin van een onderzoeksagenda die duidelijkheid moet verschaffen over welke kaders, methoden en technieken adequaat de doeltreffendheid en doelmatigheid van beleid kunnen vaststellen en welke databehoeftes daarbij komt kijken. Met dank aan de leden van de commissie, dossierhouders, departementale deskundigen, de secretaris en de auteurs van dit rapport is een begin gemaakt met een nieuwe set van methoden en technieken die kunnen helpen bij het vaststellen van de doeltreffendheid en doelmatigheid van beleid.

Bas ter Weel

Voorzitter commissie Evaluatiemethoden Systeem- en Transitiebeleid

Managementsamenvatting

Met het systeem- en transitiebeleid (S/T-beleid) probeert de overheid innovatiesystemen te verbeteren en transities richting een duurzame en digitale economie in Nederland te realiseren. S/T-beleid is de verzamelnaam voor integrale beleidsprogramma's die met coördinatieactiviteiten beogen afstemming te bereiken tussen verschillende beleidsinstrumenten om zo innovatiesystemen te verbeteren en transities richting een duurzame en digitale economie in Nederland te realiseren. Dit beleid wordt gelegitimeerd door de aanwezigheid van verschillende vormen van falen, zoals externe effecten en informatie- en coördinatiegebreken. Het is daarbij belangrijk dat de kosten van overheidsingrijpen niet groter zijn dan de kosten van het falen. Hoewel voor S/T-beleid zelf vaak een beperkt budget beschikbaar is, grijpt het beleid aan op een veelomvattende set aan beleidsmaatregelen waarmee een omvangrijk budget gemoeid is. Ook heeft het beleid een groot bereik, in de zin dat veel economische actoren en processen in aanraking komen met het beleid. Het is daarom van belang de doeltreffendheid en doelmatigheid van het S/T-beleid zo goed mogelijk te meten.

Opdracht

In 2012 verscheen het eindrapport van de Expertwerkgroep Effectmeting (commissie Theeuwes). De werkgroep adviseerde het ministerie van Economische Zaken en Klimaat (EZK) over de vraag op welke wijze en met welke evaluatiemethoden de instrumenten van het Nederlandse innovatiebeleid het best konden worden geëvalueerd. De essentie van het advies is om een aantal econometrische methoden toe te passen in evaluaties die om een effectmeting vragen. In de praktijk blijkt echter dat niet voor alle types van beleidsinterventies econometrische methoden zijn toe te passen. Vooral waar het gaat om S/T-beleid ontbreken eenduidige en gevalideerde methoden en geeft de wetenschappelijke literatuur geen duidelijke evaluatiekaders. Om de doeltreffendheid en doelmatigheid van dergelijk beleid vast te kunnen stellen moet er daarom, als aanvulling op de methoden die de Expertgroep Effectmeting heeft voorgesteld, naar andere en nieuwe evaluatiemethoden worden gezocht.

De opdracht van de commissie Evaluatiemethoden Systeem- en Transitiebeleid is de volgende: Ontwikkel een visie op de evaluatiepraktijk en -strategie van het ministerie van EZK. De commissie richt zich daarbij op advisering over methodeontwikkeling voor het evalueren van S/T-beleid. Het advies biedt een overzicht van de wetenschappelijke literatuur voor uiteenlopende beleidsvraagstukken in het innovatiebeleid van het ministerie van EZK. Het betreft hier zowel kwantitatieve als kwalitatieve

onderzoeksmethoden waarmee de effectiviteit van systeem- en transitiebeleid inzichtelijk kan worden gemaakt. Daarbij wordt ook aandacht besteed aan de rol van monitoring bij evaluatie, aan de legitimiteit van overheidsingrijpen en aan de bewijskracht van de evaluatiemethodiek van systeem- en transitiebeleid.

Dit rapport verhelpt het gebrek aan evaluatiemethoden slechts voor een deel, omdat we aan het begin staan van een onderzoeksagenda die duidelijkheid moet verschaffen over welke kaders, methoden en technieken adequaat de doeltreffendheid en doelmatigheid van S/T-beleid kunnen vaststellen en welke databehoeftes daarbij komt kijken. Voor het evalueren van S/T-beleid ontbreekt een internationale evaluatiestandaard en staat de wetenschappelijke methodeontwikkeling in de kinderschoenen. Daar waar voor instrumentevaluaties een uitgebreide literatuur met methoden en technieken beschikbaar is, ontbreekt deze literatuur voor S/T-beleid. Daarnaast zijn er nauwelijks data om beleid te evalueren, ontbreekt een *counterfactual* en is het beleid dynamisch van aard. De analyse in dit rapport zet daarom uiteen welke evaluatie-overwegingen en -methoden opportuun zijn bij het evalueren van S/T-beleid.

Systeem- en transitiebeleid

S/T-beleid is te definiëren als het beïnvloeden van de productie en consumptie en het verspreiden van publieke goederen en diensten, waarbij deze goederen en diensten het resultaat zijn van een interactief proces tussen verschillende actoren, zoals consumenten, bedrijven, universiteiten en overheid. Dit betreft beleid dat met een samenhangende set van beleidsinstrumenten een verbetering wil aanbrengen in het geheel van productiefactoren om zo een bepaald maatschappelijk doel te bereiken. Er is sprake van systeembeleid als het beleid zich richt op kenmerken van het systeem als geheel (bijvoorbeeld de samenhang tussen regelingen) en niet op geïsoleerde deelaspecten van een systeem (bijvoorbeeld het optimaliseren van een enkele innovatieregeling). Er is sprake van transitiebeleid als het aanpassen van factoren ervoor moet zorgen dat een systeem transformeert (bijvoorbeeld als nieuwe doelen moeten worden behaald of een andere werking van het systeem nodig is). Zowel bij systeembeleid als bij transitiebeleid is veelal sprake van een experimenteel en lerend proces, waarin het pad van probleem naar oplossing niet op voorhand helder is en soms zelfs de oplossing niet vooraf helder kan worden gedefinieerd. Dit betekent dat het beleid adaptief moet worden vormgegeven, zowel bij het invoeren als afschaffen ervan, omdat de reeks van prikkels en wet- en regelgeving die verstrekt of afgeschaft

wordt dient te worden aangepast op het moment dat het systeem nieuwe knelpunten kent.

Evaluatiekader

Omdat het ontbreekt aan een internationale wetenschappelijke evaluatiestandaard, stelt dit rapport eerst een evaluatiekader vast op basis van een synthese van de beschikbare wetenschappelijke literatuur en de bestaande evaluatiepraktijk. Dit kader bestaat uit zes invalshoeken: interventielogica, governanceprocessen en beleidsmix, match tussen beleids(mix) en knelpunten, systeemversterking/-transformatie, structurele verandering en maatschappelijke impact. Voor iedere specifieke invalshoek is het mogelijk om concrete hypothesen op te stellen die aangeven welke verandering er verwacht mag worden. Iedere invalshoek uit het

evaluatiekader belicht een ander aspect van de werking en doeltreffendheid van S/T-beleid. Daarbij lenen de invalshoeken over systeem- en structuurverandering en maatschappelijke impact zich voor summatieve analyses met het oog op verantwoording over bewerkstelligde veranderingen, terwijl de nadruk bij de invalshoeken over interventie-rationale, governanceprocessen en de beleidsmix en de aansluiting met knelpunten ligt op het formatief onderzoeken van de onderbouwing en uitvoering van het beleid.

Om de invalshoeken uit het evaluatiekader in de praktijk toe te passen is het zaak ze te operationaliseren door een beleidstheorie op te stellen. Hierbij biedt onderstaand kader een aantal voorbeelden van te beantwoorden vragen.

Invalshoek	Voorbeelden van vragen
Interventielogica	Wat zijn de vormen van falen die vragen om S/T-beleid (dus met focus op coördinatie en stroomlijnen)? Waar precies zitten die knelpunten, en wat is het bewijs daarvoor? Bijvoorbeeld: zijn er zonder het beleid onvoldoende (beleids)mogelijkheden om bij te dragen aan systeemverandering en/of transities, en wat verklaart waarom de markt en samenwerkende actoren er zelf niet uitkomen?
Governance en beleidsmix	Zijn er principes die aangeven waar de coördinatie vanuit het S/T-beleid op dient toe te zien, en voldoet de coördinatie aan die principes? Leidt de coördinatie tot passende aanpassingen in de beleidsmix (dat zijn wijzigingen waardoor systeemveranderingen en daaruit voortkomende activiteiten beter aansluiten bij beleidsdoelen)? In welke mate is dit terug te zien in het ontwikkelen of uitfasen van regelingen en in aanpassingen in de financiering en vormgeving van regelingen?
Match tussen beleids(mix) en knelpunten	Wat zijn de knelpunten (in termen van systeemprocessen of structuren) die systeemverandering ingewikkeld maken of transitieprocessen stroever laten verlopen, en hoe wordt hierop ingespeeld door S/T-programma's en de beleidsinstrumenten waar dat beleid op aangrijpt? Gaat de meeste energie naar de grootste bottlenecks, en worden er resultaten geboekt op het versterken van de zwakste systeemaspecten?
Systeemversterking en -transformatie	Zien we verbeteringen in de systeemprocessen die ontwikkeling en verspreiding van innovaties mogelijk maken? Is er meer kennisontwikkeling/-uitwisseling, marktformatie etc. (afhankelijk van welk proces zwak was) rondom wenselijke innovaties?
Structuurverandering	Veranderen innovatiesystemen / socio-technische / productie-consumptiesystemen structureel? Het gaat hierbij om omvorming van hele systemen, ook de delen die buiten de nauwe scope van de aan het S/T-beleid verbonden regelingen vallen. Een eerste indicatie zou zijn of er daar meer innovatie-activiteit ontstaat.
Maatschappelijke impact	Worden de tussentijdse en uiteindelijke doelen gehaald? En in welke mate is dit het gevolg van (systeem) veranderingen die toe te schrijven zijn aan de invloed van het S/T-beleid? Bijvoorbeeld: als er sectoren zijn waar CO ₂ -besparing is gerealiseerd, berust dat dan ook op innovaties waar het beleid aantoonbaar een impuls aan heeft gegeven?

Onderzoeksmethoden

Op het moment dat het evaluatiekader helpt bepalen welke aspecten van beleid onderzocht dienen te worden, is de volgende vraag welke onderzoeksmethoden het best passen. Onderstaand schema geeft een overzicht van methoden en technieken die passen bij het evalueren van de doeltreffendheid en doelmatigheid van de verschillende invalshoeken van S/T-beleid.

	Interventie-logica	Governance en beleidsmix	Match tussen beleids(mix) en knelpunten	Systeem-versterking en -transformatie	Structuur-verandering	Maatschappelijke impact
Doeltreffendheid						
Systematische reviews & meta-analyse	Systematische reviews betreft kaders en principes identificeren om beleidsuitvoering aan te toetsen. Een meta-analyse is een cijfermatige analyse.					
Lerend evalueren	Mechanismen verklaren en (samen) interpreteren, vertrekkend vanuit beleidstheorie en -uitvoering					
Case studies				Mechanismen verklaren & uitkomsten bepalen		
Outcome harvesting				Mechanismen verklaren, vertrekkend vanuit uitkomsten (hoe komt die uitkomst tot stand; wat was de rol van beleid?)		
Contribution analysis & process tracing	Causaliteit bepalen, vertrekkend vanuit de beleidstheorie (indicaties verzamelen voor optreden van een keten van uitkomsten)					
Qualitative Comparative Analysis (QCA)				Causaliteit bepalen, vertrekkend vanuit uitkomsten (welke combinatie van factoren bepaalt variatie in succes?)		
Regressieanalyse				Effectomvang meten, corrigerend voor controlevariabelen		
(Quasi-)experimentele methoden				Causaliteit bepalen & Effectomvang meten		
Doelmatigheid						
Monitoring	Ontwikkelingen bijhouden, op vooraf bepaalde indicatoren					
Productiviteits-onderzoek (DEA, SFA, benchmark)				Vergelijken van kosten en/of prestaties (of de verhouding daartussen)		
Simulaties*					Uitkomsten schatten van complexe (beleids)interacties	
MKBA						MKBA

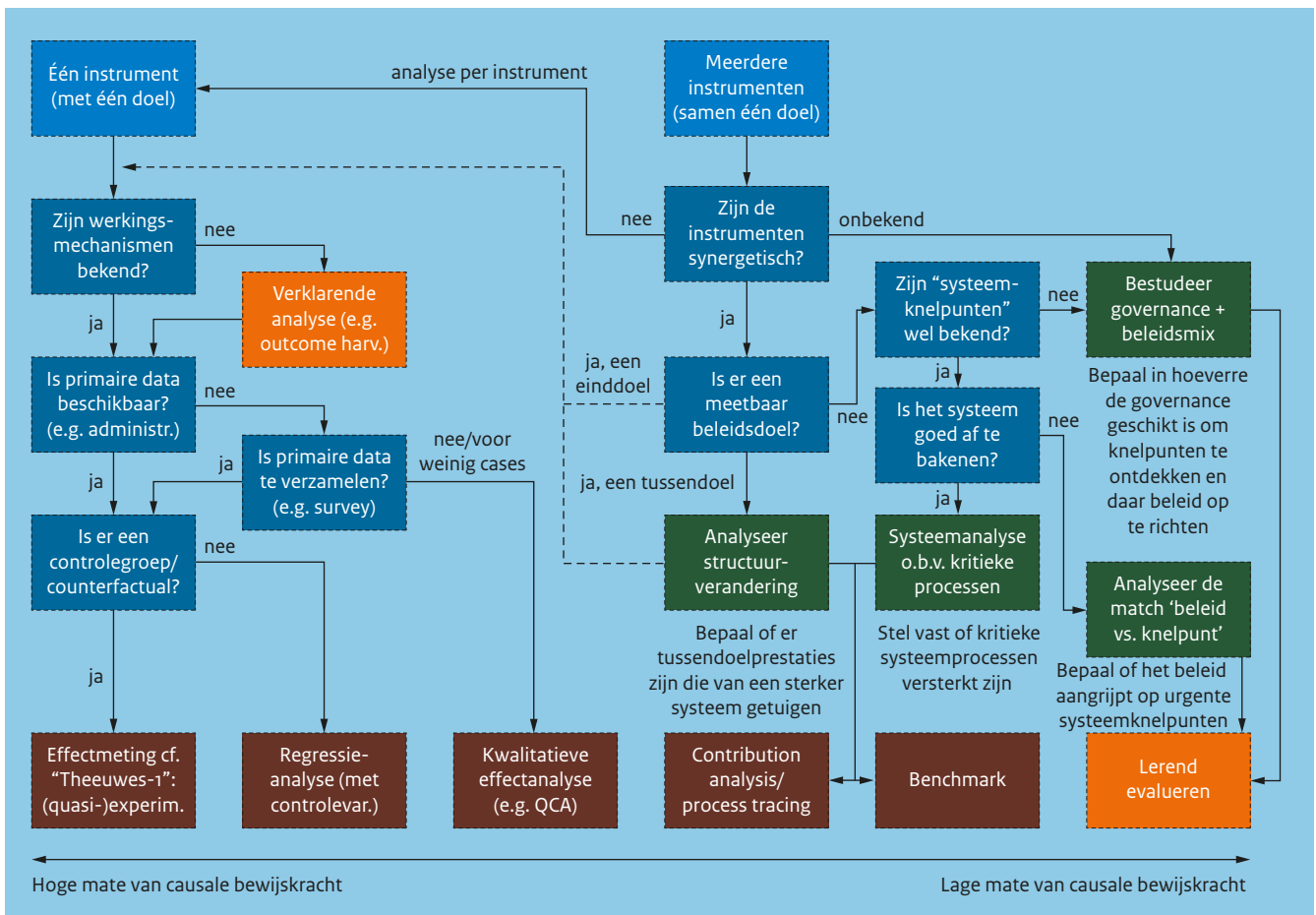
* Simulaties kunnen ook zicht geven op doeltreffendheid, maar zijn hier voor de overzichtelijkheid maar één keer opgenomen

Beslিশulp

Om evaluatoren op weg te helpen bij het selecteren van methoden is er een beslিশulp in de vorm van een stroomschema opgesteld. Door dit schema van boven naar beneden te doorlopen worden evaluatoren gewezen op de empirische methoden die het meest geijkt lijken voor een te evalueren casus. Links in het schema staan de methoden waar de causale bewijskracht het hoogst is, met de aanname dat daar ook een voorkeur voor is bij effectmetingen. In het rapport van de Expertwerkgroep Effectmeting is reeds beschreven welke concrete technieken overwogen kunnen worden bij het inzetten van (quasi-)experimentele methoden. Hoewel de as niet volstrekt ordinaal is, neemt de causale bewijskracht af voor methoden naarmate die meer rechts in de figuur staan. De te overwegen methoden zijn getoond in de donkeroranje blokjes. De meest rechtse methode, lerend evalueren, leent zich niet voor uitspraken over causale effecten. Dit methodeblokje heeft daarom dezelfde lichtere kleur als het blokje ‘verklarende analyse’. De groene blokjes verwijzen naar de analysekaders die in beeld komen afhankelijk van de eigenschappen van het te evalueren beleid, en de informatie die daarover beschikbaar is. Ze zijn gekoppeld aan de invalshoeken uit het evaluatiekader.

Vier voorbeelden

De bruikbaarheid van de in dit rapport besproken kaders en methoden en de beslিশulp worden toegepast op vier casussen die verschillen in de mate en manier waarop ze voorbeelden zijn van S/T-beleid. De variatie op dit punt maakt het mogelijk om te kunnen verkennen wat de kansen en beperkingen zijn van de aangereikte handvatten. De vier geselecteerde casussen zijn: Het Missiegedreven Topsectoren- en Innovatiebeleid, het CO₂-reductiebeleid/klimaatbeleid, de Nederlandse Digitaliseringsstrategie en het Techniepact.



Inhoud

Woord vooraf	3
Managementsamenvatting	5
1. Aanleiding	11
1.1. Evaluatiepraktijk	11
1.2. Behoefte	11
1.3. Opdracht	12
2. Startpunt van de commissie	13
2.1. Ontbreken van een internationale evaluatiestandaard	13
2.2. Bereik van het advies	15
3. Systeem- en transitiebeleid	17
3.1. Wat is systeem- en transitiebeleid?	17
3.2. Marktfalen, systeemfalen en transitiefalen: mogelijke nieuwe redenen voor beleid	18
3.3. Doeltreffendheid van enkelvoudige instrumenten en S/T-beleidsinstrumenten	21
3.4. Doelmatigheid van enkelvoudige instrumenten en S/T-beleidsinstrumenten	22
3.5. Uitdagingen voor evaluatie van S/T-beleid	24
4. Evaluatiekader systeem- en transitiebeleidsprogramma's	27
4.1. Algemene uitgangspunten	27
4.2. Beleidstheorie	29
4.3. Twee evaluatieperspectieven	29
4.4. Theoretische afwegings- en analysekaders voor S/T-beleid	31
4.5. Selectie en toepassing van de evaluatiekader-invalshoeken	36
5. Methoden voor het evalueren van systeem- en transitiebeleid	41
5.1. Overzicht van methoden	41
5.2. Inzetbaarheid van afzonderlijke methoden	43
6. Stappenplan voor het voorbereiden en ontwerpen van evaluatiestudies	49
7. Beslishulp voor het selecteren van methoden	53
7.1. Beslishulp voor beleid met één instrument en één doel	54
7.2. Beslishulp voor beleid met meerdere instrumenten en één doel	54
8. Eerste lessen op basis van verkennende casussen	57
8.1. Toepassing beslishulp per casus	57
8.2. Lessen over het afbakenen van (en aanmerken als) S/T-beleid	70
8.3. Lessen over onderzoeksmethoden	71
9. Epiloog	73
Bronmateriaal	75

1. Aanleiding

In 2012 verscheen het eindrapport van de Expertwerkgroep Effectmeting (commissie Theeuwes). De werkgroep adviseerde het ministerie van Economische Zaken en Klimaat (EZK) over de vraag op welke wijze en met welke evaluatiemethoden de instrumenten van het Nederlandse innovatiebeleid het best konden worden geëvalueerd. De essentie van het advies is om econometrische methoden toe te passen in evaluaties die om effectmeting vragen. Bij dergelijke beleidsevaluaties is het zaak om empirisch inzichtelijk te maken wat de omvang en de aard van de effecten van een beleidsinterventie zijn, en of die er ook zouden zijn geweest zonder overheidsbemoeienis en de inzet van belastinggeld. Het afgelopen decennium zijn de aanbevelingen van deze werkgroep de leidraad geweest bij evaluaties en heeft het advies de evaluatiepraktijk bij het ministerie van EZK vooruit (en ook daarbuiten) geholpen.

In de praktijk blijkt echter dat niet voor alle types van beleidsinterventies econometrische methoden zijn toe te passen. Vooral waar het gaat om systeem- en transitiebeleid ontbreken eenduidige methoden. Er moet daarom naar andere en nieuwe evaluatiemethoden worden gezocht om de doeltreffendheid en doelmatigheid van dergelijk beleid vast te kunnen stellen. Deze methoden vormen aanvullingen op de methoden die worden voorgesteld door de Expertwerkgroep Effectmeting.

Er is door het ministerie van EZK een commissie ingesteld om advies uit te brengen over evaluatiemethoden ten behoeve van systeem- en transitiebeleid. Deze commissie is breed samengesteld met experts vanuit het wetenschappelijk onderzoek en departementale deskundigen (zie Bijlage A). De commissie heeft haar werk verricht in de periode november 2000 – januari 2022. Dit rapport is daar het resultaat van.

1.1. Evaluatiepraktijk

Het uitgangspunt is dat voor instrumentevaluaties (met één interventie, één gebruikerstype en één helder omschreven en meetbaar doel) een internationale evaluatiestandaard en kwantitatieve methoden voorhanden zijn. Deze evaluatiestandaard en kwantitatieve methoden dienen te worden toegepast volgens het advies van de Expertwerkgroep Effectmeting. Voor systeemevaluaties en het beoordelen van transitiebeleid ontbreekt een evaluatiestandaard en staat de methodeontwikkeling in de kinderschoenen. Daar waar mogelijk dienen beleidsinstrumenten die deel uitmaken van een systeem of van een transitie als instrument te worden geëvalueerd volgens het advies van de Expertwerkgroep Effectmeting, daar waar het gaat om het systeem of de transitie zelf zijn andere methoden nodig.

De kwaliteit van evaluaties hangt in zeer sterke mate af van de beschikbaarheid en kwaliteit van microdatabestanden met informatie over de respons van bedrijven en individuen op het beleid. Er is noodzaak om de data-infrastructuur in Nederland verder te verbeteren, onder meer door voorafgaand aan de invoering van nieuw beleid al te investeren in deze data-infrastructuur. Dit moet zorgen voor een tijdige en adequate dataverzameling over *input*, *output*, *outcome* en *impact* van beleid dat zoveel mogelijk aansluit op en is te koppelen aan bestaande microdatabestanden. Voor beleidsinstrumenten zijn hier in de afgelopen tien jaar stappen gezet die moeten worden gecontinueerd en waar nodig geïntensiveerd, voor het evalueren van systeem- en transitiebeleid bestaat geen noemenswaardige data-infrastructuur. Deze dient te worden opgebouwd.

Bij de meeste beleidsevaluaties die in het afgelopen decennium zijn uitgevoerd, zijn slechts conclusies over de directe effectiviteit van beleidsinstrumenten getrokken. De maatschappelijke impact van beleid (inclusief externe effecten) en de doelmatigheid van beleid kan slechts zelden worden beoordeeld in een beleidsevaluatie. Dit leidt ertoe dat in veel beleidsevaluaties geen duidelijke en scherpe conclusies kunnen worden getrokken over de doeltreffendheid en doelmatigheid van beleid, ondanks dat daar bij de opdrachtgever (en in de maatschappij) wel behoefte aan is. Voor een deel hangt dat samen met gebrekkige data, voor een deel met de timing van de evaluatie die vaak te vroeg komt en voor een deel met het ontbreken van een solide instrumentarium voor deze ‘hogere orde’ effecten.

Het evaluatiestelsel is erop gericht om beleidsinstrumenten af te rekenen op al dan niet geleverde prestaties. Er bestaat in Nederland geen cultuur waarbij beleidsevaluaties worden benut om te leren in de beleidsontwikkeling. Openheid en transparantie over wat goed gaat en wat beter kan, de mogelijkheid van tussentijds bijsturen en de wens om te evalueren komen daarmee onder druk te staan. Wanneer het gaat om systeem- en transitiebeleid, waar bijsturen en experimenteren van groot belang zijn, is dat een zorgelijke ontwikkeling.

1.2. Behoeftes

Er is bij het ministerie van EZK behoefte aan het vergroten van de gereedschapskist van de evaluatiepraktijk. In het bijzonder is er behoefte aan inzicht in methoden die voor systeem- en transitiebeleid een standaard zetten. De redenen voor deze behoefte zijn de volgende.

Transitiebeleid, zoals het missiegedreven innovatiebeleid of beleid gericht op verduurzaming van de energie-intensieve industrie, verandert het beleidsperspectief door doelstellingen te formuleren van een te realiseren doel op de middellange termijn. Op korte termijn is het slechts mogelijk de directe output van een beleidsinterventie op weg naar de realisatie van het doel te meten. Daarnaast is de realisatie van het doel wel te formuleren in termen van een meetbare doelstelling, maar het beleid op weg naar de realisatie van het doel is met grote onzekerheid omgeven. De reden is dat het meest succesvolle en optimale beleid op weg naar de realisatie van een doel niet op voorhand vaststaat. Dat leidt ertoe dat het bijvoorbeeld wel vaststaat dat nieuwe sleuteltechnologieën de wereld ingrijpend zullen veranderen, en dat ze behulpzaam zullen zijn bij het realiseren van een maatschappelijke opgave, maar dat de mogelijke toepassingen van die technologieën nog niet bekend zijn, laat staan al zijn gerealiseerd in de praktijk. In veel gevallen moet de technologie worden doorontwikkeld en bijgestuurd om tot een concrete en bruikbare toepassing te komen.

Voor de realisatie van een doel zijn verschillende actoren met elkaar verbonden in een systeem en beïnvloeden beleidsinstrumenten elkaar. Er is sprake van een grote mate van afhankelijkheid tussen de actoren en beleidsinstrumenten. Beleidsinstrumenten kunnen op de korte termijn alleen worden beoordeeld op de mate waarin directe additionele output wordt gerealiseerd en wat dat bijdraagt op de weg naar de realisatie van

het doel. Het monitoren van de directe additionele output en het ontwikkelen van meetbare indicatoren die iets zeggen over de richting van de geobserveerde ontwikkeling zijn belangrijk voor de tussentijdse monitoring en bijsturing van het beleid. Ex-ante evaluatie lijkt hiervoor minder geschikt, omdat het systeem al bestaat en in functie is. Ex post evaluatie is vaak ook onmogelijk omdat in het systeem niet altijd oorzaak en gevolg helder zijn. Monitoring lijkt hier wel van toepassing. De vraag is dus hoe de doeltreffendheid en doelmatigheid van systeembeleid is vast te stellen.

1.3. Opdracht

De opdracht die aan de commissie is gegeven is de volgende. Ontwikkel een visie op de evaluatiepraktijk en -strategie van het ministerie van EZK. De commissie richt zich daarbij op advisering over methodeontwikkeling voor het evalueren van systeem- en transitiebeleid. Het advies biedt een overzicht van de wetenschappelijke literatuur voor uiteenlopende beleidsvraagstukken in het innovatiebeleid van het ministerie van EZK. Het betreft hier zowel kwantitatieve als kwalitatieve onderzoeksmethoden waarmee de effectiviteit van systeem- en transitiebeleid inzichtelijk kan worden gemaakt. Daarbij wordt gevraagd ook aandacht te besteden aan de rol van monitoring bij evaluatie, aan de legitimiteit van overheidsingrijpen (marktfalen, systeemfalen en transitiefalen) en aan de plaats van de evaluatiemethodiek van systeem- en transitiebeleid op de effectenladder.

2. Startpunt van de commissie

In het domein van innovatiebeleid voltrokken zich het afgelopen decennium twee prominente trends. Ten eerste is er vanuit de voorkeur voor *evidence-based policy* toenemende belangstelling voor gedegen econometrische effectmetingen, volgens standaarden die hoog op de Maryland-effectladder staan.¹ Deze voorkeur is te bedienen met een kwantitatieve evaluatie van beleidsinstrumenten, zoals uitgewerkt door de Expertwerkgroep Effectmeting. Voor het evalueren van afzonderlijke instrumenten (met één interventie, één gebruikerstype en één helder omschreven en meetbaar doel) bestaat een uitgekristalliseerde wetenschappelijke literatuur, is er een erkende internationale evaluatiestandaard, zijn kwantitatieve methoden voorhanden en zijn er vaak ook data en een data-infrastructuur beschikbaar. Ten tweede is er, voornamelijk in Europa, de opkomst van systeem- en transitiebeleid (S/T-beleid). Dit betreft beleid dat middels een samenhangende set van beleidsinstrumenten een verbetering wil aanbrengen in het geheel van productiefactoren (het 'systeem') om zo een bepaald maatschappelijk doel te bereiken.² Ten slotte is er toenemende mate sprake van transitiebeleid. Bij dit beleid, dat dikwijls veranderingen van systemen beoogt, ligt het maatschappelijke doel doorgaans ver in de toekomst en is het niet helder hoe het meest effectieve en efficiënte pad daarnaartoe eruitziet. Ook kunnen de beleidsdoelen (veelal tussendoelen die gesteld worden om het uiteindelijke maatschappelijke doel te bereiken) en aanpak in de loop van de tijd wijzigen omdat nieuwe informatie beschikbaar komt.

2.1. Ontbreken van een internationale evaluatiestandaard

Voor het evalueren van S/T-beleid ontbreekt een internationale evaluatiestandaard en staat de wetenschappelijke methodeontwikkeling in de kinderschoenen. Dit rapport richt zich op het ontwikkelen van richtlijnen voor het evalueren van systeem- en transitiebeleid. Deze beleidsevaluaties zijn getoond in de twee middelste kolommen in onderstaand overzicht (Tabel 2.1). Voor de volledigheid is in dit overzicht ook een kolom opgenomen voor het evalueren van instellingen die publieke belangen borgen op het terrein van innovatie, zoals de TO2-instellingen.³ Naast het periodiek evalueren van het beleid zelf, is het toevoegen van

nieuwe methoden aan de evaluatiepraktijk ook nuttig voor het verbeteren van beleidsdoorlichtingen. Door de evaluatiepraktijk uit te breiden zullen er waarschijnlijk meer delen van begrotingsartikelen geëvalueerd kunnen worden dan nu het geval is.

Behalve verschillen in de objecten van S/T-beleidsevaluatie, kunnen er ook verschillen bestaan in de evaluatiecriteria die daarbij gehanteerd worden. Dit rapport richt zich op de vraag hoe de criteria doeltreffendheid en doelmatigheid te evalueren zijn als sprake is van S/T-beleid. Hierbij komen ook deelaspecten van doeltreffendheid en doelmatigheid zoals coherentie, consistentie en uitvoering van beleid aan bod.

Eén van de aspecten die bepalend is voor het evalueren van doeltreffendheid en doelmatigheid is het soort gegevens dat gebruikt kan worden voor evaluatie. De kwaliteit van evaluaties hangt in sterke mate af van de beschikbaarheid van data die *input*, *output*, *outcome* en *impact* van beleid en beleidsinstrumenten meten. Uit een evaluatie van het Bedrijvenbeleid door het ministerie van EZK (2020), een inventarisatie van beleidsdoorlichtingen in de periode vanaf 2011 (Koopmans et al., 2019) en de Toolbox beleidsevaluaties uit 2020 blijkt dat er behoefte is aan het verder verbeteren van de data-infrastructuur, onder meer door voorafgaand aan invoering of in een vroeg stadium van nieuw beleid te investeren in de bijbehorende dataverzameling en data-ontwikkeling voor monitoring en evaluatie. Deze data-infrastructuur dient aan te sluiten bij de bestaande infrastructuur van microdatabestanden die beschikbaar is bij het CBS. Gebrek aan bruikbare data is één van de redenen dat slechts een deel van het beleid op een wetenschappelijk verantwoorde manier op effectiviteit kan worden beoordeeld (doeltreffendheid) en een nog kleiner deel op efficiëntie (doelmatigheid).

Een andere beperking is de beschikbaarheid van methoden en technieken om beleid te evalueren. De methoden en technieken die beschikbaar zijn in de wetenschappelijke literatuur leiden ertoe dat op de meeste terreinen slechts eerste-orde-uitkomsten van beleidsmaatregelen op het terrein van doelbereik en doeltreffendheid kunnen worden vastgesteld. De uiteindelijke

¹ Onderzoekers die zich buigen over *evidence-based* economisch en innovatiebeleid kijken nog maar beperkt naar de vraag *hoe* het precieze ontwerp van een instrument zich verhoudt tot de gegenereerde impact, al wordt erkend dat dit in de praktijk net zo belangrijk kan zijn als weten of beleid impact heeft (Hünermund & Czarnitzki, 2019; Duflo, 2017).

² Het beleid dat zich richt op dit samenhangende pakket, bijvoorbeeld door te proberen daar meer samenhang in te brengen, is zelf ook weer te zien als beleidsinstrument.

³ De TO2-instellingen zijn vijf organisaties voor toegepast onderzoek: Deltares, MARIN, het Nederlands Lucht- en Ruimtevaartcentrum, TNO en Wageningen Research.

Tabel 2.1. Evaluatieaanpak bij beleidsvraagstukken in het S/T-beleid

	Instrumentevaluaties	Systeembeleid-evaluaties	Transitiebeleid-evaluaties	Evaluatie publieke instituten
Aangrijpingspunt	Actoren	Werking & functies innovatiesysteem	Systeem- en structuurverandering	Instellingen en randvoorwaarden
Overheidsrol	Marktfalen	Systeemfalen	Transitiefalen	Wettelijke taken
Primaire doel	Gedragsverandering	Verbeterde werking innovatiesysteem	Realisatie missies en maatschappelijke doelen	Publieke taak uitvoeren
Primaire interventievorm	Financiële stimulans	Interactie en samenhang binnen het systeem	Maatschappelijke doel- en normstelling	Wetgeving en (basis) subsidies
Evaluatieaanpak	Natuurlijke en sociale experimenten	Werking van innovatiesysteem vaststellen	Monitoring van richting en status van ontwikkelpaden op weg naar realisatie	Doelrealisatie en benchmarking
Kenmerk	Gericht op beïnvloeding gedrag individuele actor	Gericht op systemen en clusters	Gericht op systemen en clusters	Gericht op organisaties en wetgeving
Voorbeelden	WBSO, Innovatiekrediet	Topsectorenbeleid	Missiegedreven innovatiebeleid	TO2 instellingen, universiteiten, ZBO instellingen
Richtlijnen	Expertwerkgroep Effectmeting (2012)		Dit rapport	Standaard Evaluatie Protocol

outcome en impact blijven vaak buiten beeld.⁴ Daarnaast zijn voor het meten van de doelmatigheid van het beleid nieuwe data noodzakelijk en lijken ook nieuwe methoden/modellen te moeten worden ontwikkeld. Het kunnen beschikken over betrouwbare schattingen van tweede- en derde-orde-effecten, waaronder positieve en negatieve externaliteiten van beleidsinstrumenten, is een noodzakelijke voorwaarde om een beter begrip van de doeltreffendheid en doelmatigheid te krijgen en daarover conclusies te kunnen trekken.

Bovenstaande lacunes zijn van toepassing bij de evaluatie van uiteenlopende soorten beleid. De aanleiding voor dit rapport is dat S/T-beleid specifieke eigenschappen kent die evaluatie nog eens extra lastig maken. Er bestaan vaak nog meer data-lacunes, het ontbreekt aan *counterfactuals* en er zijn vooralsnog geen eenduidige manieren voor het evalueren van dit dynamische beleid.

- *Data.* S/T-beleid richt zich op het realiseren van effectieve en efficiënte processen, regelgeving en coördinatie om een systeem van actoren en processen (bijvoorbeeld het innovatiesysteem of het energiesysteem) te verbeteren of op een transitie naar een nieuw en soms onbekend doel (bijvoorbeeld verdere digitalisering van de economie), waardoor de beschikbaarheid van data over *input*, *output*, *outcome* en *impact* vaak gering is. Er is behoefte aan gegevens die deze processen en de mate van

coördinatie monitoren, als basis voor analyses die uitwijzen in welke mate een transitie gerealiseerd is of in hoeverre het pad naar een transitiedoel effectief en efficiënt wordt bewandeld. Voorafgaand aan het uitvoeren van een beleidsevaluatie is het belangrijk dat de opdrachtgever erover nadenkt welke gegevens verzameld moeten en kunnen worden voor een meting van doeltreffendheid en doelmatigheid.

- *Counterfactual.* Ten tweede bestaat er geen *counterfactual* die de ontwikkeling van het systeem of de snelheid van een transitie meet in afwezigheid van S/T-beleid.⁵ Niet alleen is er veelal geen tweede vergelijkbaar systeem beschikbaar, ook is S/T-beleid vaak een (beperkt) onderdeel van een breder palet aan beleidsinstrumenten met elk afzonderlijke effecten of bestaat S/T-beleid uit een bundeling van beleidsinstrumenten. S/T-beleid is vaak niet gericht op traditionele beleidsinstrumenten, maar op coördinatie-activiteiten om bestaande beleidsinstrumenten te richten op de gewenste systeemverandering dan wel transitieprocessen. Als zodanig kan coördinatie natuurlijk ook worden gezien als instrument. Niettemin is het in dit soort gevallen lastig om geobserveerde uitkomsten toe te schrijven aan een heterogene set van directe en indirecte interventies. Het doel van S/T-beleid is om met coördinatie deze verzameling van instrumenten beter te laten werken. Het isoleren van het

⁴ Een voorbeeld is het effect van een subsidieregeling op de hoeveelheid onderzoek dat bedrijven uitvoeren. Het gevolg op de hoeveelheid R&D is dan meetbaar, maar niet de effecten op economische groei en productiviteit. In sommige gevallen kunnen macro-economische evenwichtsmodellen helpen om economische doorwerking van eerste-orde-effecten te bepalen.

⁵ Een voor-/nameting of een vergelijking met andere landen of sectoren kunnen behulpzaam zijn, maar vormen bijna nooit een *counterfactual* vanwege de veelheid aan andere factoren die tot verschillen leidt.

effect van meer coördinatie van het effect van de afzonderlijke elementen ('is de som meer dan het geheel der delen?') is uitdagend. Door het ontbreken van de *counterfactual* kunnen geen duidelijke en scherpe conclusies worden getrokken over de doeltreffendheid en doelmatigheid van het beleid. Dat neemt niet weg dat op kleinere schaal soms wel *counterfactuals* beschikbaar zijn, bijvoorbeeld als het S/T-beleid een goed afgebakende beleidsmaatregel of instrument omvat of wanneer de coördinatieactiviteit zelf kan worden uitgelicht als te evalueren instrument. In dat geval kan geëvalueerd worden zonder uitspraken te hoeven doen op het niveau van een systeem of transitie.

- *Dynamisch in plaats van statisch.* Ten derde is bij S/T-beleid vaak wel duidelijk wat het einddoel is en wat op dit moment het marktfalen is, maar is onduidelijk welke belemmeringen precies in de weg staan om het doel in de toekomst te bereiken wanneer nieuwe en nu nog onbekende oplossingen beschikbaar zijn. In dat geval wordt relevant of bestaand beleid processen en structuren kent om belemmeringen te signaleren, daarop in te spelen en het beleid aan te passen. Van belang is dus of de beleidsprocessen en -structuren zodanig ingericht zijn dat ze met de dynamische aard van systemen en transities kunnen meebewegen. Onderdeel van het S/T-beleid is niet alleen het verhelpen van knelpunten (zoals ontbrekende systeemfactoren of transitieprocessen), maar ook het identificeren en mitigeren daarvan. Die identificatie kan betrekking hebben op bijvoorbeeld de 'zwakke' functies in het systeem die moeten veranderen om het systeem te transformeren richting het realiseren van de doelstellingen. Op het niveau van afzonderlijke knelpunten kunnen heel gerichte acties ondernomen worden om ze te verhelpen. Die instrumenten zelf kunnen uit relatief bekende en beproefde interventies bestaan, of uit beleidsaanpakken waarbij gaandeweg moet blijken welke (experimentele) interventies werken en waarom dat het geval is.

2.2. Bereik van het advies

In tegenstelling tot de decennia aan onderzoek over het evalueren van (economisch) beleid dat gebaseerd is op één knelpunt en één instrument, is er bij S/T-beleid geen uitgebreide evaluatietraditie die kant-en-klare handvatten aanreikt voor het evalueren van dergelijk beleid. Er bestaat wetenschappelijke literatuur over het beleidsmatig versterken of transformeren van innovatiesystemen, productie- en consumptiesystemen en socio-technische systemen. Die literatuur beschrijft deze verschillende systemen, maar stelt niet de vraag in hoeverre systeemveranderingen door beleidsinterventies tot stand gebracht zijn (of hoe dat überhaupt te bepalen is) en heeft geen traditie in het evalueren van S/T-beleid.⁶ Ook richt deze literatuur zich vooral op het beschrijven van het

hele systeem en minder op de kantelpunten van een transitie. Er bestaat literatuur over de evaluatie van complexe beleidsinstrumenten vanuit het inzicht dat "the policy evaluation landscape is shifting; government is evaluating more varied policies, often made up of multiple projects, delivered in collaboration with many stakeholders, at multiple governance levels – central, regional and local." (Barbrook-Johnson et al., 2021). Deze literatuur richt zich vaak op het doorgronden van beleid (door daar beleidsmakers zelf in te betrekken) en niet op het effect of de efficiëntie daarvan.

Dit rapport verhelpt het gebrek aan evaluatiehandvatten slechts voor een deel, omdat we aan het begin staan van een onderzoeksagenda die duidelijkheid moet verschaffen over welke kaders, methoden en technieken adequaat de doeltreffendheid en doelmatigheid van S/T-beleid kunnen vaststellen en welke databehoeftes daarbij komt kijken. De analyse zet daarom uiteen welke evaluatie-overwegingen en -methoden opportuun zijn bij het evalueren van S/T-beleid. Hierbij houden we vast aan het adagium dat iedere beleidsevaluatie maatwerk is (Van der Knaap et al., 2020, p. 5).

Om tot een overzicht te komen wordt eerst gedefinieerd wat systeem- en transitiebeleid is en wat de uitdagingen zijn als het gaat om het evalueren van dat beleid (Hoofdstuk 3). Hoofdstuk 4 presenteert perspectieven en kaders die beschrijven hoe en waarop systeem- en transitiebeleid geëvalueerd kan worden. Tevens wordt er stilgestaan bij het operationaliseren van deze kaders, in termen van vragen en indicatoren. Hoofdstuk 5 geeft een antwoord op de vraag welke empirische methoden te gebruiken zijn voor het onderzoeken van die vragen en indicatoren. Hoofdstuk 6 vat de voorgaande hoofdstukken samen in een stappenplan dat concrete handvatten probeert te bieden voor het voorbereiden en ontwerpen van de evaluatie van S/T-beleid. Hierbij ligt de nadruk op het systematisch in kaart brengen van wat er bekend is over wat het beleid beoogt, hoe het hier een impuls aan geeft en welke analyses er verricht kunnen worden. Hoofdstuk 7 biedt vervolgens een beslissing voor het selecteren van de empirische methode(n) waarmee het feitelijke evaluatieonderzoek uitgevoerd kan worden. Dit wordt in Hoofdstuk 8 geconcretiseerd middels reflectie op een aantal concrete casussen relevant voor het beleidsterrein van het ministerie van EZK. Hoofdstuk 9 sluit af met een epiloog, die de in dit rapport gepresenteerde perspectieven, kaders en adviezen koppelt aan de nog te nemen stappen in het ontwikkelen en toepassen van een evaluatieaanpak voor S/T-beleid.

⁶ Een verklaring voor de beperkte aandacht voor evaluatie van S/T-beleid is dat dit soms voortkomt uit perspectieven op overheidsbeleid die afwijken van de traditionele rechtmatige en presterende overheid, maar juist gebaseerd zijn op overheidsrollen als een netwerkende en responsieve overheid die participeert in een 'verwoven dynamiek' (ministerie van Economische Zaken, 2017). Die alternatieve rollen gaan gepaard met andersoortige evaluatievragen, waarbij minder aandacht uitgaat naar verantwoordten en juist meer aandacht naar belichten hoe beleid ten uitvoering wordt gebracht, welke rol diverse partijen daarin innemen, en hoe dit verbeterd kan worden. In dit rapport houden we vast aan een zgn. verantwoordingsperspectief op evaluatie.

3. Systeem- en transitiebeleid

Dit hoofdstuk definieert wat S/T-beleid is in de context van het ministerie van EZK en hoe het werkt. Het kadert daarmee af welk type beleid valt binnen de scope van dit rapport en hoe het beleid in elkaar zit (paragraaf 3.1), welke typen falen relevant zijn en beleidsinterventie rechtvaardigen (3.2), hoe het bepalen van doeltreffendheid (3.3) en doelmatigheid (3.4) verschilt ten opzichte van instrumentevaluaties, en welke uitdagingen er daardoor bestaan voor het evalueren van S/T-beleid (3.5).

3.1. Wat is systeem- en transitiebeleid?

Systeem- en transitiebeleid gaat over het aanpassen van structuren, zoals actornetwerken of productie-consumptie-systemen, met het doel dat die structuren maatschappelijk wenselijke uitkomsten genereren. Het betreft het organiseren van mesoniveau-veranderingen die niet tot stand komen via het beïnvloeden van (microniveau-)gedrag van één soort actor. Uiteraard kunnen op microniveau instrumenten worden ingezet om een maatschappelijk gewenste gedragsverandering bij consumenten en producenten te stimuleren.

In economische termen is S/T-beleid te zien als het beïnvloeden van de productie, consumptie en verspreiding van publieke goederen en diensten, waarbij deze goederen en diensten het resultaat zijn van een interactief proces tussen verschillende actoren, zoals consumenten, bedrijven, universiteiten en de overheid. Dit resultaat komt tot stand aan de hand van een experimenteel en lerend proces, waarin het pad van probleem naar oplossing niet op voorhand helder is en soms zelfs de oplossing niet vooraf helder kan worden gedefinieerd.

Het denken in systemen is van waarde omdat het totale proces van de productie, verspreiding en toepassing van publieke goederen en diensten wordt beschouwd, evenals de impact van dit proces op de economische en sociale ontwikkelingen/transities (Lundvall, 1992). Met deze benadering kunnen alle activiteiten binnen een economisch systeem, evenals de organisaties die bij deze processen betrokken zijn, worden gezien als onderdelen van een publiek doel. Het systeem bevat feedbackloops en meerdere relaties tussen de organisaties waaruit het is samengesteld. Elk element van het systeem is dus belangrijk voor het proces als geheel en beïnvloedt en wordt beïnvloed door de andere elementen.

Er zijn verschillende manieren waarop de systeembenadering beleid kan ondersteunen en sturen. Ten eerste verschuift het de focus van het beleid van individuele instrumenten naar een

set van instrumenten en de interacties daartussen. Bij sommige interacties kunnen verschillende instrumenten worden ingezet die zich bezighouden met de productie van publieke goederen en diensten. De tweede manier waarop de systeembenadering nuttig is, is dat de aandacht wordt gevestigd op zowel wetenschappelijke en technologische input als op de processen die van belang zijn voor het realiseren van publieke goederen en diensten. Een derde manier waarop de systeembenadering waardevol is, is dat gedrag van economische actoren zoals consumenten, bedrijven, maar ook publieke organisaties wordt beïnvloed door een scala aan instrumenten. Deze instrumenten creëren prikkels voor het toepassen van de resultaten van wetenschappelijk onderzoek en het inzetten van experimenten. Daarnaast zijn informele regels, normen, gebruiken en routines belangrijk, zoals organisaties die zich bezighouden met certificeringen of octrooien. Ten vierde benadrukt de systeembenadering dat beleid op verschillende niveaus van toepassing is, zoals op supranationaal (klimaatbeleid), nationaal (topsectorenbeleid), regionaal (techniekpact), sectoraal en technologisch (digitalisering) niveau. Ten slotte verandert de focus van de analyse van de interne werking van een economisch systeem naar de manier waarop het systeem interacteert met de buitenwereld (bijvoorbeeld klimaatbeleid).

Wanneer we het innovatiedomein nemen, kunnen we drie soorten beleid onderscheiden waar evaluaties zich op kunnen richten: individuele regelingen zoals de WBSO (waar het rapport van de Expertwerkgroep Effectmeting zich op richt), individuele instituten zoals de TO2-instituten (te evalueren o.b.v. van bijvoorbeeld het Standaard Evaluatie Protocol) en programma's. Programma's zijn overkoepelende strategieën waarin meerdere beleidsprikkel in samenhang gericht worden op een bepaald doel. Die samenhang is essentieel bij doelen die niet met een (enkelvoudige) directe impuls bereikt kunnen worden, maar waarvoor het een voorwaarde is dat er eerst (via meervoudige impulsen) een systeem of transitieproces ingericht wordt dat ondersteunend is aan het op termijn te behalen einddoel. S/T-beleid valt in deze laatste categorie (zie ook Tabel 2.1).

Kenmerkend voor S/T-beleid is dus dat het gaat om integrale programma's die middels coördinatie-activiteiten beogen afstemming te bereiken tussen een reeks van (vaak ongelijksoortige) beleidsprikkel en -instrumenten. Wat het beleid uiteindelijk beoogt te formuleren is een te realiseren beleidsdoel of -missie op middellange termijn, maar de weg ernaartoe en de adequate mix van beleidsprikkel en -instrumenten is onzeker. Het is bijvoorbeeld onduidelijk wat relevante nieuwe

doorbraaktechnologieën zijn, hoe coördinatie (inclusief regelgeving en prikkels) het best kan worden vormgegeven, of welk beleid en welke prikkels afgeschaft moeten worden. Daarom is een proces van leren en testen noodzakelijk om tot een concrete en bruikbare invulling van het uiteindelijke beleid te komen. Dat proces gaat gepaard met experimenten en pilots om uit te vinden wat werkt en wat niet werkt, bijsturen van prikkels en regels, stoppen met wat voortijdig mislukt of niet efficiënt is etc. Ten slotte is het voor de realisatie van doelen en transities van belang dat actoren met elkaar verbonden zijn in netwerken.

Er is sprake van *systeembeleid* als het beleid zich richt op kenmerken van het systeem als geheel (bijvoorbeeld de samenhang tussen regelingen) en niet op geïsoleerde deelaspecten van een systeem (bijvoorbeeld het optimaliseren van een enkele regeling). Een energiebelasting is bijvoorbeeld geen systeembeleid, omdat het probeert via uniforme prijsprikkels het gedrag van individuele actoren te beïnvloeden. Wel kan het instrument onderdeel zijn van een systeem dat zich met verschillende instrumenten richt op het beprijzen van externe effecten van het gebruik van fossiele energie. Er is sprake van *transitiebeleid* als het aanpassen van factoren ervoor kan zorgen dat systemen transformeren (omdat ze nieuwe doelen moeten behalen of wanneer een andere werking van het systeem nodig is). Dit betekent dat het beleid adaptief moet worden vormgegeven, omdat de reeks van prikkels die versterkt wordt dient te worden aangepast op het moment dat het systeem nieuwe knelpunten kent of wanneer nieuwe informatie beschikbaar komt.

3.2. Marktfaalen, systeemfaalen en transitiefalen: mogelijke nieuwe redenen voor beleid

In de wetenschappelijke literatuur is marktfaalen één van de redenen voor overheidsingrijpen, die moet worden afgewogen tegen vormen van overheidsfaalen als verdringing, kosten van beleid (zoals vergunningen en belastingen), beïnvloeding van beleid door lobbygroepen of de kans dat de overheid beleid (onverwacht) wijzigt. Verschillende beleidsinstrumenten proberen marktfaalen te mitigeren. Net zoals markten kunnen faalen, is het mogelijk dat sprake is van systeemfaalen en transitiefalen. Als bijvoorbeeld wordt gekeken naar innovatiebeleid, zijn, naast het direct aanjagen van R&D-uitgaven, het versterken van innovatiesystemen en het ontwikkelen van oplossingen voor maatschappelijke uitdagingen doelen van innovatiebeleid (Kuhlmann & Rip, 2018). Ook bij klimaatbeleid is de afgelopen tijd sprake van een transitieaanpak, en valt een instrument als het beprijzen van de uitstoot van schadelijke stoffen binnen een set van instrumenten en maatregelen die als doel hebben de gestelde klimaatdoelen te bereiken. Deze opkomst van nieuwe beleidsdoelen gaat gepaard met een discussie over de vormen van faalen die overheidsingrijpen legitimeren en hoe we tegen deze vormen van faalen aankijken. Box 3.1 geeft deze discussie weer aan de hand van de innovatiecontext.

Box 3.1. Verschillende soorten faalen in de context van innovatie

Economisch gezien is het marktfaalen in innovatieprocessen het gevolg van imperfecties als coördinatieproblemen als gevolg van transactiekosten (bedrijven kunnen bruikbare innovaties of klanten niet identificeren), informatie-asymmetrie (financiers en kapitaalzoekers hebben niet dezelfde informatie waarop ze hun transactie moeten baseren), marktmacht (bestaande bedrijven kunnen vernieuwende nieuwkomers tegenhouden), of kennisspillovers. Dit laatste betreft de voordelen die innoverende bedrijven creëren voor hun concurrenten zonder dat zij hiervoor gecompenseerd worden. Innoverende bedrijven genereren daarmee positieve externaliteiten, aangezien hun eigen investeringen ook ten goede kunnen komen aan andere bedrijven. Daarnaast is innovatie risicovol, waardoor het zeker voor kleinere bedrijven lastig is om innovatieprojecten extern te financieren in een situatie van informatieasymmetrie met kapitaalverschaffers. Tegen deze achtergrond vormen generieke innovatiesubsidieprogramma's al sinds jaar en dag de kern van het innovatiebeleid in Nederland en andere OESO-landen.

Systeemfaalen kan begrepen worden als het gevolg van coördinatieproblemen, namelijk problemen die ontstaan bij complementaire goederen (Rodrik, 2008). Dit probleem is bekend als een coördinatiespel waarbij investeringen hogere rendementen (positieve externaliteiten) genereren als alle andere relevante spelers de juiste complementaire investering doen. Dergelijk coördinatiefaalen kan in diverse fases van het innovatieproces voorkomen. In de vroege fase, wanneer voor innovatie noodzakelijke nieuwe kennis gegenereerd wordt, doet systeemfaalen zich voor omdat kennis collectief gegenereerd wordt door meerdere partijen: bedrijven, universiteiten en semipublieke instellingen. Wanneer voor dit soort samenwerking en kennisdeling de juiste omstandigheden en prikkels ontbreken, is sprake van systeemfaalen. Hier speelt ook de cumulatieve natuur van het innovatieproces, waarbij toekomstige innovaties voortbouwen op huidige innovaties. Patenten slagen er niet in om deze 'intertemporele coördinatie' tot stand te brengen (Scotchmer, 1991).

In een latere fase van het innovatieproces hangt het effectief gebruik van innovaties onder meer af van de juiste investeringen in fysieke of digitale infrastructuur en in nieuwe kennis en vaardigheden om de innovaties te benutten. Omgekeerd zullen dergelijke infrastructurele investeringen alleen gedaan worden als men redelijkerwijs kan verwachten dat er ook voldoende gebruik van wordt gemaakt. Daarnaast dienen (radicale) innovaties ontwikkeld te worden in samenhang met wetten, regels, bestemmingsplannen en maatschappelijke opvattingen, bijvoorbeeld door het ontwikkelen van technische standaarden, gebruikersprotocollen en investeringen in menselijk kapitaal. Er is sprake van systeemfaalen als dergelijke coördinatieprocessen spaak lopen.

Van transitiefalen is sprake als een overgang naar een maatschappelijk wenselijk nieuw evenwicht niet of te langzaam plaatsvindt. Achterliggende oorzaak is vaak een coördinatiefalen waardoor transitiekosten (bijvoorbeeld investeringen) voor individuele partijen te hoog zijn of overheidsfalen, als de overheid niet scherp genoeg ingrijpt om partijen in beweging te zetten. Net als in de traditionele economische kijk op innovatiebeleid kan transitiebeleid gezien worden als het voortbrengen van spillovers, zij het dat het ook gaat om andere spillovers (e.g. 'coördinatie-externaliteiten' en 'adoptie-externaliteiten') dan alleen kennisspillovers (Janssen, 2022).

In transitiebeleid worden de beleidsdoelen afgeleid van maatschappelijke uitdagingen waarvoor de oplossingen meestal nieuwe publieke goederen (een publiek goed is in deze context een dienst van de overheid, maar kan ook een systeem van regels of prikkels zijn dat ervoor zorgt dat externe effecten worden geïnternaliseerd) en economische prikkels of structuren vereisen, terwijl bij systeembeleid vaak de economische concurrentiekracht in het algemeen centraal staat. De in het transitiebeleid nagestreefde verbetering van specifieke publieke goederen is politiek gearticuleerd als bijvoorbeeld een schoner milieu, het tegengaan van klimaatveranderingen, de volksgezondheid, cyberveiligheid of effectieve rampenbestrijding. Dit laat onverlet dat oplossingen voor maatschappelijke uitdagingen ook kunnen leiden tot nieuwe productinnovaties en verdienmodellen.

Daarnaast is bij transitiebeleid het beleidsdoel slechts haalbaar in de vorm van een langdurig proces dat gekenmerkt wordt door onzekerheid over de gewenste oplossingsrichtingen, die de ernst van coördinatieproblemen vergroot, zowel tussen beleidsmakers als tussen de actoren in het transitieproces in het algemeen. Dit vergt een langetermijnbeleidsperspectief en heeft gevolgen voor de manier waarop het beleid op de kortere termijn geëvalueerd kan worden. De noodzaak tot coördinatie begint vroeg, omdat belanghebbenden zich vooraf moeten committeren aan concrete maatschappelijke doelstellingen. Dit proces is niet triviaal omdat bij transitieprocessen gevestigde belangen in het gedrang kunnen komen. Transities behelzen immers niet alleen de opbouw van nieuwe systemen, maar ook de afbraak van bestaande systemen (Loorbach et al. 2017).

In de wetenschappelijke literatuur wordt onderscheid gemaakt tussen marktfaalen, systeemfaalen en transitiefalen. Weber & Rohracher (2012) introduceren deze classificatie in de context van innovatiebeleid, voortbouwend op eerdere literatuur. Het marktperspectief analyseert faalen vanuit het startpunt van een efficiënte markt. Het systeemperspectief analyseert faalen vanuit het oogpunt van een goed functionerend systeem van bedrijven, kennisinstellingen, overheden en andere partijen,

waarbij de werking van instituties, routines en gedrag moeten passen bij de gewenste werking en resultaten. Het transitieperspectief analyseert faalen vanuit de overgang naar een nieuw maatschappelijk wenselijk evenwicht. Hierbij is het van belang dat instituties, routines en gedrag worden vernieuwd, zich aanpassen of verdwijnen om het nieuwe evenwicht te bereiken.

De drie typen van faalen worden doorgaans ieder als afzonderlijk faalen gepresenteerd, wat vooral gelegen is in het feit dat er verschillende wetenschappelijke tradities aan ten grondslag liggen (Weber & Rohracher, 2012).⁷ Dat leidt ertoe dat in de praktijk systeem- en transitiefalen vaak te begrijpen en duiden zijn aan de hand van onderliggend marktfaalen dat we kennen uit de bestaande evaluatiepraktijk bij het ministerie van EZK. In die zin verschillen de drie typen faalen niet fundamenteel van elkaar. Op een aantal punten overlappen de vormen van faalen (coördinatieproblemen komen impliciet of expliciet overal terug), terwijl de kaders ook complementair zijn: wanneer er sprake is van systeem- of transitiefalen is het vaak zinvol om de vraag te stellen welke activiteiten de markt en bestaand beleid al oppakken, versus wat er blijft liggen (omdat er bijvoorbeeld geen markt voor is).

Marktfaalen

Voor de evaluatie van enkelvoudige (vaak financiële) instrumenten bestaat een internationale standaard en zijn kwantitatieve methoden voorhanden. Ingrijpen van de overheid wordt vooral gelegitimeerd door marktfaalen. Het primaire beleidsdoel is om de onder-investeringen in publieke goederen en diensten aan te pakken. De voornaamste vormen van marktfaalen zijn:

- *Marktmacht*. Bedrijven die samenspannen om de concurrentie te beperken belemmeren toetreding van nieuwe bedrijven. Andersom leidt te veel fragmentatie ook tot te weinig vernieuwing doordat bedrijven te klein zijn om te investeren in nieuwe goederen en diensten. Hierdoor is het niveau van vernieuwing van goederen en diensten te laag. Bronnen van marktmacht zijn schaalvoordelen en *lock-in* effecten;
- *Informatieasymmetrie* (averechtse selectie). Onzekerheid over het succes van investeringen en de tijdschikhorizon waarbinnen rendement kan worden verwacht leidt tot terughoudendheid van private (of publieke) partijen om te investeren. Hierdoor is er sprake van suboptimale maatschappelijke investeringen in risicovolle projecten zoals onderzoek en ontwikkeling;
- *Freeridgedrag*. Onderzoek en ontwikkeling en het genereren van nieuwe ideeën hebben kenmerken van een publiek goed. Het is namelijk niet-rivaal waardoor de private baten voor de investeerder lager zijn dan de publieke baten van het verspreiden van kennis. Hierdoor is de maatschappelijke investering in

⁷ Overheidsfaalen speelt ook een belangrijke rol bij de vormgeving van beleid. Er kan sprake zijn van *capture* van beleid door lobby's van belangengroepen of als gevolg van een informatieachterstand waardoor verkeerde beleidskeuzes worden gemaakt. We bespreken overheidsfaalen hier niet als aparte categorie omdat het geen rechtvaardiging is voor overheidsingrijpen, maar een factor die vooral de vormgeving van dat ingrijpen beïnvloedt, denk aan eisen aan governancestructuren en transparantie van besluitvorming die onafhankelijkheid versterken, of eisen aan ex ante onderbouwing van beleidskeuzes.

onderzoek en ontwikkeling suboptimaal door *freeridergedrag* als gevolg van *kennisspillovers*;

- *Externaliteiten*. Innovatie-investeringen van een bedrijf hebben maatschappelijke baten voor andere bedrijven en de rest van de samenleving, zonder dat dit bedrijf voor die externe effecten wordt vergoed. Het innoverende bedrijf houdt geen rekening met deze positieve externe effecten. Vanuit een maatschappelijk oogpunt zal het bedrijf daarom te weinig innoveren met als gevolg te weinig *rentspillovers* voor consumenten en *kennisspillovers* voor bedrijven en samenleving;
- *Transactiekosten*. Sommige markten kunnen ontbreken waardoor de financiering niet tot stand komt of zoekkosten tussen partijen die financiering aanbieden en nodig hebben kunnen zo hoog zijn dat transacties die maatschappelijk rendabel zijn niet tot stand komen. Transactiekosten kunnen leiden tot coördinatiefalen doordat marktpartijen complementaire activiteiten niet kunnen coördineren. Ten slotte is het mogelijk dat transactiekosten zo hoog zijn dat er geen maatschappelijk rendement is te behalen met overheidsingrijpen.

Systeemfalen

Voor systeemevaluaties ontbreekt een internationale standaard en is de methodeontwikkeling minder ver gevorderd. Ook is er veel minder aandacht voor causaliteit in de zin dat er meestal geen kwantitatieve methoden zijn om de impact van beleid te meten. De wetenschappelijke literatuur is meer beschrijvend van aard en legt vooral de werking van systemen bloot. Systemen kunnen last hebben van beperkingen, die kunnen worden gedefinieerd als systeemfalen als gevolg van een onderliggend marktfalen. Ingrijpen van de overheid kan daarom worden gelegitimeerd door systeemfalen. Achterliggend aan systeemfalen bestaat dus vaak marktfalen of overheidsfalen (Woolthuis et al., 20015; Wiczorek & Hekkert, 2012). De voornaamste vormen van systeemfalen zijn:

- *Gebrek aan infrastructuur*. Dit refereert aan infrastructuur die onvoldoende ontwikkeld is door de grote omvang van de benodigde investering, de lange terugverdientijd ervan en het lage private rendement op de investering. Denk bijvoorbeeld aan kennis- of energie-infrastructuur. Dit systeemfalen kan ook het gevolg zijn van *informatieasymmetrie* en *externaliteiten*;
- *Capaciteitsfalen*. Dit kan spelen als er onvoldoende capaciteit is waardoor nieuwe ontwikkelingen of activiteiten niet van de grond komen en de systeemactoren zich onvoldoende aanpassen aan nieuwe omstandigheden. Capaciteitsfalen ontstaat bijvoorbeeld als er onvoldoende toegang is tot kennis als gevolg van een gebrek aan of mismatch van kennis en vaardigheden, maar kan ook refereren aan fysieke infrastructuur die ontbreekt. *Transactiekosten* (zoals zoekkosten) en *informatieasymmetrie* (onduidelijk welke opbrengsten investeringen in kennis opleveren) zijn hier vaak een oorzaak van;

- *Netwerkfalen*. Netwerkfalen kan optreden als netwerkeffecten en netwerkexternaliteiten belangrijk zijn. Netwerkeffecten kunnen leiden tot *marktmacht* en het bewegen van het netwerk in de verkeerde richting ten opzichte van maatschappelijk doelen als gevolg van *technologische lock-in*. Bestaande netwerken belemmeren dan toetreders, of netwerken worden afhankelijk van een dominante partner. Netwerkfalen treedt ook op als netwerken zich onvoldoende ontwikkelen door te hoge transactie- of coördinatiekosten. Positieve *externaliteiten* van innovatie blijven dan onderbenut.

Veel van het innovatiebeleid in Nederland kan beschouwd worden als een vorm van systeembeleid. Daarbij gaat het om de instrumenten die gericht zijn op het creëren van een omgeving die bevorderlijk is voor innovatiemogelijkheden, het aanjagen van samenwerking, het benutten van positieve externaliteiten en het aanpassen aan een veranderende markt/omgeving. In de context van innovatie richt systeembeleid zich op de innovatieactiviteiten van bedrijven, universiteiten en overheid als belangrijkste actoren in economische en innovatieprocessen (Lundvall, 1992; Nelson, 1993).

Transitiefalen

Net zoals systemen kunnen falen is het ook mogelijk dat een transitie niet dan wel te laat tot stand komt of stagneert. Ingrijpen van de overheid wordt gelegitimeerd doordat de gewenste transities niet of niet snel genoeg via de markt tot stand komen, of dat de markt (veelal gedreven door overheidsbeleid) vanuit maatschappelijk oogpunt de verkeerde kant op beweegt, en doordat bestaand overheidsbeleid veranderingen tegenhoudt. Vormen van transitiefalen zijn (Weber & Rohrer, 2012):

- *Een ontbrekende richting van ontwikkeling*. Bij transities gaat het er niet alleen om het falen van de markt of het systeem aan te pakken, maar ook met beleid de richting van de verandering en de tijdigheid te sturen. Dat vraagt om het coherent bewegen richting dat doel door alle actoren, ofwel een gedeelde visie op de transitie. Symptomen van het ontbreken van richting zijn bijvoorbeeld een gebrek aan een gedeelde visie over het doel en de richting van de transitie; gebrek aan coördinatie tussen de actoren of transitiepaden die geacht worden het systeem in een nieuwe richting te bewegen; gebrek aan wet- en regelgeving om de richting van de transitie te sturen en die sturing vast te houden; gebrek aan financiële middelen om onderzoek en voorbeeldprojecten mogelijk te maken; opgesloten zitten in verouderde routines, instituties en gedrag en gebrek aan infrastructuur om routes voor ontwikkelpaden vast te stellen;
- *Onduidelijkheid over de vraag*. In het systeem perspectief wordt geen rekening gehouden met het marktpotentieel in termen van nieuwe producten en diensten. Een kenmerk van transities is dat er vaak tegelijkertijd maatschappelijke, organisatorische en institutionele veranderingen nodig zijn. Tekortkomingen om de acceptatie van innovaties door gebruikers en consumenten mogelijk te maken zijn daarom een type falen

dat beleidsinterventie rechtvaardigt. Symptomen van het ontbreken van vraag zijn bijvoorbeeld onvoldoende ruimte om de gebruikersbehoeften te analyseren om de acceptatie van innovaties te vergroten door te weinig afstemming met gebruikers; een gebrek aan stimulerende signalen vanuit de publieke vraag door het ontbreken van de overheid als gebruiker en gebrek aan competenties om innovaties aan de vraag en behoeftes van (eind)gebruikers te koppelen;

- *Falen van beleidscoördinatie.* De interactie van beleidsterreinen relevant voor transities heeft baat bij coördinatie, namelijk beleidscoördinatie. Hoewel het begrip coördinatiefalen in het onderzoeks- en innovatiebeleid is gebruikt als een voorbeeld van systeemfalen, verwijst het alleen naar coördinatieproblemen van innovatieactoren, niet naar coördinatieproblemen bij beleid. Het organiseren van activiteiten op bijvoorbeeld nationaal, regionaal en sectoraal niveau en tussen verschillende partijen is van belang tijdens transities;
- *Gebrek aan reflexiviteit.* Het gaat hier om de reflectie of toetsing van beleid en daarop herziening van dat beleid. In transities is het van belang om te experimenteren en te leren van mislukkingen en voort te bouwen op successen.

Transitiebeleid richt zich op maatschappelijke doelen die via systeemverandering tot stand moeten komen. Bij transities bestaat de wens om veranderingsprocessen op te starten, zelfs als op voorhand niet duidelijk is welke interventies het best werken. Er bestaat in de wetenschappelijke literatuur geen kader om transitiebeleid te evalueren.

Het theoretisch en empirisch vaststellen van het probleem waar beleid het antwoord op vormt, oftewel het benoemen en analyseren van het marktfalen, systeemfalen of transitiefalen, het liefst gecombineerd met het opzetten van een beleidstheorie, is een essentiële stap in vrijwel alle vormen van beleidsevaluatie. Als niet duidelijk is welk probleem opgelost moet worden, en waarom de overheid met het voorziene beleid mogelijk een (positief) verschil kan maken, dan is het op voorhand de vraag of het beleid een positieve impact kan hebben – zie het rapport van de Expertwerkgroep Effectmeting (2012) voor een uitgebreidere reflectie hierop. Hoewel het bij systeem- en vooral transitiebeleid gangbaarder is dat beleid een experimenteel karakter heeft, omdat er simpelweg weinig kennis is over welk beleid zou kunnen werken (en het maatschappelijke doel dat het ingrijpen alsnog rechtvaardigt), is het ook daar van belang om op voorhand een theorie te ontwikkelen t.a.v. de werking en verwachte uitkomsten van het gekozen beleid. Bovendien zouden er op zijn minst structuren moeten zijn (bijvoorbeeld een formele leercyclus) die het mogelijk maken om gaandeweg een scherper beeld te krijgen van de problemen die in de weg staan aan het bereiken van de gewenste doelen. Ook het functioneren van die structuren kan in een evaluatie onderzocht worden.

3.3. Doeltreffendheid van enkelvoudige instrumenten en S/T-beleidsinstrumenten

Hoewel het begrip doeltreffendheid eenduidig is, kent het bepalen daarvan bij S/T-beleid enkele andere uitgangspunten dan bij de evaluatie van enkelvoudige instrumenten.

Een prominent verschil tussen S/T-beleid en enkelvoudig beleid is het type actie dat de overheid onderneemt. De acties richten zich in plaats van op eigenschappen van individuele actoren in het systeem (hoeveel investeert een bedrijf in R&D?) op eigenschappen van het systeem als geheel (hoe sluiten investeringen van verschillende bedrijven op elkaar aan?). Dit betekent dat beleidsacties niet zozeer bestaan uit instrumenten die gedrag of keuzes van individuele actoren proberen te beïnvloeden, maar uit meerdere instrumenten die de samenhang van gedrag of keuzes van groepen actoren proberen te beïnvloeden. Dit maakt het beleidsinstrumentarium voor S/T-beleid anders dan interventies die zich richten op individuele actoren. Het zwaartepunt van S/T-beleid is het mobiliseren en richten van, mogelijk reeds bestaande, instrumenten en initiatieven. Het beleid is systemisch omdat het zich richt op systeemkenmerken en niet op geïsoleerde deelaspecten. Zo worden er bijvoorbeeld netwerken gecreëerd (bijvoorbeeld het MVI North Sea Energy Lab), akkoorden gesloten (zoals het Energieakkoord), agenda's opgesteld (bijvoorbeeld de Digitaliseringsagenda), pacts gesmeed (zoals het Techniepact) en overlegtafels opgezet (bijvoorbeeld versnellings tafels voor de energietransitie). Het doel is om sturing te geven aan relevante economische en maatschappelijke ontwikkelingen door het vormen, inhoudelijk richten en stimuleren van netwerken, oftewel groepen van interacterende actoren.

Daarnaast is S/T-beleid dynamischer van karakter dan enkelvoudig beleid. Bijzonder aan S/T-beleid is dat het probeert door voortdurende coördinatie van beleidsprikkelers en beleidsinstrumenten een beleidsdoel te bereiken. Dit dynamische karakter heeft twee redenen:

- *Identificatie van problemen.* Ten eerste is de identificatie van het onderliggende probleem vaak onduidelijk (het concept systeemfalen is niet helder uitgewerkt in een economisch theoretisch kader, waardoor het in de praktijk een wat vaag begrip blijft dat te plakken is zonder dat duidelijk wordt wat nu echt het onderliggende probleem is), terwijl ook onzeker is of de gekozen oplossing (stimuleren van overleg, afstemming, samenwerking of netwerkvorming) bijdraagt aan de oplossing van het probleem. Dit leidt ertoe dat systeembeleid vaak een continue zoektocht en herdefinitie is naar een antwoord op de vraag 'wat werkt om het systeem te verbeteren en wat niet?' gecombineerd met een steeds verschuivend palet aan oplossingen (wat bijvoorbeeld kan worden bediend door te experimenteren);
- *Doel.* Ten tweede speelt, vooral bij transitiebeleid, dat het beleidsdoel en het probleem verschuiven omdat de samenleving gedurende een transitie steeds op andere, nieuwe problemen

stuit en het pad van de transitie steeds beter leert kennen wat om een veranderende aanpak vraagt. Het is moeilijk om beleidsmatig te anticiperen op verwachte zwaktes (ontbrekende condities), omdat vaak niet duidelijk is welke transitiepaden tot de gewenste uitkomst leiden. Vandaar ook de noodzaak te blijven bijsturen en experimenteren met beleid om te kijken wat wel en wat niet werkt – ook om overheidsfalen te voorkomen.

Tabel 3.1 vat de verschillen tussen het beoordelen van de doeltreffendheid van beleid gericht op enkelvoudige instrumenten en gericht op S/T-beleid samen op vijf aspecten. De *counterfactual* is bij S/T-beleid meestal niet beschikbaar of uitgewerkt, het directe beleidsdoel is gericht op het systeem of de transities, instrumenten zijn vaak meervoudig of een bundeling van enkelvoudige instrumenten, dat maakt evaluatie complex, de doelgroep is dikwijls niet goed af te bakenen en het type falen kan net anders uitpakken.

3.4. Doelmatigheid van enkelvoudige instrumenten en S/T-beleidsinstrumenten

Beleid moet ook op doelmatigheid worden beoordeeld. Zowel bij de evaluatie van het EZK-beleid in de afgelopen periode als bij S/T-beleid is dit een moeizaam proces gebleken. Er bestaat in theorie een onderscheid tussen drie vormen van doelmatigheid: technische doelmatigheid, allocatieve doelmatigheid en dynamische doelmatigheid. Technische en allocatieve doelmatigheid zijn statische vormen van doelmatigheid, want ze beoordelen de doelmatigheid van beleid op een gegeven punt in de tijd. Dynamische doelmatigheid beschrijft de ontwikkeling van doelmatigheid over de tijd. In de praktijk wordt tijdens beleidsevaluaties vaak slechts gerapporteerd over technische

doelmatigheid, terwijl maatschappelijk gezien ook allocatieve doelmatigheid (de marginale waarde van een extra euro wanneer keuzes gemaakt moeten worden tussen alternatieven) zeker zo belangrijk is. Bij transitiebeleid is dynamische doelmatigheid van belang, omdat over de tijd nieuwe informatie beschikbaar komt over de wijze waarop het beleidsdoel het best bereikt kan worden (en met welk beleid dat gepaard zou moeten gaan).

Technische doelmatigheid

Het tegengaan van verspilling is een manier om technische doelmatigheid te vergroten. Verspilling van publieke middelen treedt op als de inzet hoger is dan noodzakelijk voor een te bereiken doel. Door verspilling te voorkomen, blijft het effect van beleid gelijk maar dalen de uitgaven. Andere voorbeelden van verbeteringen in technische doelmatigheid zijn het voorkomen van fouten en het stroomlijnen van processen.

Het reduceren van de uitgaven is niet de enige manier om technische doelmatigheid te vergroten. Technische doelmatigheid kan ook verbeteren door hogere kwaliteit te leveren tegen dezelfde kosten, bijvoorbeeld doordat nieuwe technologieën worden ingezet of wanneer met dezelfde middelen meer kan worden bereikt. Beleid dat gepaard gaat met lagere kosten staat dus niet per definitie gelijk aan technisch meer doelmatig beleid (het kan immers ook ten koste gaan van kwaliteit) en een verbetering in technische doelmatigheid is niet per definitie hetzelfde als lagere kosten (een investering, bijvoorbeeld in een nieuw ICT-systeem, levert een hogere kwaliteit op).

In evaluaties draait technische doelmatigheid om de verhouding tussen de baten (effecten) en de kosten van beleid. De kosten

Tabel 3.1. Verschillen in uitgangspunten bij bepalen doeltreffendheid enkelvoudig beleid en S/T-beleid.

	Enkelvoudig beleid	S/T-beleid
Counterfactual	Constructie van <i>counterfactual</i> op basis van partijen die buiten de doelgroep liggen (behandel- en controlegroep).	In evaluaties van S/T beleid word een <i>counterfactual</i> doorgaans niet expliciet uitgewerkt. Constructie van <i>counterfactual</i> is mogelijk basis van theorie, kwalitatieve beschouwende studies uit de literatuur en experimenten. ⁹
Directe doel beleid	Beïnvloeden van eigenschappen en keuzes van individuele actoren.	Beïnvloeden van collectieve eigenschappen en gedrag van groepen van actoren (in het systeem of betrokken bij de transitie).
Type beleids-instrument	Acties die aangrijpen op gedrag en keuzes van individuele actoren. Denk aan financiële prikkels, opleiding of wettelijke eisen.	Acties die aangrijpen op verzamelingen van actoren of een verzameling van enkelvoudige beleidsinstrumenten. Vaak gericht op processen en coördinatie van actoren.
Complexiteit	Beperkt. Sturen via acties die direct aangrijpen op een individuele actor.	Omvat meerdere beleidsinstrumenten, vaak op meerdere beleidsterreinen, daarnaast gericht op processen en coördinatie.
Doelgroep	Duidelijk afgebakend.	Vaak grote, heterogene groepen actoren.
Type falen	Marktfalen. Vaak externe effecten en <i>spillovers</i> .	Systeemfalen, transitiefalen. Achterliggend vaak marktfalen en verouderde instituties.

⁹ Soms zijn vergelijkingen met andere landen mogelijk, hier spelen echter meestal veel factoren die ook op de *outcome* van invloed zijn en waarvoor niet kan worden gecontroleerd.

meten de waarde van de ingezette publieke middelen en de baten de waarde van het effect dat is bereikt. Technische doelmatigheid is tussen alternatieve beleidsopties te vergelijken door te kijken naar de netto baten per uitgegeven euro, waarbij hogere baten per uitgegeven euro gelijk staan aan hogere technische doelmatigheid. In de praktijk richten analyses van technische doelmatigheid zich vaak op de kosten: kan dezelfde dienstverlening gerealiseerd worden tegen lagere kosten?

Allocatieve doelmatigheid

Allocatieve doelmatigheid is te verbeteren door de beschikbare publieke middelen daar in te zetten waar ze het meest opleveren voor de samenleving. Allocatieve doelmatigheid is van toepassing binnen systeem- en transitiebeleid en op de totale uitgaven van het systeem of de transitie.

Door publieke middelen binnen een systeem in te zetten waar ze het grootste effect opleveren wordt de allocatieve doelmatigheid van het systeem zelf verbeterd. Dit kan door beleid te voeren met de hoogste technische doelmatigheid of door beleid daar in te zetten binnen een systeem waar de doelmatigheid van de uitvoering het hoogst is (hier valt bijvoorbeeld ook het niveau waarop beleid wordt gemaakt en uitgevoerd onder: nationaal of regionaal, centraal of decentraal en top-down of bottom-up).

Ook op het niveau van het systeem als geheel speelt allocatieve doelmatigheid een rol. Stel bijvoorbeeld dat een extra euro meer welvaart oplevert als het aan de digitaliseringsagenda wordt uitgegeven dan aan klimaatbeleid. In dit voorbeeld is het allocatief doelmatig om te investeren in de digitaliseringsagenda, zelfs wanneer de euro technisch gezien (positieve netto baten) ook doelmatig te besteden is aan klimaatbeleid. De investering in de digitaliseringsagenda leidt in dit voorbeeld tot hogere gemeenschappelijke baten dan een investering in klimaatbeleid. Het bepalen van het niveau van beleidsinspanningen is dus een allocatief vraagstuk, waarbij afwegingen gemaakt moeten worden tussen de baten die het systeem oplevert en de baten van andere collectieve goederen of maatschappelijke doelen.

Allocatieve doelmatigheid van systeem- en transitiebeleid is te evalueren aan de hand van twee vragen:

- Welk systeem- en transitiebeleid is allocatief doelmatig? Om deze vraag te beantwoorden is informatie nodig over de baten en de kosten van al het S/T-beleid. Baten zijn vaak moeilijk te meten, maar ook kosten zijn dikwijls niet eenduidig toe te wijzen aan specifiek beleid. Coördinatieprogramma's gebruiken bijvoorbeeld vaak bestaande instrumenten, waardoor de toerekening van kosten niet meteen duidelijk is. In de praktijk zal het moeilijk zijn om deze vraag te beantwoorden, omdat daar de relatieve baten van verschillende beleidsinitiatieven ten opzichte van elkaar afgewogen moeten worden;
- Welk niveau van publieke uitgaven is allocatief doelmatig? Om deze vraag te beantwoorden moet in ogenschouw worden

genomen dat doelmatigheid kan variëren met beleidsintensiteit, specifiek als het gaat over de coördinatie-inspanningen die gepaard gaan met S/T-beleid. Te weinig inzetten op coördinatie is niet effectief, waardoor inspanningen tevergeefs zijn, maar te veel coördinatie leidt ertoe dat slechts insiders aan bod komen.

Methoden die kunnen worden ingezet voor het meten van allocatieve doelmatigheid zijn beperkt beschikbaar, omdat er een vergelijking gemaakt moet worden met de opbrengst van middelen voor alternatieve doelen. In de huidige praktijk is eigenlijk alleen het bestuderen van casussen een geschikte methode, omdat de overige methoden zich op één toepassingsgebied richten. Een risico hierbij is dat alle reallocaties worden beoordeeld als positief, terwijl niet duidelijk is of deze wel effectief zijn.

Dynamische doelmatigheid

Zowel technische als allocatieve doelmatigheid zijn statische dimensies van beleid, omdat ze doelmatigheid op een bepaald punt in de tijd evalueren (meestal het einde van een beleidsperiode of een verplicht evaluatiemoment). Door technologische vooruitgang is het mogelijk om in de toekomst een hogere doelmatigheid te bereiken dan met de huidige kennis en met het huidige beleid nu mogelijk is. Voor het evalueren van systeem- en vooral van transitiebeleid is dynamische doelmatigheid belangrijk, omdat het omschrijft hoe het huidige beleid de doelmatigheid in de toekomst beïnvloedt. Meer ruimte voor innovatie kan in de toekomst leiden tot nieuwe oplossingen voor bijvoorbeeld klimaatproblemen die kosten besparen of helpen bij het op een efficiënter manier behalen van klimaatdoelstellingen.

In systeem- en transitiebeleid speelt het stimuleren van veranderingen een belangrijke rol. Innovatie is gericht op nieuwe producten en diensten en vormen van coördinatie, productieprocessen, communicatie of organisatie waarmee het systeem doelmatiger gaat werken. Het kan gaan om technologische vernieuwingen, zoals digitalisering, om organisatorische innovatie, zoals logistieke processen, en externe organisatorische innovatie, zoals outsourcing van productie. De doelen van S/T-beleid bestaan uit het verbeteren van processen, waarbij het kan gaan om primaire (producten en diensten) en ondersteunende (governance) processen en het efficiënter bereiken van maatschappelijke doelen.

Doelmatigheidsbepaling neemt toekomstige baten mee en zet die af tegen de kosten. De kosten-batenverhouding kan door de tijd heen veranderen. Beleid bij aanvang van een transitie kost veel tijd en middelen, terwijl de baten ver in de toekomst liggen. Baten worden zichtbaar op het moment dat de transitie een exponentieel ontwikkelingsverloop laat zien (bijvoorbeeld op prestatie-indicatoren als gerealiseerde besparing op CO₂-uitstoot). De doelmatigheid van het beleid op verschillende punten in de tijd loopt hierdoor uiteen, zeker als het coördinatiebeleid na het realiseren van een kantelpunt afgeschaald kan worden.

Een complicatie bij het evalueren van systemen en transities is dat technische doelmatigheid (lagere kostprijs met dezelfde baten of hogere baten met dezelfde kostprijs) niet consistent hoeft te zijn met dynamische doelmatigheid. Er kan bijvoorbeeld een technologische *lock-in* en padafhankelijkheid ontstaan als op basis van technische doelmatigheid een keuze wordt gemaakt die weinig ruimte biedt aan nieuwe technologische ontwikkelingen die op termijn doelmatiger zijn (Foray, 1997).

Aanwijzingen voor doelmatig beleid

Bij de evaluatie van de doelmatigheid van S/T-beleid is het belangrijk recht te doen aan de dynamische aard van het beleid. Het evalueren van dynamische efficiëntie is echter gecompliceerd door onzekerheid. Er zijn vanuit de optietheorie ten minste drie verschillende vormen van het meten van dynamisch doelmatig beleid mogelijk: (1) inbouwen van de optie om te wachten met het invoeren van beleidsinstrumenten, (2) het opknippen van keuzes in een aantal stappen zodat het beleid tussentijds bijgesteld kan worden (of kan worden gestopt) en (3) experimenteren op kleine schaal met nieuwe mogelijkheden om te leren welk beleid het meest effectief en efficiënt is. Al deze aspecten van beleid zien op het omgaan met onzekerheid over het te voeren beleid, wat een belangrijk kenmerk is van het S/T-beleid. Om tot een oordeel over de doelmatigheid van bepaald beleid te komen, kan dan ook gekeken worden in welke mate deze elementen daadwerkelijk onderdeel vormen van het beleid, denk aan subsidieregelingen die pas worden ingezet als marktontwikkelingen daar aanleiding toe geven, tussentijdse evaluaties gevolgd door bijsturing of daadwerkelijk uitgevoerde beleidsexperimenten.

Wachten

Wanneer een technologie of oplossing voor een knelpunt beschikbaar is, hoeft deze niet meteen te worden ingezet als beleidsinstrument. Onzekerheid over de doelmatige toepasbaarheid van de technologie wordt veroorzaakt door onzekerheid over de prijs van verschillende technologieën om het knelpunt op te lossen. Een grotere mate van onzekerheid leidt hierbij tot het uitstellen van keuzes.

Doelmatigheid van beleid kan met evaluatie worden vastgesteld door te beoordelen in hoeverre informatie is verzameld om de huidige waarde van een technologie te bepalen en af te wegen tegen mogelijk efficiëntere technologie in de toekomst. Bijvoorbeeld de huidige effectiviteit van zonnepanelen om een klimaatdoelstelling te bereiken. Deze waarde fluctueert vanwege veranderingen in de markt van zonnepanelen en de beschikbaarheid van alternatieven voor zon. In het geval van zonnepanelen gaat het bijvoorbeeld om de prijs van grondstoffen, de steeds betere technologie, het aantal producenten en de mogelijkheden om alternatieven, zoals windenergie, in te zetten. Daarnaast zijn er ontwikkelingskosten om te investeren in een bepaalde oplossing of technologie (in termen van opties gaat het om de uitoefenprijs). Deze kosten verschillen per technologie en de evaluatie van doelmatigheid moet in kaart brengen in hoeverre er een afweging is gemaakt over de inzet van middelen. De afweging

die resulteert moet ook een tijdsdimensie bevatten, omdat niet te lang kan worden gewacht met het behalen van doelen. Als blijkt dat de verwachting is dat er niet binnen afzienbare tijd een efficiëntere technologie beschikbaar komt, kan worden geëvalueerd in hoeverre deze beslissing is genomen op basis van voldoende beschikbare informatie en welke kosten en baten resulteren.

Opknippen

Het beoordelen van het proces van sequentieel investeren is een tweede aspect van beleid waarnaar gekeken kan worden in het kader van dynamische doelmatigheid. Het opknippen van investeringen in technologieën met een onzeker rendement gebeurt in veel sectoren, zoals in de luchtvaart en farmacie. In transitiebeleid leidt het opknippen van het proces tot beslismomenten. Een scala aan technologieën is beschikbaar om de klimaatdoelen te behalen. De mix van het ontwikkelen en inzetten van deze technologieën bepaalt waarschijnlijk het succes van het beleid. In een eerste fase van het beoordelen van doelmatigheid moet sprake zijn van keuzes voor een aantal veelbelovende technologieën, die in een tweede fase worden getest of waarmee wordt geëxperimenteerd. De keuzes die hieruit voortkomen kunnen op technische doelmatigheid worden beoordeeld waar het gaat om de kosten van beleid, op allocatieve doelmatigheid waar het gaat om de optie te stoppen of te intensiveren en op dynamische doelmatigheid door te evalueren in hoeverre onzekerheid expliciet in de beleidsafweging in meegewogen. In een tweede fase gaat het om het bouwen van een structuur waarbinnen samenwerking tussen publieke en private partijen plaatsvindt of publieke investeringen worden gedaan. Technische doelmatigheid beoordeelt in hoeverre verspilling van middelen is voorkomen en waar mogelijk gebruik wordt gemaakt van bestaande structuren en instrumenten. Vervolgens kan de uitvoering van beleid worden geëvalueerd op technische doelmatigheid (voorkomen van verspilling en verbeteren van kwaliteit) en op de opties om het beleid te stoppen (bij onvoldoende succes) en aan te passen (bij nieuwe informatie of wanneer tussendoelen zijn bereikt).

Experimenteren

Een derde manier om tot doelmatig beleid te komen is het doen van (kleinschalige) experimenten om erachter te komen welk beleid effectief en efficiënt is. Beleid dat dergelijke methoden inzet, is met grotere kans dynamisch doelmatig dan beleid dat niet experimenteert. Belangrijk is dan wel dat de experimenten zelf ook effectief en doelmatig zijn opgezet, en dat er ook daadwerkelijk lessen getrokken kunnen worden uit de gedane experimenten op het niveau van het systeem of de transitie. Het moet vervolgens zichtbaar zijn dat de resultaten gebruikt worden om beleid vorm te geven.

3.5. Uitdagingen voor evaluatie van S/T-beleid

Het doel van beleidsevaluatie is te bepalen of het beleid op een bepaald moment doeltreffend en doelmatig is, zodat een volgende stap kan worden gezet (doorgaan, aanpassen of stoppen). Een belangrijke stap om te komen tot een evaluatie van

doeltreffendheid is het formuleren van toetsbare hypothesen aan de hand waarvan de effectiviteit van het beleid vastgesteld kan worden.

Dergelijke hypothesen gaan uit van een actie (bijvoorbeeld een subsidie, het creëren van een netwerk, het starten van een coördinerend overleg, neerleggen van een gezamenlijke toekomstvisie), een effect dat die actie beoogt te bereiken (meer innovatie, onderzoek dat beter op elkaar is afgestemd, een verschuiving in de richting van innovatieve projecten) en een mechanisme waardoor de actie tot het gewenste effect leidt (de kosten van investeren in R&D worden verlaagd waardoor er meer innovatie plaatsvindt, door betere coördinatie komen er innovaties tot stand die elkaar versterken).

Eén van de uitdagingen bij de evaluatie van S/T-beleid is het scheiden van de effecten van de individuele elementen van het effect van het S/T-beleid als geheel. Binnen S/T-beleid zitten vaak ook meerdere standaardbeleidselementen die zich richten op gedrag van individuele actoren, zoals subsidie-instrumenten. In de praktijk worden de effecten van het S/T-beleid en de effecten van enkelvoudige beleidsinstrumenten voortdurend met elkaar verward. Wat in de praktijk gebeurt is dat S/T-beleid probeert om een aantal beleidsinstrumenten meer in dezelfde richting te laten werken. Neem bijvoorbeeld de digitaliseringsagenda. Daar wordt de effectiviteit van S/T-beleid op het vlak van deze agenda bijvoorbeeld bepaald door te kijken naar uitkomstmaten die gerelateerd zijn aan de mate van digitalisering, zoals het aantal breedbandverbindingen of de mate van digitale geletterdheid. Het S/T-beleid op het vlak van digitalisering bestaat echter uit een verzameling 'gewone' subsidieregelingen en daarnaast (onder meer) uit coördinerende activiteiten en prikkels voor partijen om samen te werken. Om het effect van S/T-beleid op digitalisering te bepalen, moet bepaald worden wat de meerwaarde is van de S/T-elementen ten opzichte van de traditionele elementen. Hoeveel leveren die extra activiteiten op, ten opzichte van een wereld waarin die activiteiten niet zouden plaatsvinden? Welke functies in de werking van het systeem moeten veranderen? En wat kan de mix aan instrumenten op het terrein van digitalisering daaraan doen? Box 3.2 illustreert deze uitdagingen aan de hand van de evaluatie van de Topsectorenaanpak.

Hier bevindt zich de kern van de uitdaging voor de evaluatie van S/T-beleid. Het evalueren van wat die betere coördinatie nu precies oplevert voor een beter werkend systeem ten opzichte van een wereld waarin die coördinatie niet plaatsvindt, is namelijk vele malen complexer en ingewikkelder dan het evalueren van het effect van een enkelvoudig instrument dat zich min of meer direct richt op het beïnvloeden van gedrag van individuele entiteiten.

Essentieel is dat een combinatie van interventies moet leiden tot het bevorderen van omstandigheden die tot systeemverbeteringen of wenselijke transities leiden. In hoeverre beleid hier een doeltreffende en doelmatige bijdrage aan levert vergt inzicht in het samenspel van alle gemobiliseerde of bijgestuurde

interventies alsook de uiteenlopende systeemveranderingen die zij bewerkstelligen. Vanwege de meervoudige interventies, het relatief grote belang van governancestructuren en processen (t.o.v. instrumenten of een begrotingsartikel) en de lange effectketen is de scope van evaluaties daardoor ruim.

Stern et al. (2012) constateren dat het gebrek aan geschikte methoden voor evaluatie van systeembeleid terug te voeren is op de aanzienlijke complexiteit van het geheel aan acties en interventies dat daarbij komt kijken. Het vaststellen van causale effecten wordt bemoeilijkt door de grote schaal, brede scope, geneste structuur, multidimensionaliteit en het multilevelkarakter van systeembeleid. Daarnaast verandert het beoordelingscriterium. Het gaat niet om doelrealisatie, maar om functieveranderingen in het systeem die bepalen in hoeverre de werking van het systeem verbetert of de transitie de goede kant opgaat. S/T-beleid wordt in deze context meer beoordeeld op basis van de beweging die tot stand wordt gebracht dan op basis van het niveau dat op een zeker moment in de tijd is bereikt.

In algemene zin geldt dat het verrichten van zuivere effectmetingen vergt dat er zicht moet zijn op wie welke interventie van toepassing is geweest en op wie niet, en hoe hun prestaties zich door de tijd hebben ontwikkeld qua niveau en richting (liefst in relatie tot situatie vóór de interventie). Bij S/T-beleid wordt door het systeem als geheel doorgaans niet aan deze voorwaarden voldaan (hoewel voor subsystemen of onderdelen van het systeembeleid effectmeting soms wel mogelijk is), bijvoorbeeld doordat:

- De exacte doelgroep waarop beleid van toepassing is vaak niet helder gedefinieerd is. Het is niet altijd duidelijk wie vanuit het beleid iets 'krijgt' of in een bepaalde mate beïnvloed wordt. Het feit dat de aanpak leunt op een beleidsmix hoeft geen probleem te zijn, zolang voor elk instrument bekend is wie welke (positieve of negatieve) prikkel heeft gekregen. Beleid dat dient om de condities voor innovatie te vergroten grijpt echter niet (altijd) aan op organisaties; denk aan regelgeving, infrastructuur of kennis die een bijdrage moet leveren op innovaties t.b.v. een bepaalde transitie. Effectmeting vergt een beeld van verschillen in de mate waarin individuele of groepen van organisaties hierdoor geraakt worden, omdat die differentiatie de basis vormt voor het verklaren van prestatieverschillen. Als beleid meerdere instrumenten omvat die zich richten op systeemfactoren (i.p.v. steun aan specifieke doelgroepen) kan het erg lastig worden om te bepalen wie hoeveel voor- of nadeel van het beleid heeft gehad, en wie dus een goede experimentele of controlegroep kunnen vormen;
- Het vinden van een *counterfactual* is zoals vastgesteld vaak lastig of zelfs onmogelijk, omdat er doorgaans sprake is van slechts één systeem/transitie per domein. Bij interventies op organisatie-niveau is het doorgaans eenvoudiger om behandelde en niet-behandelde organisaties te identificeren die over dezelfde tijdsperiode vergeleken kunnen worden. Voor metingen op het niveau van ultieme doelen als het bewerkstelligen van een

Box 3.2. Evaluatie van de Topsectorenaanpak illustreert uitdagingen

De evaluatie van de Topsectorenaanpak illustreert de uitdagingen waarmee een evaluatie van S/T-beleid gepaard gaat (Janssen, 2019):

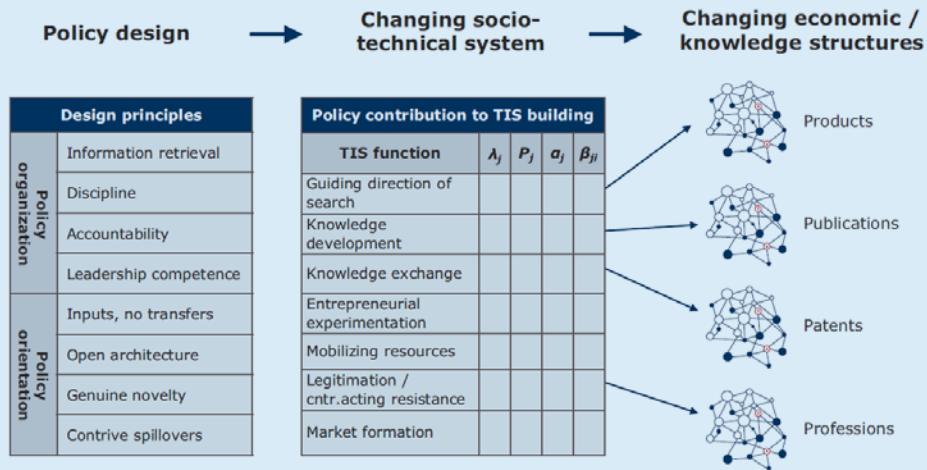


Fig. 6. Overall assessment scheme.

De evaluatie van de governancestructuur had de vorm van het identificeren en vervolgens beoordelen van beleidsprincipes die voor deze specifieke beleidsaanpak van belang werden geacht. Een literatuuranalyse liet zien dat de literatuur acht beleidsprincipes onderscheidde die relevant worden geacht voor het creëren van condities waardoor overheden, bedrijven en kennisinstellingen samen nieuwe kansrijke innovatiepaden kunnen verkennen en versterken. Deze beleidsprincipes waren echter niet gebaseerd op empirisch onderzoek dat liet zien dat, wanneer aan dit principe voldaan was, innovatie-uitkomsten zouden verbeteren.

De beleidsevaluatie bestond daarnaast uit het beoordelen van de samenhang van de initiatieven ten behoeve van systeemversterking. Leidend was de vraag of de initiatieven consistent waren met de voornaamste knelpunten in de systemen waar de Topsectoren zich op richtten. De analyse vond plaats aan de hand van een raamwerk van sleutelprocessen die van belang zijn voor een goed functionerend technologisch innovatiesysteem. In het onderzoek is ook een poging gedaan om uitspraken te doen over hoe goed de interventies werkten. Gezien de veelheid van interventies voor ieder van de negen Topsectoren was het niet mogelijk hier een empirische analyse van te doen.

Waar een effectmeting bij de Topsectorenaanpak betrekking op zou moeten hebben was niet evident. De uiteindelijke beleidsdoelen bestonden uit het vergroten van de publiek-private uitgaven aan PPS-projecten, het verhogen van de algemene R&D-intensiteit in Nederland en een stijging van Nederland in de ranking van competitieve landen. Het eerste beleidsdoel houdt het meest verband met datgene wat er beleidsmatig feitelijk ondernomen is. Ten tijde van de evaluatie was het niet mogelijk om iets te zeggen over de werkelijke gegenereerde uitkomsten, o.a. omdat de recente invoering van de systeemaanpak nog onmogelijk te relateren was aan macro-uitkomsten als een hogere R&D-intensiteit.

circulaire maakindustrie resteren al snel enkel ruwe vergelijkingen door de tijd of met andere landen; deze lenen zich slecht voor het doen van uitspraken over de specifieke invloed van beleid;

- Er niet altijd overeenstemming is over wat nu precies het beleidsinstrument is dat geëvalueerd moet worden als men het over het evalueren van systeembeleid heeft. Vaak bestaat systeembeleid uit een combinatie van verschillende instrumenten op uiteenlopende beleidsterreinen, waarbij instrumenten ook nog eens heel uiteenlopend van aard zijn. Systeembeleid richt zich dan op coördinatie of samenhang tussen de instrumenten om de werking van het systeem te verbeteren. Evaluatievragen met betrekking tot systeembeleid moeten daarom duidelijk worden geformuleerd in termen van de instrumenten die in samenhang moeten worden

geëvalueerd en in hoeverre de invloed van het beleid op de samenhang of coördinatie zelf onderdeel is van de evaluatie. Zelfs als er overeenstemming is over concrete interventies die van belang zijn voor systeemversterking of een transitie, kan het nog lastig zijn welk aspect/deel hiervan gekoppeld kan worden aan het S/T-beleid. Alleen al het budget dat gemoeid is met bepaald systeembeleid kan daardoor onduidelijk zijn, hetgeen belemmerend is voor het doen van uitspraken over de doelmatige inzet van middelen;

- Tot slot is S/T-beleid soms gericht op het aanjagen van ontwikkelingen die nu nog niet plaatsvinden of op het realiseren van transities die worden gekenmerkt door niet-lineaire effecten, waarbij het effect van beleid in de tijd sterk kan fluctueren. Beide aspecten maken het ingewikkeld om te zeggen of een uitkomst al (voldoende) onderweg is.

4. Evaluatiekader systeem- en transitiebeleidsprogramma's

Dit hoofdstuk presenteert een integraal evaluatiekader voor S/T-beleid. Allereerst worden uitgangspunten geïntroduceerd, om te zorgen dat duidelijk is welke aannames en principes leidend zijn bij het evalueren van S/T-beleid. Vervolgens presenteren we stapsgewijs perspectieven en kaders uit de wetenschappelijke literatuur die bruikbaar kunnen zijn voor evaluatie.

4.1. Algemene uitgangspunten

Er is een aantal uitgangspunten voor evaluatie.

Beleidstheorie. Er moet sprake zijn van een falen. Aan de hand van het falen moet een heldere omschrijving van het beleidsinstrument dat wordt ingezet worden opgesteld, het beleidsdoel dat met het beleidsinstrument wordt nagestreefd moet worden gedefinieerd en de mechanismen waardoor beleidsinspanningen tot het gewenste beleidsdoel zouden moeten leiden moeten worden bepaald.⁹ Het is daarbij relevant om te ontleden hoe het spectrum van beleidsinspanningen en -instrumenten eruit ziet, en voor ieder daarvan na te gaan welk deel van de beleidstheorie erbij hoort als het gaat om de keten van *input – throughput – output – outcome – impact*. Daarnaast moet helder zijn wat de financiële gevolgen zijn voor de overheid (en indien relevant voor andere actoren) en hoe overheidsfalen wordt voorkomen. Dit maakt het mogelijk om te schakelen tussen verschillende observatieniveaus; effecten lager op de as van validiteit van beleidsattributie (causaliteit) zijn vaak eenvoudiger in kaart te brengen dan hogere ordes.

Belangrijk in deze eerste stap is om te analyseren in hoeverre het beleidseffect slechts de optelsom van de effecten van afzonderlijke beleidsinspanningen is of dat sprake is van synergie.

Een cruciaal ingrediënt van de beleidstheorie is om vast te stellen wat nodig is om een conclusie te kunnen trekken over doeltreffendheid en doelmatigheid van het S/T-beleid.¹⁰ We onderscheiden de volgende niveaus waarop het bewijs voor hypothesen kan worden getoetst: consistentie, falsificatie, correlatie en causaliteit. Van consistentie is sprake als waarnemingen consistent zijn met het veronderstelde mechanisme (er worden minder dieselmotoren verkocht). Van falsificatie is sprake indien het mogelijk is om tegenbewijs te vinden voor de beleidstheorie (verkoop van dieselmotoren is gestegen ondanks het beleid). Van correlatie is sprake als waarnemingen van het ingezette beleidsinstrument samenvallen met gewenste uitkomsten (een daling van de uitstoot van CO₂ valt samen met een daling van de verkoop van dieselmotoren). Van causaliteit is sprake als waarnemingen de oorzaak zijn van gewenste uitkomsten (door het beprijzen van CO₂-uitstoot zijn dieselmotoren minder aantrekkelijk geworden en is de verkoop ervan gedaald). Per niveau (*input – throughput – output – outcome – impact*) kan deze vierslag gemaakt worden, waarbij dient te worden gestreefd naar causaliteit. Tabel 4.1 geeft dit weer.

Tabel 4.1. Effectenmatrix om te bepalen welke effectladder-trede per geïdentificeerd mechanisme mogelijk is

	Input	Throughput	Output	Outcome	Impact
1. Consistentie					
2. Falsificatie					
3. Correlatie					
4. Causaliteit					

⁹ Dit dient naast economisch ook volgens de Comptabiliteitswet te gebeuren. In artikel 3 van die wet staat dat voorstellen, voornemens en toezeggingen een toelichting dienen te bevatten waar het onder andere gaat om de doelstellingen en de effectiviteit en efficiëntie die worden nagestreefd.

¹⁰ Bekend is de Maryland Scientific Methods Scale. In de context van evaluaties spreekt men ook wel van de 'effectladder'. De laddermetafoor geeft aan dat er een hiërarchie bestaat in hoe geschikt bepaalde onderzoeksmethoden zijn om uitspraken te doen over beleidseffecten.

Hierbij is er een hiërarchie, waarbij causaliteit boven correlatie staat, correlatie boven falsificatie en falsificatie boven consistentie. Bij enkelvoudige beleidsinstrumenten, waarbij het beleidsdoel helder gedefinieerd is, de bereikte populatie bekend is, de *counterfactual* te identificeren is en het instrument niet continu bijgesteld wordt, zijn er in principe goede voorwaarden om causaliteit vast te stellen. Er is immers zicht op een *counterfactual*, zoals een controlegroep van organisaties die zo veel mogelijk lijkt op de behandelde groep, en van wie eveneens het prestatieverloop bekend is. Ook is er in omvang en kwaliteit voldoende data beschikbaar om het instrument te evalueren op doeltreffendheid.¹¹

Data. Welke data beschikbaar zijn in het kader van een beleidsevaluatie, bepaalt welke aspecten van beleid gemeten kunnen worden. Systeembeleid bestaat uit instrumenten en grijpt daarnaast vaak aan op coördinatieproblemen en coördinatie-activiteiten. Het bestaan van structuren met een goed functionerende governance is daar een voorbeeld van. Een ander voorbeeld is in hoeverre beleidsaanpassingen resulteren in geïntensiveerde of herziene coördinatie. Dit maakt het mogelijk om aan de hand van de effectenmatrix te identificeren op welk niveau een evaluatie zich kan richten en welke methoden ingezet kunnen worden.

Type evaluatie. Evaluaties van S/T-beleid vragen niet enkel om een standaard effectmeting, aangezien het beleid weliswaar zijn legitimatie ontleent aan de outcome die het genereert, maar zich vooral richt op een combinatie van instrumenten (die elkaar dienen te versterken en/of aan te vullen) en het sturen van de processen die leiden tot deze outcome. De vraag die voorligt is tweeledig. Ten eerste is de vraag in hoeverre interventies leiden tot de gewenste bijsturing van processen en bijdragen aan het gewenste transitiepad. Ten tweede is het de vraag in hoeverre het bijsturen van de processen tot de gewenste outcome leidt. Beide zijn ook vormen van effectmeting, maar op andere outcome-variabelen. Het probleem bij evaluaties van S/T-beleid zit vaak in de tweede vraag, omdat die vraag meestal niet causaal te beantwoorden is met de beschikbare methoden.

Het evalueren van S/T-beleid is voor een deel gericht op input en proces, omdat de overheid probeert een maatschappelijk probleem aan te pakken. Een randvoorwaarde voor doeltreffendheid is dan dat de overheid processen inricht die het aannemelijk maken dat een relevante bijdrage wordt geleverd aan het oplossen van het probleem (Boon & Edler, 2018). Een oordeel over de ingerichte processen vormt daarmee een stap in de evaluatie van S/T beleid. De inschatting van die aannemelijkheid vindt in de praktijk plaats door met theoretische en kwalitatief-analytische methoden een beeld te schetsen van voorwaarden waaraan effectief beleid zou

moeten voldoen. Evaluatie is in dat geval een exercitie in het nagaan of het beleid ook daadwerkelijk aan deze voorwaarden voldoet. De beperking van deze aanpak is dat het probleem van het meten van de effecten van beleid slechts verschuift. Het is namelijk belangrijk dat aannemelijk wordt gemaakt dat de voorwaarden die in de literatuur als effectief worden benoemd in de toepassing in de specifiek Nederlandse context ook effectief zijn (externe validiteit). De wetenschappelijke literatuur over effectiviteit van S/T-beleid is grotendeels niet-empirisch of gebaseerd op gefragmenteerde casuïstiek. Het scoort dus laag op causaliteit in Tabel 4.1.

Er wordt daarom in toenemende mate gepleit voor formatief evalueren (Molas-Gallart et al., 2021), al dan niet in combinatie met een summatieve evaluatie met het oog op verantwoording (Magro & Wilson, 2019). Het concept formatief evalueren heeft zijn oorsprong in het onderwijs. Bij formatief evalueren staat de voortgang richting het einddoel en feedback over het leerproces centraal, terwijl bij summatief evalueren een oordeel over het wel of niet (in voldoende mate) bereiken van het einddoel centraal staat. Vanwege veranderende omstandigheden is het bij S/T-beleid vaak onvoldoende om te oordelen of beleid in het verleden doeltreffend is geweest, aangezien effectiviteit op een bepaald moment in de tijd geen garantie geeft dat het beleid bij nieuwe systeem- en transitiekelpunten nog steeds doeltreffend is. Een voorbeeld is de constatering dat de Topsectorenaanpak heeft geholpen bij het aanjagen van nieuwe samenwerkingsrelaties, maar dat het blijven opbouwen hiervan afnemende meeropbrengsten genereert die maar beperkt bijdragen aan het bereiken van de beoogde beleidsdoelen. Op het moment dat er nog meer nieuwe samenwerkingsstructuren tot stand worden gebracht, is het effectiever andere systeemfactoren onder handen te nemen. Het is daarom vooral zaak om te weten hoe het systeem ervoor staat, welk beleid daarbij past en of organisatie en processen effectief functioneren. Om deze doeltreffendheid te beoordelen bieden eerdere beleidservaringen en wetenschappelijke literatuur handvatten. Dit perspectief sluit aan bij het idee dat methoden die hoog scoren op causale bewijskracht kunnen worden aangevuld met andere evaluatiemethoden.¹²

Een eigenschap die gedeeld wordt in enkele van de alternatieve evaluatiemethoden is dat er meer ruimte is voor het betrekken van beleidsuitvoerders in het doorgronden van de beoogde en reële werking van beleid, en de factoren die daarop van invloed zijn (Barbrook-Johnson et al., 2021). Beleidsuitvoerders kunnen bijvoorbeeld participeren in het definiëren (en toepassen) van criteria voor procesevaluaties, hoewel niet duidelijk is dat dit wenselijk is vanuit de invalshoek van een onafhankelijke evaluatie. Het gezamenlijk ontwerpen en uitvoeren van evaluatieonderzoek past binnen de aanpakken van *developmental evaluation* (Patton,

¹¹ Overigens is de praktijk hier minder goed dan de theorie. Beleid dat in principe volgens deze systematiek geëvalueerd zou kunnen worden, wordt dat vaak niet. Dit heeft er mee te maken dat de inspanningen voor een dergelijke evaluatie groot zijn (en dus zijn dergelijke evaluaties kostbaarder), de risico's dat de evaluatie onwelgevallige uitkomsten oplevert ook groot zijn. Ook zijn deze evaluaties vaak wat eendimensionaal en qua scope beperkt, waardoor beleidsmakers niet het gevoel hebben handvatten aangereikt te krijgen waardoor ze het beleid kunnen verbeteren.

¹² Van der Steen (2020). Opgavegericht evalueren: methoden die passen bij het beleidsvraagstuk. P. 35 in: Beleidsevaluatie in theorie en praktijk (Van der Knaap et al., 2020).

2011) en lerend evalueren (Verwoerd et al., 2020). Door Van der Knaap et al. (2020) wordt gesproken over responsieve evaluatie (waarbij richting en duiding continu gezocht moeten worden in de belangen van diverse stakeholders) en van contextueel-realistische beleidsevaluatie (met aandacht voor de vraag via welke mechanismen en onder welke omstandigheden beleid werkt). Arnold et al. (2018) staan stil bij de handvatten die dergelijke evaluatieperspectieven bieden voor evaluatie van complexe innovatie- en transitie-beleidsaanpakken. Ze wijzen erop dat bij systeembeleid voortschrijdend inzicht een belangrijke rol speelt, omdat op voorhand niet duidelijk is via welke tussendoelen en beleidsinspanningen het ultieme beleidsdoel behaald kan worden. Het monitoren van de voortgang en het behalen van doelen is daarom van groot belang voor een oordeel over doeltreffendheid en ook in het bepalen van (dynamische en allocatieve) doelmatigheid.

4.2. Beleidstheorie

In deze paragraaf bespreken we een generieke beleidstheorie die voor iedere evaluatie van systeem- en transitiebeleid als startpunt zou moeten dienen (Figuur 4.1, op basis van Janssen et al., 2022).

Voor daadwerkelijke toepassingen zal deze uiteraard moeten worden verfijnd om voldoende aanknopingspunten te geven voor evaluaties. Voordat een evaluatie plaatsvindt is een noodzakelijke voorwaarde dat er een uitgewerkte beleidstheorie is opgesteld. Uitwerking van een beleidstheorie vindt bij voorkeur plaats voordat beleidsmakers met S/T-beleid starten. Zonder een van tevoren opgestelde beleidstheorie bestaat het risico dat een evaluatie uiteindelijk niet meer is dan een beschrijving of poging tot legitimatie van wat er heeft plaatsgevonden (ook omdat er waarschijnlijk onvoldoende data beschikbaar zijn). De beleidstheorie voor S/T-beleid bestaat uit een aantal elementen.

- *Te behalen maatschappelijk doel.* Het startpunt van de beleidstheorie wordt gevormd door een maatschappelijke wens of behoefte die via de bestaande gang van zaken niet vervuld kan worden (bijvoorbeeld duurzame energieopwekking). Het is belangrijk om als startpunt het relevante falen te bepalen dat ervoor zorgt dat het maatschappelijk doel niet zonder overheidsfalen tot stand komt (bijvoorbeeld niet of onvoldoende beprijzen van milieuvuiling leidt tot een gebrek aan duurzame energieopwekking, waardoor er teveel milieuvuiling is).
- *Interventies.* S/T-beleid probeert hier verandering in te brengen, in de vorm van een integraal programma dat middels synergetische beleidsinstrumenten verandering probeert te bewerkstelligen (bijvoorbeeld de energietransitie).
- *Coördinatie van inspanningen middels het S/T-beleidsprogramma.* De coördinatie grijpt aan op bestaande en nieuwe

beleidsinitiatieven die op een systeem inwerken en het veranderen en optimaliseren. De beleidsinitiatieven kunnen specifiek zijn voor een bepaalde systeemfactor (zoals kennisontwikkeling, marktforming en menselijk kapitaal) of een systemische scope hebben (zoals overleg en coördinerende financiële prikkels).

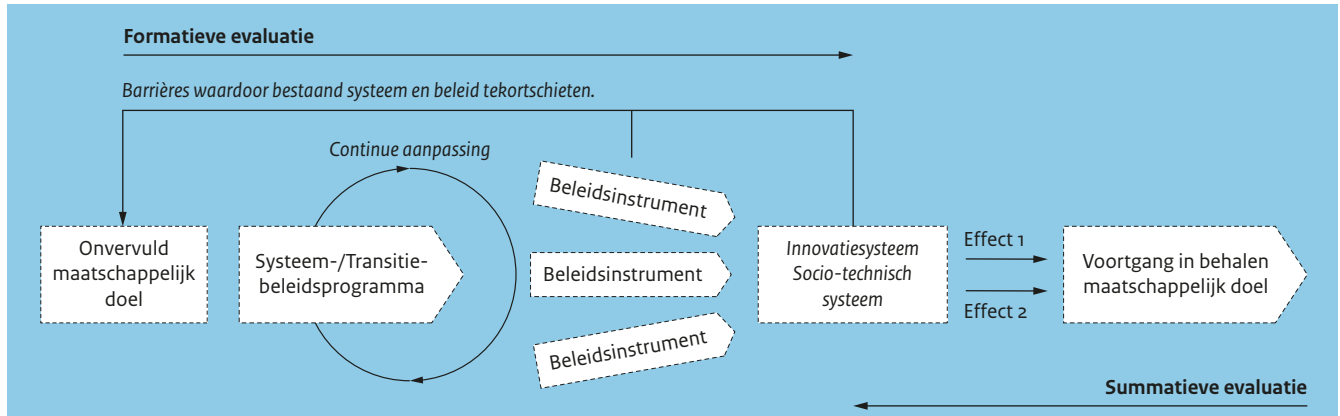
- *Continue beleidsmixaanpassing op basis van nieuwe kennis.* Er is een continue aanpassings- of feedbackloop waarbij het programma wordt aangepast om de doelen te bereiken. Sommige instrumenten zijn effectiever dan andere en nieuwe technologische en maatschappelijke ontwikkelingen veranderen de effectiviteit van de interventies.
- *Meten van (tussen)uitkomsten.* S/T-beleid probeert middels synergetische instrumenten veranderingen aan te brengen in een systeem. Het versterken of transformeren van een systeem is al een tussenuitkomst op zich. Dit dient vervolgens te leiden tot effecten die gerelateerd zijn aan een beleidsdoel. Bijvoorbeeld: onderzoeks- en marktpartijen werken vaker samen bij het doen van onderzoek en richten zich daarbij op onderwerpen die als relevant zijn gedefinieerd door de overheid, of er verschijnen bijvoorbeeld na verloop van tijd laaggeprijsde zonnepanelen op de markt die helpen bij het verduurzamen van de energieopwekking.¹⁵
- *Monitoren van transitiepaden.* Helemaal rechts in Figuur 4.1 bevinden zich de observaties over hoe de transitie feitelijk verloopt. Er moeten indicatoren worden opgesteld die het transitiepad monitoren om een beeld te krijgen van de voortgang. Deze indicatoren zeggen op zichzelf nog niets over de invloed van beleid.

4.3. Twee evaluatieperspectieven

In de literatuur over S/T-beleid bestaat een aantal studies dat pleit voor het combineren van evaluatiemethoden. Turnheim et al. (2015) beschrijven bijvoorbeeld welke complementaire analyses relevant zijn voor het onderzoeken van verschillende schalen waarop transitiepaden zich ontwikkelen (initiatief-gebaseerde niches, en het socio-technische systeem als geheel) en Arnold et al. (2018) stellen evaluatie-aanpakken voor die zijn gekoppeld aan verschillende niveaus van beleidsverantwoordelijkheden (zoals ministeries, uitvoeringsorganen, programmamanagers en projectleiders). Een recente OESO-studie naar missiegedreven innovatiebeleid (Larrue, 2021, p. 14) stelt dat, ondanks een duidelijke trend naar systeembeleid, de bestaande evaluatie-instrumenten en -technieken vooral geschikt zijn voor individuele programma's en instrumenten. Het gebrek aan systeembrede evaluatiemethoden is vooral te wijten aan de complexiteit van effectmeting en attributie bij beleid dat wordt gekenmerkt door de grote schaal en reikwijdte, de geneste structuur waardoor beleid

¹⁵ Relevant is dat hier op diverse niveaus beleidsdoelen kunnen zijn, zoals het reduceren van CO₂-uitstoot, het realiseren van meer zonnepanelen, een groter aandeel duurzame energieproductie of het vergroten van de kosten van CO₂-uitstoot. Al deze doelen kunnen legitiem onderdeel vormen van de beleidstheorie.

Figuur 4.1. **Generieke beleidstheorie voor S/T-beleid, met daaraan gekoppeld twee evaluatieperspectieven***



Janssen, M., Bergek, A., Wesseling, J. (2022). Evaluating systemic innovation and transition programmes: Towards a culture of learning. *PLOS Sustainability and Transformation*, 1(3): e0000008. n

onderdeel is van een groter geheel aan doelen en een multidimensionaal en gelaagd ontwerp.

De beperking van de in de systeem- en transitieliteratuur voorgestelde evaluatiemethoden is weer dat deze niet primair gericht zijn op de verantwoordingsvraag of beleid effectief is. Het uitgangspunt bestaat vooral uit het nalopen van principes waarvan verondersteld wordt dat die een voorwaarde vormen voor effectief beleid. Op het moment dat beleid aan deze principes getoetst wordt is de kwaliteit van die toets sterk afhankelijk van de onderbouwing van die principes. Een voorbeeld van een principe is dat de sturing bij innovatiebeleid in handen moet zijn van ‘open’ structuren waarin zeer uiteenlopende actoren betrokken zijn (Schot & Steinmueller, 2018). Hoewel de literatuur aanwijzingen geeft dat dit principe helpt voorkomen dat een kleine groep gevestigde partijen de dienst uitmaakt (netwerkfalen), bestaan er situaties waarin dit principe niet effectief is (falsificeerbaar). Principes zijn uitgangspunten waarvan wordt aangenomen dat ze een leidraad vormen, maar het zijn geen wetmatigheden. Een ander voorbeeld van een principe is dat instrumenten consistent moeten zijn met elkaar en met de overkoepelende beleidsstrategieën waarbij ze horen (Rogge & Reichardt, 2016). Dit is een principe waarvan erkend wordt dat het in de praktijk lastig is om aan te voldoen. Beleidsstrategieën putten immers vaak uit bestaande instrumenten en processen die een nieuwe optimale beleidsmix in de weg staan. Het gaat er bij formatieve evaluatie dan ook niet enkel om om vast te stellen in hoeverre het beleid voldoet aan relevante principes (het antwoord zal vaak ontkennend zijn), maar ook om te bepalen waar onvolmaaktheden zitten en hoe deze imperfecties kunnen worden opgelost of verminderd.

Uitgaande van een generieke beleidstheorie kan worden gesteld dat formatieve evaluaties zich vooral lenen om een beeld te vormen van de (te beïnvloeden) processen, structuren en mechanismen die tot systeemverandering en transitie moeten leiden. Summatieve evaluaties trekken conclusies over in hoeverre er als gevolg van beleid ontwikkelingen optreden die voor een beleidsdoel van belang zijn.

Formatieve of normatieve evaluatie

Formatieve evaluatie richt zich op (tussentijds) evalueren van de voortgang richting een einddoel en daarbij te leren over de werking van beleid, opdat daarin bijgestuurd kan worden. Bij aspecten als onderzoeksontwerp, dataverzameling, interpretatie en het ‘ontvankelijk maken’ voor uitkomsten wordt vaak gekozen voor participatieve methoden (bijvoorbeeld een intensieve betrokkenheid van beleidsmakers en -uitvoerders). Daarbij is wel van belang dat de onafhankelijkheid van het onderzoek goed wordt geborgd, zowel in de vraagstelling als in de uitvoering.

Formatieve evaluatie leent zich onder andere om antwoord te geven op vragen aangaande de kwaliteit en geschiktheid van sturingsmechanismen en governance waarop het S/T-beleid is gebaseerd. Er is meestal geen kant-en-klaar beleidskader dat aangeeft welke proceseigenschappen en -indicatoren relevant zijn voor een beleidsaanpak. Eerst moet worden uitgezocht wat de ontwerpprincipes of normen zijn waar het proces aan getoetst kan worden. Twee essentiële stappen zijn: het *identificeren* van principes met behulp van literatuuronderzoek en/of raadpleging van partijen die nauw betrokken zijn geweest bij het ontwerp (en bijstellen) van het beleid; en het *toetsen* van de principes met meer betrokkenen (niet alleen ontwerpers), liefst met een groep van partners en stakeholders die kritisch zijn en onafhankelijke deskundigen.

Uitgangspunt van de formatieve evaluatie is om te achterhalen welke activiteiten en structuren er door beleidsmakers zijn opgezet om aan de ontwerpprincipes te voldoen. Als het bijvoorbeeld belangrijk is dat de beslisstructuur ruimte laat voor inspraak vanuit partijen met oplossingsrichtingen die niet opereren in het hart van het systeem, dan is het onvoldoende om enkel te polsen hoe betrokkenen hierover denken. Cruciaal is om de principes te kennen en als evaluator zelf een beeld te vormen door te onderzoeken welke activiteiten ondernomen zijn om deze inspraak in beeld te krijgen (met bijvoorbeeld netwerkevenementen), om die inspraak te wegen (met bijvoorbeeld het instellen van een neutrale adviesraad om de inspraak te waarborgen) en om

besluitvormingsprocedures en interventies hiervoor open te stellen (met bijvoorbeeld instrumenten).

Summatieve of positieve evaluatie

Effectmeting is een krachtig instrument. (Quasi-)experimentele ontwerpen dienen om een zo zuiver mogelijke effectmeting mogelijk te maken, waardoor het gerechtvaardigd kan zijn om de analyse toe te spitsen op slechts een deel van de regeling of het instrument, van de bereikte groep, van de jaren waarin het beleid liep, van de doelen/effect-ordes uit de beleidstheorie etc. Dit is bijvoorbeeld aan de orde als een competitieve regeling gebruikmaakt van scores en er een vergelijking wordt gemaakt tussen indieners die net een onvoldoende score ontvangen ten opzichte van indieners die net voldoende scores. De focus op een zo valide mogelijke effectmeting kan in dat geval ten koste gaan van een completer (maar minder robuust) inzicht van wat de regeling in de volle breedte heeft opgeleverd. Om die reden kan het wenselijk zijn om binnen de rubriek outcome-evaluaties met aanvullende evaluatietechnieken te werken om de effectmeting completer te maken.

Op het moment dat van verschillende regelingen bekend is hoe hun individuele effectiviteit eruitziet, geeft dat ten dele zicht op de mate waarin er sprake is van systeemverandering. Idealiter wijst een effectmeting uit of de gegenereerde uitkomsten optellen naar omstandigheden die gunstig zijn voor het uiteindelijke beleidsdoel waaraan het beleid moet bijdragen. Soms is het mogelijk om dit empirisch na te gaan. Als bijvoorbeeld bekend is welke partijen of innovatierichtingen baat zouden moeten hebben bij het beleid, is de vraag of zij ook merkbare voordelen hebben ondervonden van de in gang gezette systeemverandering. In hoeverre die systeemverandering middels attributie-analyse weer te herleiden is tot beleidsinspanningen is relevant, maar ook als dat niet het geval is kan het zinvol zijn om in ieder geval te weten op welke punten het systeem voldoende werkt en op welke niet. Er bestaat een spanning als het gaat om de doeltreffendheid van beleid en het meten in hoeverre bepaalde condities bestaan (ongeacht bewijs voor beleidseffect).

4.4. Theoretische afwegings- en analysekaders voor S/T-beleid

Er bestaat geen universeel S/T-evaluatiekader. Met een synthese van de beschikbare wetenschappelijke literatuur en evaluatie-praktijk zijn wel enkele invalshoeken te onderscheiden die relevant zijn bij het evalueren van S/T-beleid (Janssen et al., 2022). Daarbij gaat het om de zes elementen zoals getoond in Figuur 4.2: (A) interventie-rationale, (B) governanceprocessen/-structuren en de beleidsmix, (C) aansluiting tussen beleid(mix) en knelpunten, (D) systeemversterking/-transformatie, (E) structurele verandering en (F) maatschappelijke impact.

Voor iedere specifieke invalshoek is het mogelijk om concrete hypothesen op te stellen die aangeven welke verandering er op dat aspect van de beleidstheorie verwacht mag worden. Hierbij kan de effectladder worden benut om na te gaan op welk niveau

(input – acties – output – outcome – impact) conclusies kunnen worden getrokken; de mate waarin deze conclusies causaal zijn; en welke data nodig zijn om conclusies over doeltreffendheid en doelmatigheid te kunnen trekken. In de praktijk zullen niet alle horizontale of verticale doel-ordes ingevuld kunnen worden. Belangrijk is om te laten zien voor welke elementen welke conclusie kan worden getrokken en hoe de gehele balans van mogelijke conclusies over de doeltreffendheid en doelmatigheid van S/T-beleid eruitziet (Figuur 4.2).

A. Interventie-logica

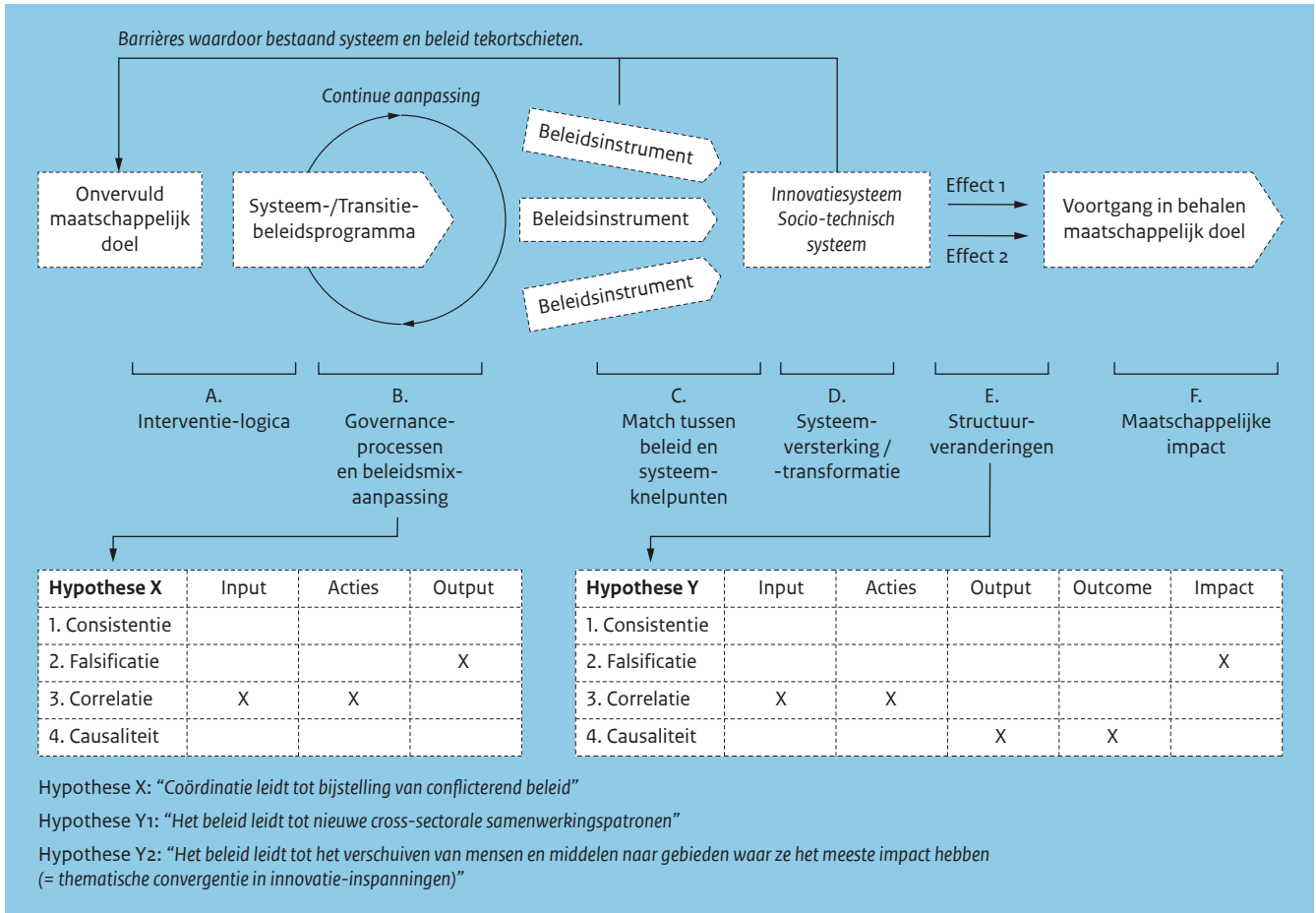
De evaluatievraag die leidend is voor de interventie is of het beleid legitiem is. Om dit te bepalen kunnen er verschillende soorten falen als leidraad genomen worden (zie de discussie in paragraaf 3.2). Net zoals bij regulier beleid is het onderzoeken van de interventielogica niet een kwestie van het benoemen van welk type falen hier mogelijk aan de orde zou kunnen zijn, maar het onderzoeken van bewijs dat een bepaald probleem zich voordoet en in hoeverre met beleid dit probleem kan worden verminderd of zelfs opgelost. Het is eenvoudig te *veronderstellen* dat ondernemers een bepaalde duurzame technologie niet toepassen omdat ze bijvoorbeeld de voordelen ervan niet goed onderkennen (marktfalen: informatie-asymmetrie) of omdat daarvoor een infrastructuur onderontwikkeld is of opstartkosten te hoog zijn (transitiefalen: ontbreken van een heldere richting). Beide voorbeelden van falen vragen om een verschillende interventie. Dit betekent dat het benoemen van een potentieel relevant type falen het startpunt is van het aantonen dat de interventie aangrijpt op een knelpunt dat het ook daadwerkelijk kan verhelpen of verminderen.

Wanneer het gaat om S/T-beleid zijn beleidsmakers gebonden aan kaders, een veelheid aan doelen en interactie met partijen met andere doelen. De mate van beleidscoördinatie en reflexiviteit zullen nooit optimaal zijn of in ieder geval niet voor lang. Het identificeren van de redenen waarom een markt-, systeem- of transitiefalen zich voordoet en waarom het verhelpen daarvan effectief is, vraagt om een integraal programma in plaats van een gerichte individuele interventie. Het op deze manier reconstrueren van de beleidstheorie komt daarmee deels neer op het achterhalen (en beoordelen) van de logica van de gekozen beleidsopzet. Het moet vanuit de interventielogica bezien voor de hand liggen om meerdere prikkels te combineren in één programma en deze prikkels dienen elkaar te versterken, waardoor op systeemniveau gecoördineerd moet worden.

B. Governance-processen/-structuren en de beleidsmix

De wetenschappelijke literatuur biedt handvatten om de kwaliteit van de ingerichte governance-processen en -structuren te beoordelen en de mate waarin die effectief zijn voor het inrichten van een beleidsmix. Voor de governance en inrichting geldt dat consistentie en coherentie sleutelwoorden zijn: er mogen geen interne tegenstrijdigheden bestaan in hoe het beleid is opgezet en wat het nastreeft. Een oordeel over welke governance in de praktijk effectief is, is een empirische vraag die in de wetenschappelijke

Figuur 4.2. Grafische weergave van evaluatiesystematiek



literatuur op kwalitatieve wijze wordt beantwoord. Daarvoor moet vaak eerst vastgesteld worden aan welke principes de vormgeving van het beleid moet voldoen om überhaupt effectief te kunnen zijn. Zo hebben Grillitsch et al. (2019) de Zweedse strategische innovatieprogramma's onderzocht door te kijken of de governance in lijn is met een viertal uit de transitieliteratuur verkregen criteria waar het gaat om transitiefalen (zie paragraaf 3.2).

Box 4.1 bespreekt kort welk algemeen kader gehanteerd kan worden bij het evalueren van de governance van de meeste vormen van S/T-beleid. Dit komt neer op het nalopen van consistentie op verschillende niveaus van beleidsontwikkeling en de coherentie daartussen. Rogge en Reichardt (2016) onderscheiden drie niveaus: strategisch, thematisch en implementatieniveau. Niet alleen moeten op elk niveau onderdelen vrij zijn van tegenstrijdigheden en elkaar wederzijds ondersteunen; ook tussen niveaus moet het beleid coherent zijn.

Naast het evalueren van de governanceprocessen zelf, is het ook mogelijk om te evalueren hoe die governanceprocessen leiden tot aanpassingen in de beleidsmix. Bij S/T-beleid is reflexiviteit belangrijk: het vermogen om prikkels aan te passen op basis van monitoringsinformatie en op basis van nieuwe knelpunten die zich

kunnen voordoen. Wederom is er geen uniforme standaard voor het bepalen van 'goede' en 'slechte' beleidsmix-aanpassingen, maar zijn er wel voorbeelden. Recentelijk heeft het Fraunhofer Instituut een rapport gepubliceerd waarin evaluatoren zijn nagegaan of beleidsinstrumenten goed samenhangen met de 'impact pathways' die leidend zijn voor de missies uit de Duitse HighTech Strategy 2025. Het (maatwerk-)analysekader is hier gericht op het bepalen van consistentie, in dit geval tussen de herzieningen in instrumenten enerzijds, en visies op hoe doelen bereikt kunnen worden anderzijds.

C. Aansluiting tussen beleid(mix) en knelpunten

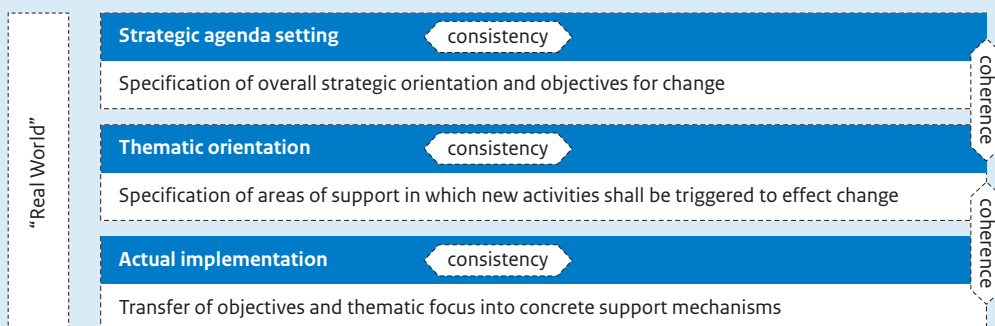
De derde invalshoek gaat over de aansluiting van de beleidsmix op knelpunten in de vorm van zwakke/ontbrekende systeemfactoren of transitieprocessen. Dit vergt een beeld van wat deze knelpunten op detailniveau zijn. Evaluatie volgens deze invalshoek dient uit te wijzen of de governance en interventies resulteren in prikkels die qua reikwijdte en dosering passen bij de knelpunten die gewenste veranderingen in de weg staan.

Van Mierlo et al. (2010) bestuderen systeemknelpunten in de glastuinbouw – zie Figuur 4.3. Zij brengen knelpunten in kaart door te kijken naar actoren en elementen in het innovatiesysteem.

Box 4.1. Handvatten voor procesevaluatie van S/T-beleid

Een recent artikel van Kroll (2019) reikt een structuur aan om systematisch na te gaan in hoeverre de governance van de beleidsimplementatie effectief verloopt.

Hij onderscheidt drie niveaus in de vertaling van beleidsambities naar beleidsimplementatie. Behalve dat activiteiten en structuren op ieder niveau consistent moeten zijn (geen interne tegenstrijdigheden), moeten ze tussen verschillende niveaus ook coherent zijn. Dit kan in het geding komen als gevolg van actor-gebaseerde uitdagingen (incompetentie, irrationaliteit), maar ook door functionele uitdagingen als het gebrek aan informatie, middelen, mandaten, en organisatorische vaardigheden.



Figuur 4.3. Voorbeeld van analyse van aansluiting tussen voorziene beleidsmix (de cirkels) en systeemknooppunten (de vierkanten). Bron: Van Mierlo et al., 2010.

	Actors	Producers & producers interest groups	Suppliers / advisors	Industry / chemical means	Knowledge institutes	Governmental bodies	Produce retailers	
Features								
Knowledge infrastructure					Knowledge development slow and underutilized			
Physical infrastructure								
Hard institutions		Covenants / agreements between stakeholders not binding				Violations not retraceable due to high costs		
Soft institutions						EU Water framework directive (KRW)		
Interaction		Lack of interactions						
Market structure		No financial stimulus to change practices					New demand for residue-poor products	

Op ongeveer dezelfde wijze is er in de evaluatie van de Topsectorenaanpak gekeken naar zwakke functies en sleutelprocessen in de technologische innovatiesystemen die corresponderen met de Topsectoren en de mate waarin geïnitieerde acties daarbij aansluiten (EZK, 2020). Een recenter voorbeeld is een onderzoek naar barrières in het missiegedreven

innovatiesysteem rondom duurzame scheepvaart (Wesseling & Meijerhof, 2021).

D. Systeemversterking/-transformatie

Er is een traditie van studies die systeemversterking onderzoekt door te kijken naar structuren en ‘sleutelprocessen’ in technologische innovatiesystemen (TIS) (Hekker et al., 2007; Bergek et al., 2008). De nadruk ligt op het begrijpen van innovatie- en diffusieprocessen en op het verrichten van diagnoses die uitwijzen op welk systeemprobleem (beleids)inspanningen zich het best kunnen toeleggen. Deze lens is ook te gebruiken om te kijken of er op een bepaalde tijdsinterval positieve ontwikkelingen hebben plaatsgevonden op de systeemaspecten waarin verbetering het hardst nodig is. Vanuit de effectladder bezien gaat deze analyse om het aantonen van correlaties of om het falsificeren dat de meest noodzakelijke verbeteringen hebben plaatsgevonden.

Hoewel er in deze literatuur sporadisch over evaluatie wordt gesproken, zijn er voorbeelden van studies die deze opzet kennen. Zie bijvoorbeeld de in Box 3.2 behandelde evaluatie van de Topsectorenaanpak, en de recente evaluatie van het vergelijkbare Australische Industrial Growth Centre Initiative. Ook de bij invalshoek B besproken Zweedse evaluatie is in de evaluaties gebruikgemaakt van het TIS-kader, zoals Figuur 4.4 ook laat zien (Åström et al., 2022). Hetzelfde geldt voor de evaluaties van het NWO Take-off programma en de Nederlandse variant van de SBIR; ook daarin is aan de hand van de TIS-sleutelprocessen gekeken naar de systeemeffecten van het ingezette beleid (Dialogic, 2017).

E. Structurele verandering

Beleid verandert de werking van een systeem. De vraag is wat een S/T-programma hiermee probeert te bewerkstelligen. Dit is bijvoorbeeld het creëren van beleidsondersteuning ter versterking van innovatiesystemen waar specifieke innovatiepaden bij gebaat zijn. Voor effectmeting met het oog op structurele verandering is van belang of die systemen bijdragen aan het succes van een selectie van innovatiepaden. Figuur 4.5 toont hoe een evaluatie-opzet eruit zou kunnen zien. Als het beleidsdoel is om meer focus en versnelling aan te brengen in specifieke innovatiepaden dan dient effectmeting zich toe te spitsen op de volgende vragen:

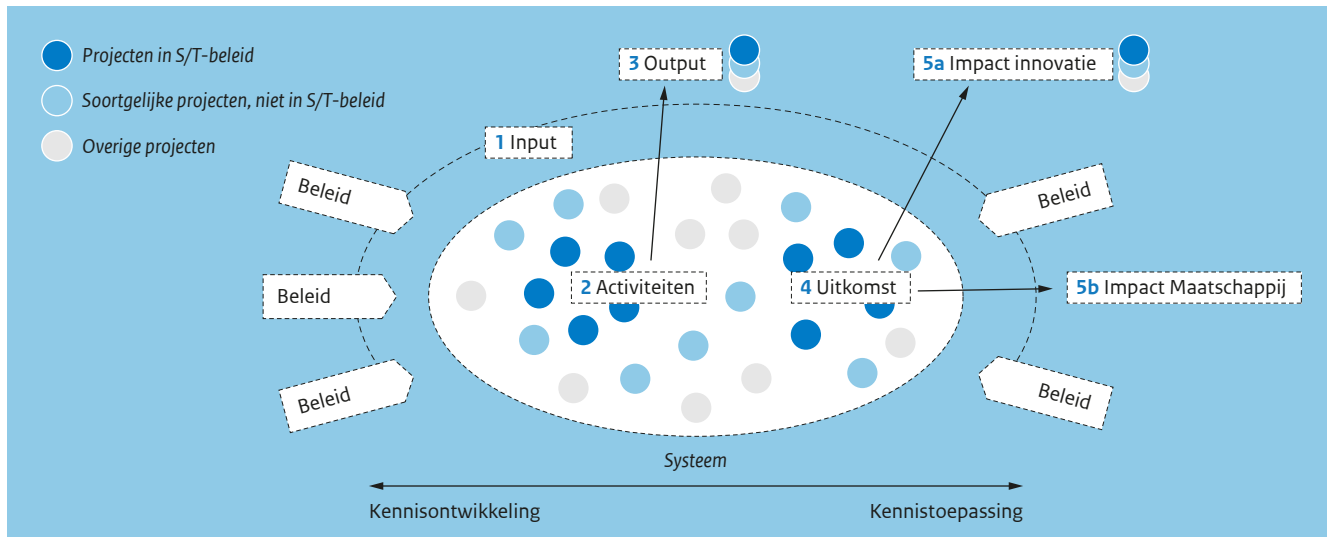
1. Input: Is er meer thematische en organisatorische en administratieve samenhang tussen potentieel complementaire beleidsinstrumenten en -inspanningen?
2. Activiteiten: Is er sprake van thematische convergentie van de door diverse regelingen gesteunde innovatieprojecten (in relatie tot de geprioriteerde thema’s)?
3. Output: Ontwikkelen de aan geprioriteerde thema’s gerelateerde projecten zich sneller of succesvoller (in termen van bijvoorbeeld patenten, publicaties, prototypes) dan andere innovatieprojecten?

Figuur 4.4. Voorbeeld van analyse van de mate waarin TIS-sleutelprocessen door beleid versterkt zijn.

Bron: Arnold (2022)

Function	Current practice	Potential with SIP 1.0
TIS functions	(Scale: 1 = low, 5 = high)	
Entrepreneurial experimentation with new technologies, markets and business opportunities	2	3
Knowledge development, via R&D and learning-by-doing	3	4
Knowledge diffusion through networks	3	4
Directionality	2	3
Market formation by opening market space or articulating demand	2	2
Market formation by creating protected space for niche innovations	1	1
Legitimation	3	4
Resource mobilisation	1	1
Developing positive externalities	2	2
Other change agency functions		
Creating arenas for priority setting	2	3
Building actor networks or coalitions	3	4
Developing guiding visions	1	4
Action at the political and policy levels	1	4
‘Creative destruction’, phase-out management	1	1
Reflexivity	2	4

Figuur 4.5. Structuur voor effectmeting op diverse doelordes (1-5), behorend bij de S/T-beleid om innovatie te richten



- 4. Uitkomst: Zien we dat er ook *buiten* de regelingen meer innovatie-activiteiten ontstaan die zich toeleggen op de toepassing van oplossingen zoals gevonden *in* de regelingen?
- 5a. Impact Innovatie: Gaan de onder 4. genoemde innovatie-activiteiten ook gepaard met betere economische prestaties van de organisaties die deze activiteiten ondernemen?
- 5b. Impact Maatschappij: Worden de door het S/T-beleid gesteunde innovaties ook relatief sneller toegepast en opgeschaald?

Uitgangspunt bij deze vragen is steeds dat het verhelpen van een falen, zoals een gebrek aan richting, zich primair zou moeten uiten in het creëren van meer momentum voor specifieke innovatierichtingen (geoperationaliseerd als meer samenhang in innovatie-inspanningen), vanwege de verbindende coördinatie-activiteiten en de toegespitste beleidsprirrels. Een vraag is of er binnen de regelingen meer aandacht komt voor innovatierichtingen en of de beleidsprirrels en innovatieprojecten zoveel in gang zetten (systeemversterking bewerkstelligen) dat organisaties aanhaken en erin investeren.

Een manier om dit proces te observeren is door te kijken naar de mate waarin bedrijfsspecifieke innovatieprojecten, in bijvoorbeeld de WBSO-regeling, zich naar die richtingen voegen. Aangezien de WBSO een relatief generieke innovatieregeling is, zonder inhoudelijke sturing, zou het effect van S/T-beleid kunnen blijken uit een geobserveerde toegenomen bereidheid van bedrijven om R&D te verrichten op het vlak van producten en processen die passen bij de beleidsmatig gesteunde innovatiepaden. Is een dergelijke toename niet waarneembaar, dan is het niet waarschijnlijk dat het beleid effectief is. Wanneer de toename wel waarneembaar is dient onderzocht te worden wat de bijdrage van beleid is. In het verlengde hiervan moet worden onderzocht

in hoeverre complementaire bedrijfseconomische prestaties van belang zijn (zie effect 5a Impact innovatie).

Een vereiste voor deze evaluatieaanpak is dat duidelijk moet zijn wat de oplossingsrichtingen zijn en hoe innovatieprojecten binnen de daaraan gelieerde regelingen en vooral buiten die regelingen zich tot die richtingen verhouden. Voor de vergelijking dient het labelen van projecten nauwgezet te gebeuren, omdat labelen het mogelijk maakt te kijken of de gekozen innovatierichtingen zich sneller en beter ontwikkelen dan de niet-gekozen richtingen. Veelbelovend op dit punt is het uitgangspunt van S/T-beleid om duidelijke richtingen leidend te laten zijn voor het bundelen van inspanningen. Bij evaluatie moet blijken in hoeverre deze richtingen duidelijk zijn en om welke richtingen het gaat.

Een beperking van deze analyseopzet is dat het voorbijgaat aan het feit dat niet voor alle S/T-programma's de relevante oplossingen het gevolg zijn van systeemverandering of van transities. Daarnaast wordt in de analyses niet expliciet gekeken naar veranderingen in de vraag naar nieuwe producten/processen/diensten. Een groter aantal innovatieprojecten is nog geen garantie dat die innovaties omarmd worden en effectief zijn om de maatschappelijke doelen te bereiken.

F. Maatschappelijke impact

De laatste invalshoek voor evaluatie omvat het onderzoeken van ontwikkelingen op de doelvariabelen waar het vanuit het S/T-beleid uiteindelijk allemaal om te doen is.

Veranderingen in doelindicatoren worden vaak gebruikt om uitspraken te doen over doelmatigheid. Hierbij kijkt men bijvoorbeeld naar de hoeveelheid CO₂(equivalent)-reductie per euro beleidskosten. Om werkelijk iets te zeggen over doeltreffendheid en doelmatigheid van beleid is het echter

zaak om te weten in hoeverre de prestaties veroorzaakt worden door de verstrekte impulsen. Het kan ook zo zijn dat exogene ontwikkelingen hiervoor verantwoordelijk zijn.¹⁴ Andersom is het ook mogelijk dat het beleid qua logica, governance, beleidsmix en zelfs systeemveranderingen goed in elkaar steekt, maar dat onvoorziene exogene ontwikkelingen er juist voor zorgen dat dit zich toch niet vertaalt in gewenste prestaties. Zoals eerder beschreven is het vaststellen van causaliteit bij S/T-beleid ingewikkeld vanwege de veelheid aan synergetische beleidsprikkelers die ook ieder afzonderlijk een effect hebben en vanwege andere ontwikkelingen die samen van invloed zijn op het beoogde doel. Een tweede probleem is dat het effect mogelijk niet lineair is in de beleidskosten en over de tijd. Bij het beoordelen van prestaties op doelvariabelen is het daarom van belang om aandacht te besteden aan het verwachte verloop, waarbij beleid zelf ook rekening moet houden met onzekerheid. Prestaties kunnen zich bijvoorbeeld aanvankelijk positief ontwikkelen, maar daarna stagneren. Beleid werpt dan op korte termijn vruchten af, maar is waarschijnlijk op langere termijn minder doeltreffend en doelmatig.¹⁵ Het omgekeerde kan ook het geval zijn; op korte termijn ziet men nog geen direct resultaat maar op de lange termijn wel.

Samengevat is het bij het evalueren van maatschappelijke impact zaak om op betrouwbare wijze vast te stellen in hoeverre ontwikkelingen van gemonitorde doelindicatoren het gevolg zijn van ingezet beleid. Indien dat niet mogelijk is, kunnen geobserveerde prestaties soms vergeleken worden met situaties elders. Bij dergelijke vergelijkingen dienen evaluatoren – als een onderliggende causale analyse niet mogelijk is – voorzichtig te zijn met het toewijzen van geobserveerde verschillen aan het gevoerde beleid.

4.5. Selectie en toepassing van de evaluatiekader-invalshoeken

Afwegingen ten aanzien van invalshoek-keuze

Iedere invalshoek uit het voorgestelde evaluatiekader belicht een ander aspect van de werking en doeltreffendheid van S/T-beleid. Daarbij lenen de 'hogere' invalshoeken (D-F) zich het best voor summatieve analyses met het oog op verantwoording over bewerkstelligde veranderingen, terwijl de nadruk bij de 'lagere' invalshoeken A-C ligt op het formatief onderzoeken van de onderbouwing en uitvoering van het beleid. Strikt genomen behoort die laatste categorie van invalshoeken doorgaans niet tot het domein van effectmetingen. Zoals beschreven in

hoofdstuk 3 is het bij S/T-beleid soms beperkt mogelijk om het effect te isoleren van programma's die middels betere coördinatie en beleidsafstemming een impuls willen geven aan systemen die uiteindelijk van belang zijn voor het bereiken van een maatschappelijk wenselijk doel. In die gevallen wordt het relevanter om uitspraken te kunnen doen over successen en mislukkingen in het organiseren van beleid dat op zijn minst wezenlijke systeem-/transitieknelpunten weet te adresseren (ook al kunnen we niet goed causaal vaststellen hoe dat vervolgens uitpakt). Dit geldt des te meer als S/T-beleid zelf ook maar weinig duidelijk aanwijsbare prikkels verstrekt aan 'het veld', en in plaats daarvan vooral dient om de aansluiting tussen bestaande (complementaire) beleidsinitiatieven te versterken. In hoeverre het beleid doeltreffend is, is dan te onderzoeken door te kijken naar verbeteringen in de samenhang van beleidsinitiatieven. Is die er, dan is met invalshoeken D-F ook de uitwerking van die samenhang te bepalen. Is de samenhang er bij voorbaat niet, dan wordt het al direct minder relevant om te kijken naar de doorwerking.

Bovenstaande logica staat of valt bij de mate waarin er principes zijn waar 'samenhangend beleid' aan moet voldoen. In theorie kan het voorkomen dat het verbeteren van de aansluiting tussen beleidsinitiatieven mislukt, maar dat er om onverwachte redenen tóch vooruitgang wordt geboekt op een transitie. Deze mogelijkheid onderstreept niet alleen het devies om in ieder geval te proberen om invalshoeken D-F toe te passen, maar ook het belang van vervolgonderzoek naar goed functionerende governanceprocessen en beleidsmixen: des te meer we weten over hoe coördinatie van S/T-beleid dient te verlopen, des te beter we concrete casussen daarop kunnen toetsen.

In algemene zin is het bij S/T-beleid wenselijk om zoveel mogelijk invalshoeken te evalueren, aangezien juist uit de consistentie tussen deel-analyses kan worden afgeleid tot op welke hoogte beleid doeltreffend is: zijn er bijvoorbeeld al problemen bij het versterken van een beleidsmix, gaat het veranderen van het systeem naar behoren, en loopt het (mogelijk vanwege exogene factoren) spaak bij het bewerkstelligen van structuurveranderingen?¹⁶ Bevindingen uit verschillende deel-analyses kunnen aanleiding vormen voor gerichte verdiepingen: mocht systeemverandering op een bepaald aspect excelleren of juist achterblijven, is dit dan te relateren aan eigenschappen van de beleidsopzet en -uitvoering?

Uiteraard is het om praktische redenen niet altijd mogelijk om alle invalshoeken te bestrijken in één evaluatiestudie. Als er

¹⁴ Het is denkbaar dat beleid gericht op maatschappelijke uitdagingen soms aanhaakt bij ontwikkelingen die al een zeker momentum hebben, en die daardoor ook de maatschappelijke druk leveren waardoor het beleid tot stand komt. In dat geval is er na enkele jaren wellicht correlatie tussen beleidsinvoering en prestaties op doelindicatoren, zonder dat dit laatste ook een beleidsgevolg hoeft te zijn. Het is ook mogelijk dat het beleid een negatieve invloed heeft, bijvoorbeeld wanneer nieuw opgelegde coördinatiestructuren bestaande coördinatie-mechanismen in de weg gaan zitten. Dit scenario onderstreept het belang van een scherpe analyse van de interventie-rationale, aangezien die moet blootleggen of er werkelijk problemen zijn die beleidsinterventie behoeven (zie invalshoek A).

¹⁵ Voor een voorbeeld, zie het PBL-rapport "Grote opgaven in een beperkte ruimte" (april 2021), waarin wordt gewaarschuwd dat honderden miljoenen investeringen voor aanpassingen in de landbouw op middellange termijn weer overboord kunnen.

¹⁶ Deze logica om te traceren tot waar in de beleidstheorie er wenselijke veranderingen zijn komt overeen met de principes van de *process tracing*-methodologie; zie daarvoor ook hoofdstuk 5.

gekozen moet worden is het zinvol om in ieder geval rekening te houden met twee zaken. De eerste is de aard van het beleid, waarvoor geldt dat programma's met een nadruk op coördinatie en beleidsmix-afstemming (in plaats van nieuwe prikkels verstrekken) relatief meer gebaat zijn bij inzichten op basis van invalshoeken A-C. Het beleidsdoel bij een programma als het Missiegedreven Topsectoren- en Innovatiebeleid is dat er meer synergie komt in uiteenlopende middelen en initiatieven die benut kunnen worden voor het bijdragen aan maatschappelijke uitdagingen. De eerste vraag om te beantwoorden is of er inderdaad meer momentum ontstaat, wat onder andere kan blijken uit consistentie in de criteria van diverse maatregelen, focus in de onderwerpen waar actoren hun eigen investeringen op richten, of netwerkvorming rondom die onderwerpen (zie volgende paragraaf). Het onderzoeken van het vermogen om beleid aan te passen op problemen is belangrijker naarmate die problemen verschuiven, wat zeker bij transities vaak het geval kan zijn. Op het moment dat S/T-beleid zelf ook concrete prikkels verstrekt, zoals bij het klimaatbeleid het geval is, wordt het relevanter om op basis van invalshoeken D-F ook goed te kijken naar de invloed van die prikkels. De tweede factor om rekening mee te houden is de fase van de beleidscyclus. Als er op grond van de Regeling Periodiek Evalueren elke vijf jaar verantwoordend onderzoek dient plaats te vinden wijst dat automatisch naar invalshoeken D-F, terwijl het voor invalshoeken A-C voor de hand ligt om op continue basis te monitoren welke beleidsaanpassingen worden ingezet en welke reacties hierop volgen. Omdat coördinatie-intensief S/T-beleid vaak berust op het bewerken van de 'marges' van een groot aantal beleidsactiviteiten en -initiatieven (bijv. beslissingen ten aanzien van onderwerpen die extra of juist minder gewicht krijgen) kan het verstandig zijn voortdurend te reflecteren op wie en wat er gemobiliseerd wordt: niet alleen omdat er dan al eerder ingegrepen kan worden, maar ook omdat het niet eenvoudig is om achteraf een beeld te construeren van de cumulatie van al die kleine aanpassingen.

Operationalisering van de invalshoeken

Om de invalshoeken uit het evaluatiekader in de praktijk toe te passen is het zaak ze te operationaliseren. Bij traditioneel innovatiebeleid is dat vaak een kwestie van het zoeken van indicatoren voor doelen als 'het vergroten van R&D-intensiteit/-samenwerking'. Zijn de indicatoren bekend, dan is de volgende stap het identificeren van een databron met variabelen waarmee de indicator gemeten kan worden.

R&D-intensiteit wordt in Nederland vaak benaderd door te kijken naar de WBSO-uren van een bedrijf (waar het CBS vertrouwelijke informatie over heeft), door gebruik te maken van de antwoorden van een bedrijf op relevante vragen uit de *R&D-enquête / Community Innovation Survey* (ook bekend bij het CBS, voor een selectie van bedrijven), of door evaluatoren een gerichte enquête uit te laten zetten.

Bij S/T-beleid is de stap naar indicatoren (laat staan databronnen) ingewikkelder. Op de eerste plaats is er een grotere variëteit in

doelen waar indicatoren bij passen, omdat de ambities kunnen gaan over systemen en transities op een groot aantal onderwerpen. Daarnaast zijn dergelijke doelen een stuk abstracter dan het vergroten van R&D-investeringen of de directe resultaten daarvan. Het begrip systeem kan betrekking hebben op innovatiesystemen, productie-consumptiesystemen (wat wordt er geproduceerd en aangeschaft, en door wie?) of op socio-technische systemen. Zeker voor die laatste interpretatie geldt dat het moeilijk is om er met data goed vat op te krijgen. De uiteindelijke transitieopgave waar S/T-beleid aan moet bijdragen is soms wel goed te monitoren (bijvoorbeeld de hoeveelheid gereduceerde CO₂-uitstoot in een bepaalde sector), maar als dit niet te attribueren is aan het beleid zal er teruggegrepen moeten worden op indicatoren voor tussendoelen. Het wordt dan essentieel om goed te begrijpen via welke mechanismen het uiteindelijke doel tot stand zou moeten komen, en wat dus relevante tussendoelen zijn om de evaluatie op te richten. Een derde en reeds uitvoerig besproken complicatie is dat die werking bij S/T-beleid vaak niet eenduidig af te leiden is uit de officiële beleidsstukken. In dat geval moet er eerst geleerd worden (over het beleid) voordat het zinvol is om te gaan meten.

Hieruit volgt dat het operationaliseren van de evaluatiekader-invalshoeken begint met het stellen van de juiste vragen. De antwoorden op die vragen zijn geen kwantitatief meetbare indicatoren, maar wel indicaties voor hoe het beleid ervoor staat. Tabel 4.2 geeft een overzicht van de enkele voorbeeldvragen per invalshoek. Daarmee is het meteen ook een samenvatting van wat de respectievelijke invalshoeken te bieden hebben.

Aangezien de voorbeeldvragen verschillende vormen van S/T-beleid dienen te bestrijken zijn ze algemeen geformuleerd. Om het geheel wat concreter te maken geven we een beknopte illustratie aan de hand van één van de casussen die ook in hoofdstuk 8 aan bod komen: het Missiegedreven Topsectoren- en Innovatiebeleid (MTIB).

- Invalshoek A: Het beoordelen van de interventie-rationale voor het MTIB vraagt om antwoord op de vraag waarom het oprichten van coördinatiestructuren als themateams en missieteams, inclusief hun mandaten inzake het realiseren van het Kennis- en Innovatieconvenant (KIC), een geschikte manier zou zijn om een netto positieve impuls te geven aan innovatie ten behoeve van maatschappelijke doelen zoals vervat in de 25 missies. Uit de wetenschappelijke literatuur is bekend dat markten er soms niet in slagen synergetische investeringen in nieuwe oplossings-/transitiepaden te realiseren, en dat verschillende overheden elkaar kunnen tegenwerken bij het ondersteunen hiervan. Op voorhand kan er dus al eenvoudig betoogd worden dat 'innovatie t.b.v. transities' om overheidsinterventie vraagt op grond van transitiefalen als *directionality failure*, *reflexivity failure* en *policy coordination failure*. De vraag is echter of dergelijke vormen van falen ook echt aan de orde zijn voor alle 25 missies, of voor de 4 missiethema's waar ze onder vallen. Mogelijk zijn er ook allerlei krachten die juist vanzelf al zorgen dat marktpartijen en andere stakeholders (inclusief overheden) vanuit hun

Tabel 4.2. **Overzicht van voorbeeldvragen per invalshoek uit het voorgestelde evaluatiekader**

Invalshoek	Voorbeelden van vragen
A: Interventie-rationale	Wat zijn de vormen van falen die vragen om S/T-beleid (dus met focus op coördinatie en stroomlijnen)? Waar precies zitten die knelpunten, en wat is het bewijs daarvoor? Bijvoorbeeld: zijn er zonder het beleid onvoldoende (beleids) mogelijkheden om bij te dragen aan systeemverandering en/of transities, en wat verklaart waarom de markt en samenwerkende actoren er zelf niet uitkomen?
B: Governance-proces en beleidsmix	Zijn er principes die aangeven waar de coördinatie vanuit het S/T-beleid op dient toe te zien, en voldoet de coördinatie aan die principes? Is er commitment vanuit relevante partijen? Leidt de coördinatie tot passende aanpassingen in de beleidsmix (dat zijn wijzigingen waardoor systeemveranderingen en daaruit voortkomende activiteiten beter aansluiten bij beleidsdoelen)? In welke mate is dit terug te zien in het ontwikkelen of uitfasen van regelingen en in aanpassingen in de financiering en vormgeving van regelingen?
C: Beleid vs. knelpunten	Wat zijn de knelpunten (in termen van systeemprocessen of structuren) die systeemverandering ingewikkeld maken of transitieprocessen stroever laten verlopen, en hoe wordt hierop ingespeeld door S/T-programma's en de beleidsinstrumenten waar dat beleid op aangrijpt? Gaat de meeste energie naar de grootste bottlenecks, en worden er resultaten geboekt op het versterken van de zwakste systeem-aspecten?
D: Systeemverandering	Zien we verbeteringen in de systeemprocessen die ontwikkeling en verspreiding van innovaties mogelijk maken? Is er meer kennisontwikkeling/-uitwisseling, marktformatie etc. (afhankelijk van welk proces zwak was) rondom wenselijke innovaties?
E: Structuurverandering	Veranderen innovatiesystemen / socio-technische / productie-consumptiesystemen structureel? Het gaat hierbij om omvorming van hele systemen, ook de delen die buiten de nauwe scope van de aan het S/T-beleid verbonden regelingen vallen. Een eerste indicatie zou zijn of er daar meer innovatie-activiteit ontstaat.
F: Maatschappelijke impact	Worden de tussentijdse en uiteindelijke doelen gehaald? En in welke mate is dit het gevolg van (systeem)veranderingen die toe te schrijven zijn aan de invloed van het S/T-beleid? Bijvoorbeeld: als er sectoren zijn waar CO2-besparing is gerealiseerd, berust dat dan ook op innovaties waar het beleid aantoonbaar een impuls aan heeft gegeven?

eigen overwegingen al gedrag vertonen dat bijdraagt aan een missiedoel als “Het vak van veiligheidsprofessional behoort in 2030 tot de top 10 van meest aantrekkelijke beroepen in Nederland.”. In die zin is het relevant om voor ieder van de missiedoelen te verifiëren (bijv. met behulp van deskstudie en interviews met beleidsmakers, domein-actoren en domein-experts) wat precies de gronden zijn om aan te nemen dat S/T-beleid zoals het MTIB een geijkte beleidsstrategie is. Behalve het nalopen van mogelijk transitiefalen is het zaak daarbij ook te kijken naar onderliggend markt- en systeemfalen: als het waar lijkt dat er niet automatisch al convergentie ontstaat rondom een maatschappelijk wenselijke oplossingsrichting, is de vervolgvraag immers waaróm dat zo is. De antwoorden hierop zouden voor de 25 missies behoorlijk kunnen verschillen, wat illustreert hoe lastig het kan zijn om over de algehele kwaliteit (of bij andere invalshoeken: het effect) van S/T-beleid te spreken.

- Invalshoek B: Gezien het karakter van het MTIB - momentum creëren door een beleidsmix te stroomlijnen en uiteenlopende actoren te mobiliseren/verenigen - ligt het voor de hand om uitvoerig aandacht te besteden aan de inrichting van governance-processen en -structuren. Aangezien het concretiseren hiervan de nodige kennis van het MTIB vereist is het lijstje met relevante voorbeeldvragen opgenomen in de casusuitwerking in sectie 8.1. Die vragen zijn verbijzonderingen van de voorbeelden

uit Tabel 4.2, met als gemene deler de aandacht voor de thematische samenhang in beleid, agenda's en innovatie-/diffusie-inspanningen die het MTIB teweeg moet brengen. Aanvullend kan er ook geanalyseerd worden in hoeverre het MTIB werkelijk leeft onder de partijen die in de governance-structuren vertegenwoordigd zijn, en die daar namens een achterban bijdragen aan het blootleggen van knelpunten en het oppakken daarvan. Dit kan bijvoorbeeld blijken uit *commitment* in de vorm van tijd en middelen die partijen investeren in het participeren in governance-activiteiten. Of de *captains of industry/science* (etc.) blijven deelnemen aan de overleggen van thema-teams, of niet, zegt iets over de waarschijnlijkheid dat het MTIB bijdraagt aan het mobiliseren en afstemmen van de acties van relevante stakeholders.

- Invalshoek C&D: Om te bepalen of de coördinatieactiviteiten vanuit het MTIB ook de juiste prioriteiten stellen is het zaak deze te koppelen aan de systeemknelpunten die zich voordoen bij de thema's die centraal staan in de MTIB-programmering. Het kennen van die knelpunten is tevens essentieel voor het bestuderen van systeemveranderingen. Voortbouwend op de evaluatie van de Topsectoren-aanpak kan er in beide gevallen gekozen worden voor gebruik van de dimensies en indicatoren behorende bij het raamwerk van *Technological Innovation Systems* (TIS). Zoals in sectie 4.4 besproken onderscheidt dit raamwerk enkele sleutelprocessen of 'functies' die van belang zijn voor

het goed functioneren van een op een specifieke technologie gerichte innovatiesysteem. Het lastige aan maatschappelijke missiedoelen is echter dat er niet op voorhand al één technologie aan te wijzen is, en dat het behalen van die doelen ook om andersoortige interventie vraagt. Om beter aan te sluiten bij een dergelijke context, die kenmerkend is voor S/T-beleid in bredere zin, wordt er momenteel gewerkt aan de conceptualisering en operationalisatie van ‘missiegedreven innovatiesystemen’ (MIS) – zie Elzinga et al. (2020). Tabel 4.3 op de volgende pagina toont de MIS-functies en enkele voorbeelden van onderliggende indicatoren. De uitdaging is in de praktijk vooral om goed af te kunnen bakenen voor welk systeem/thema die indicatoren gemeten moeten worden.

- Invalshoek E: De invalshoek over structuurveranderingen kent geen algemene conceptuele basis. In het geval van het MTIB ligt het voor de hand om in ieder geval te kijken naar veranderingen in samenwerkingsstructuren, oftewel R&D- en innovatienetwerken. Voor de originele Topsectorenaanpak is na afloop van de formele evaluatie een studie verricht om te bepalen of die netwerken er anders uit zijn gaan zien nadat de Topsectoren en TKI’s hun intrede deden. – zie Janssen et al. (2020). Hieruit bleek onder andere dat partijen over grotere geografische afstanden zijn gaan samenwerken, en dat partijen die tot het ecosysteem van eenzelfde TKI behoren over grotere cognitieve afstand zijn gaan samenwerken (er wordt kennis gecombineerd die normaliter minder vaak gecombineerd werd). Op eenzelfde manier kan nu ook bekeken worden of de introductie van het MTIB weer veranderingen teweeg heeft gebracht. Een eerste analyse op basis van publiek-private R&D-projecten laat zien dat partijen die samenwerken aan een project op het gebied van de missiethema’s Energietransitie & Duurzaamheid of Gezondheid & Zorg over het algemeen minder cognitieve afstanden overbruggen in vergelijking met andere publiek-private R&D-projecten – zie Janssen en Abbasiharofteh (2022). De vraag is hoe dit zich in de loop der jaren verder ontwikkelt, en of meer en diverser wordt samengewerkt op de oplossingsrichtingen die in de MTIB-programmering centraal staan (in vergelijking met soortgelijke maar niet-uitverkoren richtingen).
- Invalshoek F: Uiteindelijk beoogt het MTIB bij te dragen aan het behalen van de missiedoelen. Het meest directe beleidsdoel is echter dat er dankzij het MTIB een versnelling optreedt in de ontwikkeling en toepassing van innovaties die aan die doelen zouden kunnen bijdragen. Dit is af te meten aan de mate waarin private gedragingen ver buiten het MTIB zich ook voegen naar de prioriteiten uit de MTIB-programmering. Als het gaat om innovatie kan hiervoor op geautomatiseerde wijze gekeken worden naar de omschrijvingen van de WBSO-projecten van individuele bedrijven, maar wellicht zijn er per missie ook andere geschikte databronnen. Zie wederom de casusuitwerking in hoofdstuk 8 voor meer detail op dit punt.

Tabel 4.3. Voorbeelden van indicatoren voor het analyseren van missiegedreven innovatiesystemen (Elzinga et al., 2020).

Functie	Voorbeelden van indicatoren
Aanscherpen van probleem-focus	Zichtbaarheid van de uitdaging in media Zichtbaarheid van de uitdaging in bedrijfsstrategie Zichtbaarheid van de uitdaging in strategie brancheorganisaties Zichtbaarheid van de strategie in onderzoeksprogramma's Mate waarin uitdaging op de agenda staat van relevante organisaties (prioriteit) Specificiteit van de beschrijvingen van de uitdaging in strategische documenten Stimulerende wet- en regelgeving ten aanzien van generieke uitdaging
Aanscherpen van oplossings-focus	Benoemen van verwachtingen rondom specifieke oplossingen Maatschappelijke discourse in media rond specifieke oplossingen Stimulerende wet- en regelgeving voor specifieke oplossing
Coördinatie	Vorming van coalities rond de transitie Ontstaan van leidende organisaties die het veld samenbrengen Strategische documenten die verschillende oplossingsrichtingen vergelijken / analyseren Ontstaan van coherente visies op de transitie
Experimenteren door ondernemingen	Aantal (nieuwe) ondernemers die experimenteren met relevante oplossingen Aantal projecten die zijn gestart/gestopt Aantal bedrijven die zijn gestart/gestopt
Kennisontwikkeling	Aantal (wetenschappelijke en professionele) publicaties Aantal kennisprojecten gestart/gestopt Impact van publicaties (citaties, gebruik in samenleving) Aantal nieuwe onderzoeksgroepen Aantal nieuwe onderzoeksprogramma's Aantal nieuwe leerstoelen Aantal nieuwe patenten
Kennisverspreiding	Aantal symposia en conferenties Aantal kennisnetwerken Aantal gemeenschappelijke publicaties Aantal kennis-delende activiteiten tussen verschillende actoren (bedrijven, consumenten, kennisinstellingen, overheden)
Marktcreatie	Bedrijfsactiviteiten die consument verleiden tot adoptie Regelgeving die verplicht om oplossingen toe te passen Fiscale of subsidie-instrumenten die oplossingen financieel aantrekkelijker maken Standaarden / labels die oplossingen onderscheidend maken en adoptie stimuleren / verplichten
Mobiliseren van middelen	Beschikbaar stellen van financiële middelen voor innovatie, onderzoek, pilots, investeringen door overheden, bedrijven en andere organisaties Opleiden en beschikbaar maken van personeel met juiste kennis en capaciteit Beschikbaar stellen van benodigde infrastructuur en grondstoffen
Legitimiteit creëren	Lobby en media uitlatingen die dienen om angst / weerstand rond nieuwe trajecten te verminderen Balans in lobbyactiviteiten voor en tegen de missie met bijbehorende oplossingen
Druk zetten op het huidige regime	Regelgeving die stelt dat bestaande activiteiten afgebouwd dienen te worden Afschaffen van subsidies / voordelen voor bestaande activiteiten Vrijwillige afspraken met marktpartijen om van koers te veranderen Maatschappelijke discourse / activisme tegen huidige praktijken Acties gericht op benoemen 'verkeerd' consumentengedrag

5. Methoden voor het evalueren van systeem- en transitiebeleid

Op het moment dat het evaluatiekader helpt bepalen welke aspecten van beleid onderzocht dienen te worden, is de volgende vraag welke onderzoeksmethoden het best passen. Paragraaf 5.1 geeft een overzicht van verschillende opties en paragraaf 5.2 vervolgt met een bespreking van de inzetbaarheid voor het evalueren van S/T-beleid.

5.1. Overzicht van methoden

Deze sectie biedt een overzicht van empirische onderzoeksmethoden. Hierbij is gebruikgemaakt van de inventarisatie en beschrijving van onderzoeksmethoden in de online Toolbox Beleidsevaluaties. Monitoring is niet opgenomen als methode, aangezien het registreren van ontwikkelingen op relevante indicatoren dient om andere (meer evaluatieve) methoden te

voeden. Monitoring van beleid is een noodzakelijke voorwaarde en een manier om relevante data voor evaluatie te verzamelen.

Doeltreffendheid

Tabel 5.1 geeft weer welke methoden het meest geschikt lijken voor het onderzoeken van de doeltreffendheid van S/T-beleid.

- De methoden zijn te categoriseren op basis van *het soort empirische data* dat benodigd is voor toepassing. In twee gevallen zijn de methoden vooral theoretisch¹⁷ van aard: systematische reviews & meta-analyses en simulaties. Meta-analyses en simulaties zijn kwantitatief, terwijl systematische reviews kwalitatief zijn. Vervolgens zijn er uiteenlopende kwalitatieve en kwantitatieve methoden die allemaal een andere toepassing kennen.

Tabel 5.1. Kenmerken van onderzoeksmethoden met relevantie voor het evalueren van doeltreffendheid

Methode	Data			Analyse-doel		
	Theoretisch	Kwalitatief	Kwantitatief	Werkings-mechanismen identificeren	Causaliteit bepalen	Effect-omvang meten
Systematische reviews & meta-analyse	X		X	X		
Simulaties	X			X		
Lerend evalueren		X		X		
Outcome harvesting		X		X		
Case studies		X		X	X	
Qualitative Comparative Analysis (QCA)		X		X		
Contribution analysis		X	X	X		
Process tracing		X	X		X	
Regressieanalyse			X			X
(Quasi-)experimentele methoden			X		X	X

¹⁷ Met theoretisch wordt bedoeld dat deze analyses niet uitgaan van data die in het te evalueren systeem zelf worden gegenereerd maar modelmatig of op basis van analyses in een andere context.

- Om dit inzichtelijk te maken toont de tabel ook voor welke *analysedoelen* de methoden zich lenen. Voor doeltreffendheidsevaluaties zijn in ieder geval drie analysedoelen belangrijk: het verklaren van mechanismen (de manieren waarop beleid en eventueel andere factoren/condities bijdragen aan het behalen van doelen), het bepalen van causaliteit (in hoeverre het is vast te stellen of en wanneer A tot B leidt), en het meten van de (meestal gemiddelde) effectomvang van een beleidsinterventie.

Ten opzichte van simulaties kennen systematische reviews en meta-analyses een ander analysedoel. Het raadplegen van bestaande literatuur kan helpen om beter te begrijpen wat de (normatieve) uitgangspunten zijn van hoe beleid eruit zou moeten zien. Daarnaast is het analyseren van bestaande studies een manier om aannemelijk te maken in hoeverre beleidseffecten optreden, mits er natuurlijk eerdere studies zijn naar hetzelfde soort beleid en dezelfde beleidsdoelen. De causaliteit van S/T-beleid wordt hiermee niet getoetst, maar naarmate er meer relevant eerder onderzoek is kan wel beter ingeschat en beredeneerd worden hoe waarschijnlijk het is dat het voorliggende S/T-beleid een causaal effect heeft. Simulaties lenen zich daarentegen beter om mechanismen te verkennen, in het bijzonder als het gaat om het in elkaar grijpen van gedragingen die afzonderlijk bekend zijn maar niet in samenhang.

Er is ook een aantal kwalitatieve methoden dat kan worden ingezet voor het bepalen van doeltreffendheid. Deze kwalitatieve methoden verschillen op een aantal punten in aard en toepasbaarheid:

- *Lerend evalueren* is een manier om, samen met beleidsmakers en stakeholders, te onderzoeken hoe beleid zich verhoudt tot (vaak complexe) maatschappelijke vraagstukken. Daarbij gaat het in de eerste plaats om het verkrijgen van een beter begrip van de werking die is toebedacht aan gekozen governancestructuren, beleidsactoren, en beleidsinterventies, om zo te (re)construeren hoe het beleid werkt en welke knelpunten overblijven of opdoemen. Deze methode helpt om te verklaren via welke mechanismen beleid kan bijdragen aan het bereiken van het beoogde doel, maar zegt niet of dat ook gebeurt (of in welke mate daar sprake van is);
- *Outcome harvesting* is (eveneens) een methode om werkingsmechanismen in kaart te brengen. Ten opzichte van lerend evalueren staat niet het beleid zelf, maar de uitkomst van het beleid centraal. Door vanuit de uitkomst van dat beleid terug de (causale) keten die is gevolgd in te gaan, oftewel door te onderzoeken welke factoren verklaren hoe een geobserveerde uitkomst tot stand komt, kan achterhaald worden wat de relatieve rol van beleid is ten opzichte van andere factoren die voor de uitkomst van belang zijn;
- *Casestudies* zijn een combinatie van een onderzoeksmethode en een dataverzamelmethode en kunnen voor vele doeleinden ingezet worden. Door voor één of enkele casussen (projecten,

missies of transities) na te gaan wat de rol van het gevoerde beleid is, kan een beter beeld ontstaan van de mechanismen die uitkomsten tot stand brengen. Voor de onderzochte casussen is soms ook te achterhalen of sprake is van een causaal verband tussen beleid en uitkomst, maar dit effect kan in casestudies vaak niet kwantitatief gemeten worden en gegeneraliseerd worden naar andere casussen (gebrek aan externe validiteit);

- *Qualitative Comparative Analysis (QCA)* is een methode om middels een systematische vergelijking van een aantal casussen te bepalen welke (combinatie) van factoren geassocieerd is aan een bepaalde uitkomst. De methode vergt dat de casussen variëteit kennen in de prestaties op die uitkomst en dat voor alle casussen bekend is of bepaalde condities wel of niet aanwezig zijn. Bij die condities kan de aan- of afwezigheid van (een bepaald aspect) van beleid horen, maar het is ook mogelijk dat beleid aanwezig is bij alle onderzochte casussen. Een conditie die niet varieert over casussen, zal niet worden meegenomen in de analyse (want heeft geen verklarende waarde om verschillen tussen de casussen te duiden). Wanneer er wel variatie is, helpt de QCA-methode om een correlatie te vinden tussen een configuratie van condities en het al dan niet behalen van een beleidsdoel, maar is het niet mogelijk om iets te zeggen over de rol van het beleid zelf;
- *Contribution analysis & process tracing* zijn methoden om causale relaties in beeld te brengen. Net als bij lerend evalueren vormt bestaand beleid het vertrekpunt van de effectmeting. Bij beide methoden bestaat de essentie uit het verzamelen van kwalitatieve en/of kwantitatieve gegevens waarmee de bestaande beleidstheorie wordt afgelepen om vast te stellen of veronderstelde veranderingen en doorwerkingen aanwezig zijn. Dit geeft geen zicht op de kwantitatieve omvang van een causaal effect, maar legt wel bloot in hoeverre er ontwikkelingen zijn die consistent zijn met wat de beleidstheorie nastreeft. In tegenstelling tot contribution analysis wordt er bij process tracing gewerkt met het toetsen van hypothesen die gebaseerd zijn op de beleidstheorie.

Het voornaamste verschil tussen de twee kwantitatieve methoden in Tabel 5.1 is dat regressieanalyses primair in te zetten zijn om correlaties tussen beleidsimpuls en doelbereik te meten (en daarbij zo goed mogelijk corrigeren voor overige factoren die ook van invloed zijn op de doelvariabelen), terwijl (quasi-)experimentele methoden gekenmerkt worden door hun potentie om causaliteit zo betrouwbaar mogelijk vast te stellen.

De uitdaging bij doeltreffendheidsevaluaties is doorgaans om conclusies te trekken die relevant zijn ten aanzien van de uiteindelijke beleidsdoelen die worden nagestreefd, én die een valide beeld geven van causale verbanden. Vaak is het lastig om bij de evaluatie van S/T-beleid beide ambities te realiseren met één enkele methode en is het effectief meerdere methoden te combineren. In dergelijke ‘mixed methods’-analyses kunnen de diverse methoden zich op verschillende manieren tot elkaar verhouden, wat ook bepaalt of ze gelijktijdig of volgorde-lijk

moeten worden ingezet. Zo kunnen methoden gecombineerd worden om observaties te trianguleren om de betrouwbaarheid van bevindingen te bestendigen. Het is ook mogelijk via formatieve methoden eerst zicht te krijgen op werkingsmechanismen en op basis daarvan een summatieve effectmeting te ontwerpen.

Doelmatigheid

Evaluaties die beogen conclusies te trekken over doelmatigheid zijn gebaat bij enkele aanvullende onderzoeksmethoden (Tabel 5.2). Methoden die kunnen worden ingezet voor het meten van technische doelmatigheid behoren veelal tot de familie van productiviteitsonderzoeksmethoden – zoals data envelopment analysis (DEA), stochastic frontier analysis (SFA) en benchmarking (bijvoorbeeld door vergelijkingen te maken met hoeveel er in het buitenland wordt geïnvesteerd in soortgelijk S/T-beleid, en welke ontwikkelingen daar geobserveerd worden in output- en outcome-variabelen). Deze methoden zijn gericht op het bepalen en vergelijken van geobserveerde prestaties en/of kosten, en eventueel de verhouding daartussen.

Kwalitatieve methoden lenen zich ook het best voor het bestuderen van dynamische doelmatigheid, omdat hiervoor een aantal processen zichtbaar moet zijn. Bijvoorbeeld in hoeverre sprake is van het opknippen van beslissingen in stappen, de mate waarin geëxperimenteerd is en in hoeverre er een optie tot uitstel van beleid is opgenomen. Daarnaast kan er ook gebruik worden gemaakt van maatschappelijke kosten-batenanalyses (MKBA's) om de beoogde effecten in de toekomst af te wegen tegen de inspanningen op dit moment, als er op basis van bijvoorbeeld reeds bekende elasticiteiten of modellen empirische inschattingen mogelijk zijn over te verwachten effecten. Dit geeft een ex ante beeld van de totale beleidsinspanning die op punten in de tijd anders kan zijn. Ook wordt inzichtelijk van welke door beleid beoogde mechanismen de effecten onbekend zijn en waar dus effectmeting noodzakelijk is. Daarnaast kan - niet alleen ex ante maar ook in een ex post evaluatie - gebruik worden gemaakt van MKBA's. Dit laat zien of de voordelen (baten) van het beleid de nadelen (kosten) overtreffen. Met een "MKBA als denkkader" kan

worden vastgesteld welke effecten onbekend zijn en waar dus effectmeting noodzakelijk is.

Allocatieve doelmatigheid is in de praktijk moeilijk te beoordelen, omdat het vraagt om een weging van de inzet van middelen ten opzichte van andere beleidsterreinen. Casestudies lijken hier nog het meest voor geschikt, omdat beleidsterreinen integraal moeilijk te vergelijken zijn maar op basis van cases wel lessen kunnen worden getrokken over de efficiency van de bijdrage van beleidsinspanningen.

5.2. Inzetbaarheid van afzonderlijke methoden

In deze sectie gaan we nader in op enkele van de onderzoeksmethoden. Voor iedere methode volgt een bespreking van hoe die toegepast kan worden in de context van S/T-beleid. Uitgebreidere beschrijvingen en algemenere toelichtingen op de aard, toepassing, eisen, beperkingen, kosten en doorlooptijd van de methoden zijn te vinden op de website van de Toolbox Beleidsevaluaties.

Systematische reviews & meta-analyse

Beschrijving

Een systematische review is een specifieke vorm van literatuuronderzoek waarbij op een gestructureerde manier de resultaten van een aantal wetenschappelijke artikelen over een bepaald specifiek onderwerp gericht worden samengevat en beoordeeld. Een meta-analyse is een uitbreiding op de systematische review. Het is een statistische methode om de kwantitatieve uitkomsten van meerdere studies te vergelijken en te combineren.

Relevantie voor S/T-beleid

Een systematische review kan bruikbaar zijn om na te gaan of er bestaand empirisch bewijs en/of een theoretische basis is voor het bestaan van bepaalde knelpunten, en om werkingsmechanismen te identificeren via welke specifieke beleidsaanpakken die knelpunten te verhelpen. Het onderzoeken hiervan komt neer op het nalopen van geloofwaardigheid van

Tabel 5.2. Kenmerken van onderzoeksmethoden met relevantie voor het evalueren van doelmatigheid

Methode	Data			Analyse-doel		
	Theoretisch	Kwalitatief	Kwantitatief	Technische doelm. meten	Allocatieve doelm. meten	Dynamische doelm. meten
Simulaties	X		X	X		
DEA			X	X		
Stochastic Frontier analysis			X	X		
Benchmarking			X	X		X
MKBA			X			X
Casestudies		X		X	X	X

de beleidstheorie. Deze methode is geschikt voor het spiegelen van de interventielogica (invalshoek A) en de match tussen beleid en knelpunten (invalshoek C) aan wetenschappelijke inzichten. Systematische reviews zijn daarnaast bruikbaar om kaders te ontwikkelen (van normatieve aard) die aangeven volgens welke principes transitie-governance georganiseerd zou moeten worden (invalshoek B). Dit is met name behulpzaam bij het opzetten van een beleidstheorie. Meta-analyses kunnen worden gebruikt om de gemiddelde omvang van effecten te bepalen die in de literatuur zijn gevonden. Dit geeft een indruk van de mogelijke effecten van het te evalueren beleid.

Simulaties

Beschrijving

In een simulatie bootsen onderzoekers de werkelijkheid na door die werkelijkheid te vereenvoudigen met een model. In dat model is een aantal spelregels beschreven waar de individuele actoren (agenten) in het model (mensen of bedrijven) aan moeten voldoen. Vervolgens 'spelen' deze actoren het 'spel' volgens de regels en ontstaat er een bepaalde uitkomst, zoals een gedeelde consensus over een bepaald probleem of een evenwicht op een markt in het geval van economische modellen. Door de spelregels te veranderen kan worden bestudeerd welke invloed bepaalde factoren kunnen hebben op de uitkomsten. Simuleren is dus bij uitstek een methode om te leren, zowel over het gedrag van actoren als over de waarschijnlijke impact van interventies op dat gedrag. Er zijn twee soorten simulaties die respectievelijk gebruikmaken van mensen (spelsimulaties) of van rekenmodellen (computersimulaties). Beperkingen van dergelijke simulaties zitten in modelaannames en in representativiteit van een laboratoriumomgeving voor de werkelijkheid.

Relevantie voor S/T-beleid

S/T-beleid grijpt per definitie aan op systemen. Kenmerkend voor veel systemen is dat er sprake is van complexiteit als gevolg van een niet-lineaire feedback-loop en een samenloop van impulsen of gedragingen. Denk bijvoorbeeld aan (technologische) innovatiesystemen die de ontwikkeling en diffusie van een maatschappelijk wenselijke innovatie pas goed kunnen bevorderen als een reeks van complementaire sleutelprocessen allemaal in stelling is gebracht. Zolang processen nog zwak ontwikkeld zijn komt een ontwikkelingspad mogelijk maar langzaam op gang, terwijl het verhelpen van het zwakst ontwikkelde proces plots tot een grote versnelling kan leiden. Iets dergelijks kan zich ook voordoen bij het ontwikkelen van complementaire (energie)technologieën die van elkaar afhankelijk zijn, en dus elkaars ontwikkeling kunnen tegenhouden maar ook versnellen. In dit soort situaties kan het waardevol zijn om ex ante te onderzoeken hoe de samenloop van kenbare afzonderlijke dynamieken en gedragingen eruitziet. Zo is het in computersimulaties mogelijk om de relatieve invloed van

een enkele transitiefactor te analyseren in het geheel van transitiefactoren, om daarmee iets te zeggen over de waarschijnlijkheid dat structuurveranderingen en maatschappelijke impact zich zullen voordoen bij een bepaalde beleidsprikkel (invalshoeken E en F). Een alternatief is om met spelsimulaties een combinatie van gedragingen op gang te brengen. Dit is ook een manier om toekomstscenario's te ontwikkelen die beleidsvorming kunnen ondersteunen.

Lerend evalueren

Beschrijving

Lerend evalueren is de Nederlandse benaming voor een evaluatieaanpak die geïnspireerd is op in Sectie 4.1 besproken methoden als *developmental innovation*, waarbij o.a. participatie van beleidsmakers zelf een kenmerkend aspect is. Bij dit soort evaluaties werken evaluatoren en beleidsbetrokkenen samen om met het evaluatieonderzoek bruikbare en betekenisvolle inzichten aan te dragen voor beleidsontwikkeling en/of -uitvoering (Folkert et al., 2018). Doordat beleid onderzoeken en beleid ontwikkelen met elkaar verbonden zijn kan deze methode het best in de ex durante fase ingezet worden, bijvoorbeeld om verstrekte beleidsimpulsen te beoordelen op hun bijdrage aan langetermijndoelen zoals een transitieopgave. Lerend evalueren is in de eerste plaats een methode om kennis te genereren over welke beleidsinstrumenten wanneer (waar en onder welke omstandigheden) werken, maar biedt geen causaal bewijs voor effectiviteit. Daarmee is het vaak ook een vorm van 'realistisch' evalueren: ondanks het ontbreken van causaal bewijs is het nog altijd nuttig zijn om beter te doorgronden waarvan die (niet gemeten) effectiviteit afhankelijk is. Hoewel lerend evalueren vooral formatief ingezet wordt, kunnen ook summatieve elementen ingebracht worden; verantwoording is mogelijk zolang evaluatoren voldoende onafhankelijkheid in acht nemen bij het ontwerp en de interpretatie van het onderzoek.¹⁸

Een voorbeeld van lerend evalueren is de rol die onderzoekers van het Fraunhofer ISI Instituut vervullen voor het Duitse missiebeleid. Deze onderzoekers zijn op continue basis bezig met eigen analyses en interacties met allerlei stakeholders in het Duitse innovatiesysteem, en brengen daar af en toe analyses over uit (naar enkele van die tussentijdse evaluaties verwijzen we elders in het rapport). Ze hebben een expliciete rol als 'luis in de pels'; middels hun vragen en analyses dwingen ze beleidsmakers om scherp te blijven op het beargumenteren en demonstreren van de relatie tussen beleid (ontwerp en uitvoering) en uitkomst. Het continue kritische meedenken door de onderzoekers verschild van het model waarbij overheden op specifieke momenten een studie laten verrichten door externe onderzoekers die op relatieve afstand tot het beleid staan.

¹⁸ Verantwoording krijgt een bredere betekenis in een lerende onderzoekssetting. Naast de 'reguliere' betekenis van verantwoord en op doeltreffend,-doelmatigheid, is een lerende evaluatie met name geschikt voor horizontale (tussen betrokken partijen) en interne (beleidsconsistentie) verantwoording van beleid (Regeer et al. 2016).

Een voordeel van lerend evalueren is dat de uitvoerende onderzoekers gedurende een langere periode het ontwikkelen en uitvoeren van beleid volgen, en al een goede indruk hebben van de complexiteit ervan voordat ze zich aan een meer specifieke analyses wijdten. Dit in tegenstelling tot externe onderzoekers die gevraagd worden binnen enkele maanden doeltreffendheid te onderzoeken, maar in de praktijk eerst veel tijd en capaciteit moeten inzetten op het begrijpen van de details die daarvoor essentieel zijn. Een risico van dit model is dat de evaluatie en beleidsproces met elkaar vereenzelvigd raken, terwijl scheiding van belang is om kwaliteit en geloofwaardigheid van evaluatie te waarborgen.

Relevantie voor S/T-beleid

Lerend evalueren is bruikbaar op het moment dat de werking van het S/T-beleid dermate complex is dat de onderliggende beleidstheorie niet goed uit bestaande beleidsstukken is af te leiden. Het inzicht verkrijgen in hoe het beleid uiteindelijk tot impact moet leiden kan dan gebaat zijn bij het combineren van evaluatiekennis van onderzoekers (bijvoorbeeld als het gaat om het ontwikkelen van een beleidstheorie en het identificeren van bijbehorende indicatoren) en inhoudelijke kennis van beleidsbetrokkenen. De methode is dus geschikt als onduidelijk is via welke mechanismen eventuele effecten bereikt worden, en er behoefte is aan het structureren van relaties tussen inputs en effecten van diverse ordes (invalshoek A, interventielogica). Zo levert lerend evalueren ook een basis op die bruikbaar is in de andere invalshoeken om beleid en uitkomsten aan te toetsen.

Verder vraagt de dynamiek van een transitie om een lerende houding, zodat waar nodig bijgestuurd kan worden om beleid te richten op de dan dominante knelpunten. Lerend evalueren kan een onderdeel zijn van de sturing op een transitie; er is een continu toetsend en verantwoordend element dat de beleidsopzet en -uitvoering versterkt (invalshoek B en C). Gedurende het evaluatieproces kunnen deze inzichten ertoe leiden dat bijstelling plaatsvindt.

Vanwege de focus op hoe beleid uitgevoerd wordt, en welke directe resultaten dat oplevert in het licht van een langetermijndoel of ambitie, is het bij lerend evalueren mogelijk om conclusies te trekken over de doorwerking van S/T-beleid. De methode richt zich dan ook op transitiestappen die dichtbij in de tijd liggen en waarbij voor volgende stappen mogelijk beleidsverandering nodig is – en dus reflexiviteit vereist is. Het gaat hierbij vooral om het boven tafel krijgen van argumenten die verklaren hoe het gevoerde beleid aangrijpt op ‘tweede orde’-effectcategorieën als systeemversterking en systeemtransformatie (invalshoek D). De methode is geschikt voor uitspraken over de uitvoerbaarheid van transitiedoelen die ver weg in de tijd liggen (20 – 50 jaar).

Casestudies

Beschrijving

Een casestudie is een onderzoeksstrategie waarmee één of enkele ‘gevallen’ diepgaand bestudeerd worden in hun natuurlijke omgeving, waarbij het geval vaak in ruimte en tijd wordt afgebakend. In beleidsonderzoek zijn bijvoorbeeld complexe besluitvormingsprocessen nogal eens object van een casestudie. Een casestudie wordt gekenmerkt door een klein aantal onderzoekseenheden, een arbeidsintensieve benadering, een voorkeur voor diepgang boven breedte van het onderzoek en een beredeneerde, geïnformeerde keuze van gevallen. Een casestudie kan beschrijvend, toetsend of verklarend zijn, maar de externe validiteit van de bevindingen is vaak beperkt.

Relevantie voor S/T-beleid

In de context van S/T-beleid is het niet vanzelfsprekend waar het begrip ‘case’ precies betrekking op heeft: het kan o.a. gaan over organisaties, projecten, agenda’s, instrumenten, (innovatie) systemen, (deel)missies of transitiepaden. Een potentieel nuttige afbakening is die van een innovatie- of oplossingspad, wat bijvoorbeeld kan draaien om een technologische innovatie en daaraan gerelateerde businessmodelinnovaties. Om te begrijpen hoe belangrijk beleid is voor een innovatie als ‘elektrisch vliegen’ helpt het soms om los te komen van specifieke actoren en projecten, en op het niveau van de innovatie te reconstrueren welke ontwikkelingen van belang waren voor groei of stagnatie. Een techniek als *history-event-mapping* is geschikt om op gestructureerde wijze inzichtelijk te maken wat er in de loop der tijd is voorgevallen in relatie tot de innovatie, en welke oorzakelijke verbanden er (voor de betreffende case) te leggen zijn met het beleid.

Met deze methode kan een beeld ontstaan van de knelpunten waar de innovatie mee te maken heeft gehad (invalshoek C), of welke systeemcondities bevorderlijk dan wel hinderlijk waren (en of beleid daar van invloed op is geweest; invalshoek D). Een dergelijke analyse is ook te verrichten op het niveau van projecten, maar vaak zegt dat wat minder omdat projecten sterk gebonden zijn aan individuele regelingen (die weer hun eigen scope kennen, bijv. in termen van ontwikkelingsfase). Voor S/T-beleid is juist het mesoniveau van systemen van belang; die moeten zodanig ingericht zijn dat innovaties alle ontwikkelingsfasen tot aan grootschalige adoptie kunnen doorlopen. Vandaar dat casestudies zich wellicht beter kunnen richten op concrete innovaties, die uiteraard wel vaak berusten op allerlei onderliggende projecten.

Outcome harvesting¹⁹

Beschrijving

Outcome Harvesting (OH) is een methode om in complexe situaties te bestuderen hoe een bepaalde uitkomst het gevolg is van een keten van oorzaken (tussenuitkomsten). Onderscheidend kenmerk van OH is het open karakter van de methode. Die gaat niet uit van een vooraf gespecificeerde beleidstheorie, maar definieert empirisch geïdentificeerde veranderingen als een 'uitkomst' (bijvoorbeeld een groei in circulaire praktijken) en probeert vervolgens te achterhalen hoe een interventie (bijvoorbeeld een plotselinge wijziging in een beleidsmaatregel) heeft bijgedragen aan deze specifieke verandering. De Most Significant Change (MSC) methode is een beknopte variant van OH, waarbij op caseniveau bewijs wordt verzameld over factoren die bijdragen aan een uitkomst. De focus ligt bij zowel OH als MSC op het inventariseren van uitkomsten, wat nuttig kan zijn als inputs, activiteiten, outputs en resultaten op voorhand niet specifiek benoemd en meetbaar zijn. Een belangrijk verschil met monitoring is dat voor die laatste methode juist wel bekend moet zijn welke indicatoren er gemeten moeten worden om zicht te houden op de beleidsinzet en het doelbereik.

Relevantie voor S/T-beleid

Voor complexe beleidsprogramma's is vaak geen accurate beleidstheorie beschikbaar. Specifiek voor S/T-beleid geldt dat dit deels ook inherent is aan het feit dat er, in tegenstelling tot bij klassieke innovatiebeleidsinstrumenten, initieel vaak niet een concreet probleem is aan te wijzen dat verklaart waarom een maatschappelijk wenselijk doel niet gehaald wordt. Juist het aanjagen van activiteit moet dan uitwijzen waardoor een innovatie of zelfs transitie gehinderd wordt, en hoe beleid daarop kan inspelen. In die zin is OH potentieel bruikbaar in de evaluatie van S/T-beleid. Op het moment dat er programma's en instrumenten zijn ingericht om een impuls te geven aan systeemversterking of -transformatie kan onderzoek gedaan worden naar de manieren waarop de verstrekte impuls relateert aan die prestaties. Coördinatie faciliteren is bijvoorbeeld vaak een centraal element van S/T-beleid, maar evaluaties moeten uitwijzen op welke manier die coördinatie uiteindelijk tot impact leidt. Dit kan bijvoorbeeld schuilen in nieuwe R&D-samenwerkingsverbanden (privaat-publiek, privaat-privaat), afstemming tussen bedrijven en overheden, stroomlijning van ontwikkeling en adoptie etc. De OH-methode, die in algemene zin zicht geeft op bedoelde en niet-bedoelde uitkomsten, sluit dus goed aan bij het vertrekpunt dat er bij S/T-beleid niet altijd op voorhand in detail gespecificeerd kan worden hoe het beleid tot wenselijke veranderingen gaat leiden.

Doordat OH is toe te passen door voor succesvolle projecten na te gaan waar zij baat bij hebben gehad, geeft de methode geen zicht op structurelere veranderingen op een schaal die groter

is dan alleen de onderzochte projecten. Het is ook niet goed mogelijk om uitspraken te doen over het tot stand brengen van maatschappelijke impact. Vertrekpunt is immers om te verklaren wat de oorzaken zijn voor een positief gevolg, en OH biedt geen manier om de resulterende theorie te falsificeren.

Contribution analysis & process tracing

Beschrijving

Contribution analysis en Process tracing zijn kwalitatieve methoden gericht op het onderzoeken ('openen') van de black box van beleid. Ze kunnen inzicht geven in de werking van beleid en in het optreden van gewenste ontwikkelingen. Daarbij ligt de nadruk op het onderzoeken van veranderingen die zich volgens een beleidstheorie zouden moeten voordoen, zonder dat er uitspraken kunnen worden gedaan over de causaliteit. De bruikbaarheid van dergelijke analyses schuilt in het kunnen nalopen tot op welke hoogte in de *theory of change* er wenselijke ontwikkelingen waarneembaar zijn, en waar precies het stagneert, zonder dat voor de werking (bijdrage) van beleid definitief bewijs wordt geleverd. Beide methoden worden vaak gebruikt binnen casestudies. Er zijn ook verschillen:

- Contribution Analysis (CA) kan meer inzicht geven in de specifieke bijdrage van beleid aan een uitkomst door het beter begrijpen en verklaren van de redenen van bepaalde effecten en de rol van beleid en andere factoren daarbij. Dit gebeurt door de praktijk te toetsen aan de beleidstheorie, waarbij wordt gekeken naar de mate waarin (tussentijdse) resultaten zich voordoen, assumpties te onderzoeken, implementatie te controleren, andere beïnvloedende factoren te testen en alternatieve verklaringen uit te sluiten;
- Process tracing (PT) start met het opstellen van hypothesen over oorzaak en gevolg uit de beleidstheorie, waarbij voor elke hypothese beschreven wordt welke processen en mechanismen zich zouden voordoen indien de hypothese klopt. Ook wordt beschreven welke gegevens en welk bewijs hiervoor te vinden is. Door een combinatie van kwantitatieve en kwalitatieve dataverzamelmethode wordt vervolgens het proces en de resultaten van de (werking van de) praktijk blootgelegd en wordt getoetst in hoeverre de theorie strookt met de praktijk. Ook alternatieve hypothesen worden geformuleerd en kunnen worden getoetst.

Het belangrijkste verschil tussen beide methoden is dat het toetsen van hypothesen bij process tracing meer gericht is op het identificeren van de processen die in de praktijk spelen en die ervoor kunnen zorgen dat beleid in een bepaalde context wel of niet werkt, terwijl contribution analysis meer gericht is op het relatieve belang van verschillende verklaringen. Een gemeenschappelijk kenmerk is dat beide methoden een aanpak

¹⁹ In de Toolbox Beleidsevaluaties wordt ook 'Outcome Mapping' besproken, maar hoewel die methode gerelateerd is aan Outcome Harvesting leent die zich meer voor het vooraf plannen van een verandertraject (i.p.v. het achteraf nagaan van welke veranderingen zich hebben voorgedaan).

bieden om op zoek te gaan naar antwoorden op de vraag hoe en waarom effecten tot stand komen, zonder dat er uitspraken worden gedaan over de omvang van de effecten. Een gedetailleerde beleidstheorie is vereist voor toepassing van deze methode, als structuur om doorwerking van beleid op te plotten.

Relevantie voor S/T-beleid

Een aanname bij het voeren van S/T-beleid is doorgaans dat coördinatie – en dus governance – essentieel is voor systeemversterking of -transformatie en transitie. Contribution analysis en process tracing helpen om inzichtelijk te maken hoe eigenschappen van de gekozen governanceprocessen en beleidsmix optellen naar een keten van wenselijke of onwenselijke uitkomsten, waar ook systeemversterking en structuurveranderingen onderdeel van kunnen zijn (oftewel: de methode bestrijkt invalshoeken B-F).

Het nalopen van die effectketen geschiedt stapsgewijs op basis van de beleidstheorie. Als bijvoorbeeld 'openheid' een belangrijk governanceprincipe is dan zou dit consequenties moeten hebben voor de diversiteit van actoren die betrokken is bij het schrijven van roadmaps, hetgeen weer van invloed kan zijn op de samenstelling van projectteams, de variatie in oplossingsrichtingen in een projectportfolio etc. En indien een 'focus op selectief gekozen topics' een belangrijk governanceprincipe is dat ook ter harte wordt genomen in de uitvoering, dan zou dit zich moeten manifesteren in een toegenomen mate van selectiviteit en coherentie in de innovatiepaden die centraal staan in pogingen een transitie te ontketenen, wat weer terug te zien zou moeten zijn in bijvoorbeeld de scope van investeringen en leningen of van experimenten die ontstaan.

De methoden zijn potentieel interessant voor de evaluatie van S/T-beleid, omdat het nalopen van causale ketens niet alleen antwoord geeft op de vraag óf er zich veranderingen voordoen (die consistent zijn met de beleidstheorie), maar ook op welk punt in de beleidstheorie deze veranderingen plaatsvinden. Het zou immers kunnen voorkomen dat het invoeren van S/T-beleid geassocieerd kan worden met aanpassingen in de strategieën van invloedrijke actoren, maar dat de veranderingen stagneren op het moment dat de activiteiten van de actoren zich moeten vertalen in een verandering als een sprong in de productie van duurzame waterstof. Kennis van waar de keten tussen beleid en gewenste uitkomsten 'breekt' is relevant voor zowel formatieve als summatieve evaluaties. Een beperking van de twee methoden is dat ze weinig ruimte bieden voor 'verrassingen', aangezien een vooraf gedefinieerde beleidstheorie het uitgangspunt vormt. In dat opzicht is de OH-methode effectiever als werkingsmechanismen nog ter discussie staan.

Qualitative Comparative Analysis (QCA)

Beschrijving

Qualitative Comparative Analysis (QCA) is een methode om te achterhalen waarom een beleidsmaatregel onder bepaalde omstandigheden wél het beoogde resultaat behaalt, en onder andere omstandigheden niet. Doel van de methode is het identificeren van (combinaties) van condities die noodzakelijk en/of voldoende zijn voor het wel of niet optreden van een bepaalde uitkomst. Op systematische wijze worden casussen vergeleken. Dat hoeft niet noodzakelijkerwijs één model op te leveren; de methode staat open voor meerdere (rivale) verklaringen over de relatie tussen conditiesets en uitkomsten die vervolgens getoetst kunnen worden. QCA is vooral geschikt om te begrijpen *wat* potentieel invloed heeft gehad op een uitkomst.

Relevantie voor S/T-beleid

QCA leent zich voor analyse van een complexe beleidsrealiteit. Eén van de principes van de methode is het kunnen omgaan met causale complexiteit (condities die alleen in combinatie met andere condities een bepaalde uitkomst produceren, A hoeft niet lineair tot B te leiden en er kan sprake zijn van causale asymmetrie). In de complexe beleidsrealiteit van systemen en transities kan middels QCA gezocht worden naar de condities die van invloed zijn op het optreden van bepaalde effecten. Dit vergt dat er meerdere vergelijkbare casussen zijn, waarbij in de uitkomstvariabele voldoende variëteit is om te kunnen onderzoeken welke combinatie(s) van condities geassocieerd zijn met een betere uitkomst. Daarnaast mag het aantal mogelijk relevante (contextuele) condities niet te groot zijn, dient de variëteit op contextuele condities laag te zijn én dient de variëteit op de 'causale' condities weer hoog te zijn.

Caseselectie is binnen het S/T-beleid een uitdaging, zeker waar het gaat om evaluaties op mesoniveau (van systemen of transities). Op het oog lijkt de methode geschikt om vergelijkingen te maken tussen systeemversterkende netwerken, zoals gevonden in het Techniekpact of zoals bij de Smart Industry Fieldlabs, maar als die netwerken in veel opzichten van elkaar verschillen is het lastig om te voldoen aan condities gerelateerd aan variëteit van contextuele condities. In het geval van de missies uit het Missiegedreven Topsectoren- en Innovatiebeleid (MTIB) is het daarom lastig om de 25 missies te vergelijken en de vraag te stellen welke factoren samenhangen met geobserveerde prestaties op het behalen van de gestelde (zeer ongelijksoortige) missiedoelen. In plaats daarvan is QCA beter geschikt om onderzoek te doen naar één missie waar vanuit parallelle maar gelijksoortige sporen aan gewerkt wordt, bijvoorbeeld wanneer meerdere regio's zich allemaal op eenzelfde manier toeleggen op het aanjagen van de circulaire economie. In deze situatie kan voor die regio's vergeleken worden welke condities (variatie in de aanwezigheid van beleidspririkkels, industrie, actieve burgers etc.) gerelateerd kunnen worden aan hogere prestaties op circulaire-economie-indicatoren. Een soortgelijke toepassing is QCA gericht op het verklaren van

verschillen in de mate waarin de Regionale Energie-Strategieën (RES'en) van de 30 energieregio's in Nederland succesvol zijn.

Regressieanalyse

Beschrijving

Regressieanalyse is een statistische analyse die het mogelijk maakt verbanden tussen variabelen inzichtelijk te maken – zie Expertwerkgroep Effectmeting (2012). Hierbij kan zowel de aanwezigheid en omvang van het verband als het type verband (positief/negatief) worden onderzocht. In essentie duiden de resultaten van een regressieanalyse alleen op correlatie (gecontroleerd voor factoren die die correlatie kunnen vertroebelen), en geeft het geen uitsluitel over causaliteit. Wel kan met verschillende methoden causaliteit aannemelijk worden gemaakt, bijvoorbeeld door het toepassen van paneltechnieken en op verschillende manieren gebruik te maken van discontinuïteiten. Voor regressieanalyse is voldoende kwantitatieve data een vereiste, waarbij het kan gaan om data uit monitoring of survey-onderzoek maar vaak ook administratieve data zoals die bijvoorbeeld te vinden zijn in de CBS microdata.

Relevantie voor S/T-beleid

In geval van S/T-beleid zijn regressieanalyses het best inzetbaar bij interventies die op een bepaald onderdeel van een systeem aangrijpen, zonder dat er veel andere moeilijk kenbare/meetbare factoren op dat aspect van invloed zijn. Kennisontwikkeling over een bepaald onderwerp (zoals sensortechnologie) kan bijvoorbeeld belangrijk zijn voor een systeem of transitie, en het is denkbaar dat er gerichte interventies zijn die op betrouwbare wijze gerelateerd kunnen worden aan ontwikkelingen in die kennisproductie. Lastiger is het doen van uitspraken over de mate waarin toegenomen kennisproductie doorwerkt in snellere ontwikkeling, productie en uitrol van een bepaalde innovatie. Die uitkomstaspecten hangen bij S/T-beleid namelijk vaak van andere systeemfactoren af. Om te bepalen of regressieanalyses inzetbaar zijn is het dus zaak om eerst te verifiëren of een beleidsdoel echt vraagt om het versterken of transformeren van een systeem, of dat er een rechte causale lijn getrokken kan worden tussen beleidsimpulsen en beleidsuitkomsten.

(Quasi-)experimentele methoden

Beschrijving

Experimentele en quasi-experimentele onderzoeksmethoden dienen om causale verbanden vast te stellen tussen een verandering in een onafhankelijke variabele (zoals betere voorlichting over het werken in de techniek aan een groep van jonge kinderen en veranderingen in de afhankelijke doelvariabele instroom in het techniekonderwijs) en een afhankelijke variabele. Op deze manier kan men het effect van een maatregel aantonen voor deze groep ten opzichte van een andere groep. Concreet betekent een experimentele onderzoeksmethode dat er twee vergelijkbare groepen nodig zijn, waarbij de ene groep (experimentele groep) wel wordt blootgesteld aan een interventie

en de andere groep (controlegroep) niet, waarbij deze laatste groep willekeurig wordt samengesteld; 'randomized trials' zijn hier een voorbeeld van. Bij een quasi-experimentele methode wordt een groep deelnemers die een specifieke interventie ondergaat, vergeleken met een controlegroep die interventie niet ondergaat. De toewijzing van deelnemers aan de groepen is echter niet willekeurig, vandaar de term quasi-experimenteel. Denk bijvoorbeeld aan beleid dat in bepaalde regio's eerder in wordt gevoerd dan in andere. Het verschil in uitkomst tussen de twee groepen kan daarmee worden toegeschreven aan de interventie of de maatregel.

Relevantie voor S/T-beleid

In S/T-beleid is het belangrijk om te experimenteren. Dit maakt het mogelijk om partiële effecten van het beleid te toetsen middels een kleinschalig experiment om zo de effecten van beleidsprikkels te leren kennen. Daarnaast kunnen de effecten van individuele instrumenten worden getoetst door vooraf een helder design op te stellen – zie Expertwerkgroep Effectmeting (2012). Voor het S/T-beleid als geheel is deze methode waarschijnlijk vaak hooguit relevant voor slechts een subset van de beleidsmix en eerder op een output- dan op een outcome-niveau. De reden is dat outcomes te algemeen geformuleerd zijn en worden beïnvloed door verschillende factoren en instrumenten.

Monitoring

Beschrijving

Bij monitoring worden gegevens over bijvoorbeeld organisaties, projecten, systemen of interventies periodiek en met een vast patroon verzameld, opgeslagen, verwerkt en gerapporteerd. Met deze methode kunnen bijvoorbeeld de prestaties van een organisatie gevolgd worden en worden bijgestuurd wanneer dat nodig is. Monitoring is niet per definitie gebonden aan de inzet van een beleidsmaatregel of doelen van beleid of uitvoering, het is een voorwaarde om te kunnen meten. De gebruikte indicatoren kunnen ook belangrijke contextuele informatie bieden van een beleidsmaatregel.

Relevantie voor S/T-beleid

Continue en langdurige monitoring is een voorwaarde om beleidsinput, -voortgang, -output en -outcome in beeld te krijgen. Het systematisch bijhouden van de beleidsinzet en uitkomsten daarvan is van belang om de validiteit van de beleidstheorie op verschillende momenten in de tijd in beeld te brengen. In transities is monitoring noodzakelijk om de dynamiek van een transitie in beeld te brengen. Er is tijdens een transitie behoefte aan indicatoren die duiden op de fase waarin een transitie zich bevindt en op welke manier beleid daarbij past en/of aangepast moet worden. Tevens is het in het geval van een transitie van belang dat de mate waarin het einddoel in zicht is en blijft gemeten wordt. Ten slotte verhoogt het monitoren van tussendoelen (systeem- of transitieparameters) de betrouwbaarheid waarmee conclusies kunnen worden getrokken over de impact van beleid.

6. Stappenplan voor het voorbereiden en ontwerpen van evaluatiestudies

Bij S/T-beleid is het vaak niet evident hoe het beleid (via het versterken of transformeren van systemen) uiteindelijk dient bij te dragen aan de gestelde beleidsdoelen. Vaak moet er eerst verhelderd worden, bijvoorbeeld wat het S/T-beleid precies omvat (coördinatiestructuren, een beleidsmix, beide), wat de achterliggende beleidstheorie is, en om welk systeem het gaat en hoe dat is afgebakend.

In de voorgaande hoofdstukken is een reeds concepten en kaders geïntroduceerd die samen een basis kunnen bieden voor het gestructureerd nalopen van de essentie van S/T-beleid en de methodologische mogelijkheden om dat te evalueren. In dit hoofdstuk vatten we dit samen in een stappenplan. Naar verwachting zijn beleidsmakers zelf het best in staat om de eerste stappen te doorlopen, en neemt richting de latere stappen de rol van evaluatoren geleidelijk toe.

Stap 1: Wat is de doelstelling van het beleid?

S/T-beleid is erop gericht de productie en verspreiding van publieke goederen en diensten te realiseren, waarbij deze goederen en diensten het resultaat zijn van een interactief proces tussen verschillende actoren, zoals bedrijven, universiteiten en de overheid. Zonder ingrijpen van de overheid is sprake van een ineffectief systeem of van een te trage transitie. Het ultieme doel van het beleid moet worden geformuleerd en ook het pad ernaartoe, inclusief het transitiepad, de verandering van de werking van het systeem (en de functies) en de termijn waarop bepaalde doelen van beleid bereikt dienen te worden.

Stap 2: Welke falens zijn reden geweest voor overheidsingrijpen?

Markt-, systeem- en transitiefalens moeten in beeld worden gebracht aan de hand van de verschillende vormen van falen uit Hoofdstuk 3. Het identificeren van dergelijke falens is een manier om concreet te maken wat de rationale is voor overheidsingrijpen, oftewel wat de specifieke problemen zijn die het beleid dient te adresseren (om effectief en efficiënt van invloed te zijn op een maatschappelijk einddoel) en welke type beleid daarbij hoort. Tevens moet worden getoetst in welke mate de overheid met beleid

het falen kan oplossen of verminderen of dat er een gereede kans op overheidsfalen is.

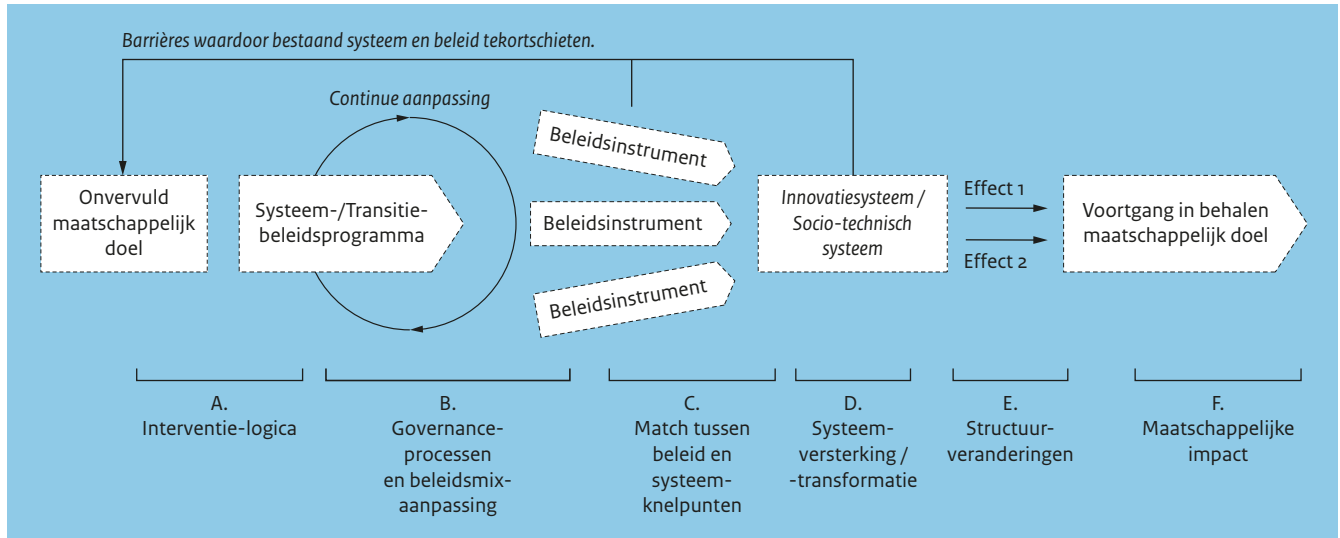
Stap 3: Hoe werkt het systeem of hoe is de transitie vormgegeven?

Het evalueren van S/T-beleid begint met het afbakenen van het systeem of de transitie en de looptijd van de evaluatieperiode. Een belangrijke vraag is of er een *top-down* aanpak gevolgd kan worden waarbij de instrumenten op een begrotingsartikel over een bepaald onderwerp leidend zijn (zoals bij een beleidsdoorlichting), of dat er een *bottom-up* aanpak gevolgd wordt waarbij het beleid wordt afgebakend door te kijken naar alle beleidsinstrumenten die tot een systeem of transitie behoren.

Hieruit volgt de interventielogica. Deze logica bestaat uit het uitdiepen van de kenmerken die bepalend zijn voor wat het beleid beoogt te veranderen aan de werking van een systeem of het bewerkstelligen van een transitie en hoe die beoogde doelen tot stand komen. Om dit proces te structureren moeten acht vragen beantwoord worden:

1. Definitie. Wat is de definitie van het systeem of de transitie?
2. Doel. Wat zijn de specifieke problemen die het beleid probeert te verhelpen, en wat zijn de redenen daarvoor?
3. Beleid. Wat zijn de beleidsactiviteiten en -instrumenten die vanuit het S/T-beleid ingezet worden, wie voert deze uit en wat is het budget?
4. Bijdrage. Op welke manier dragen deze inspanningen en instrumenten bij aan de beoogde doelen en op welke termijn?
5. Samenhang. Welke samenhang bestaat tussen de inspanningen en instrumenten?
6. Monitoring. Hoe relateren eventuele monitoringsactiviteiten en -indicatoren aan de werkingsmechanismen en doelen?
7. Overheidsfalen. Wat zijn redenen en indicaties voor het niet werken van het beleid?
8. Aard van het beleid. In hoeverre is het beleid dynamisch of adaptief?

Figuur 6.1. **Ordering van de zes invalshoeken voor de evaluatie van S/T-beleid**



Stap 4: Welk deel van het systeem of de transitie moet geëvalueerd worden?

Een systeem of transitie wordt meestal niet als geheel geëvalueerd, omdat het bij een systeemevaluatie gaat om de vraag in hoeverre verschillende functies in het systeem effectief zijn en bij een transitie de te behalen doelen ver in de toekomst liggen. Er zijn zes invalshoeken voor evaluatie (Figuur 6.1). Welke invalshoek het meest relevant is zal onder andere te maken hebben met waar in de beleidscyclus een S/T-programma staat. Op het moment dat het programma pas net is ingezet kan er mogelijk al benoemd worden aan welke instrumenten het in de eerste instantie vooral richting zal willen geven, maar het evalueren van die coördinatie of coherentie kan pas plaatsvinden wanneer er een periode verstreken is waarin de beleidsmix van kracht is. Het verbeteren of transformeren van systemen vergt doorlooptijd. Het evalueren van de legitimiteit van de interventie (is er een duidelijk knelpunt dat overheidsingrijpen rechtvaardigt?) en de governanceprocessen kan al aan het begin van de inzet van een nieuw programma, maar kunnen ook relevant zijn als dat programma al langer loopt.

Stap 5: Formuleer hypothesen over wat het beleid zou moeten bereiken en koppel deze aan de beleidstheorie

Er moeten hypothesen worden geformuleerd over wat het beleid teweeg zou moeten brengen. Hypothesen waarborgen dat de evaluatie zich kan toespitsen op een verondersteld mechanisme en een beoogd doel.

- Kernhypothesen: stel het hoofddoel van het systeem of de transitie op.
- Hypothese directe effecten: stel een hypothese op die concreet maakt welk effect wordt verwacht van het te evalueren deel van het beleid.
- Hypothese indirecte effecten: stel een hypothese op over uitstralingseffecten of hogere orde effecten van het beleid.

- Hypothese maatschappelijke effecten: stel een hypothese op over de impact van het beleid.

Op ieder van de zes invalshoeken zijn hypothesen denkbaar, ze zullen alleen een wat ander karakter hebben. Dat wil zeggen, de hypothesen over de linkerkant van Figuur 6.1 gaan bijvoorbeeld vooral over of de coördinatie en de instrumentenmix adequaat georganiseerd zijn en in hoeverre het systeem of de transitie de gewenste richting hebben en volgens het gewenste tempo verlopen, terwijl de rechterkant aan de orde zal stellen wat de effecten van beleid zijn op systemen, structuren en einddoelen.

De hypothesen kunnen veelal gekoppeld worden aan een mechanisme in de beleidstheorie en vormen in die zin ook een toetsing of de mechanismen die de beleidstheorie aanneemt zich ook in de praktijk voordoen. Bij hypothesen die betrekking hebben op de linkerkant van Figuur 6.1 is het mogelijk dat er maar weinig gezegd kan worden over hogere orde doelen. Het analyseren van output in de zin van gerealiseerde governanceprocessen en hun invloed op de instrumentenmix is waarschijnlijk vaak nog wel haalbaar, maar de doorwerking van governanceprocessen op eerste-orde en hogere orde doelen is empirisch moeilijker of niet te bepalen. Aan de hand van de beleidstheorie is het tevens mogelijk dat er additionele hypothesen bestaan, waardoor terug moet worden gegaan naar stap 5.

Stap 6: Identificeer kader(s) en bijbehorende indicatoren om prestaties (op diverse doelorden) aan te toetsen, liefst op zo hoog mogelijke treden van de effectladder

De volgende stap is het vaststellen van de evaluatiemethode. Bij klassieke effectmetingen zoals besproken in het rapport van de Expertwerkgroep Effectmeting (2012) gaat het dan vaak om het kiezen of construeren van een controlegroep, of een andersoortige *counterfactual* (bijvoorbeeld huidig presteren afzetten tegen presteren in het verleden). Idealiter worden dat soort technieken

Tabel 6.1. **Overzicht van onderzoeksmethoden en de invalshoeken waarvoor ze het best inzetbaar lijken**

	Interventie-logica	Governance en beleidsmix	Match tussen beleids(mix) en knelpunten	Systeem-versterking en -transformatie	Structuur-verandering	Maatschappelijke impact
Doeltreffendheid						
Systematische reviews & meta-analyse	Systematische reviews betreft kaders en principes identificeren om beleidsuitvoering aan te toetsen. Een meta-analyse is een cijfermatige analyse.					
Lerend evalueren	Mechanismen verklaren en (samen) interpreteren, vertrekkend vanuit beleidstheorie en -uitvoering					
Case studies			Mechanismen verklaren & uitkomsten bepalen			
Outcome harvesting			Mechanismen verklaren, vertrekkend vanuit uitkomsten (hoe komt die uitkomst tot stand; wat was de rol van beleid?)			
Contribution analysis & process tracing		Causaliteit bepalen, vertrekkend vanuit de beleidstheorie (indicaties verzamelen voor optreden van een keten van uitkomsten)				
Qualitative Comparative Analysis (QCA)			Causaliteit bepalen, vertrekkend vanuit uitkomsten (welke combinatie van factoren bepaalt variatie in succes?)			
Regressieanalyse				Effectomvang meten, corrigerend voor controlevariabelen		
(Quasi-)experimentele methoden				Causaliteit bepalen & Effectomvang meten		
Doelmatigheid						
Monitoring		Ontwikkelingen bijhouden, op vooraf bepaalde indicatoren				
Productiviteits-onderzoek (DEA, SFA, benchmark)				Vergelijken van kosten en/of prestaties (of de verhouding daartussen)		
Simulaties*					Uitkomsten schatten van complexe (beleids)interacties	
MKBA						MKBA

* Simulaties kunnen ook zicht geven op doeltreffendheid, maar zijn hier voor de overzichtelijkheid maar één keer opgenomen

ook bij de evaluatie van S/T-beleid, zeker als het gaat om de doelen output, outcome en impact.

Daar waar dit niet mogelijk is zal er een beroep moeten worden gedaan op alternatieve afwegingskaders. Uitgaande van een formatieve evaluatieaanpak kunnen hiervoor stakeholders worden geraadpleegd die een beeld hebben van hoe het beleid zou moeten functioneren. Een alternatieve route om tot afwegingskaders te komen is het selecteren en eventueel combineren van theoretische raamwerken. Enkele voorbeelden hiervan zijn besproken in Hoofdstuk 4. Het voordeel hiervan is dat de raamwerken relatief wat beproefder zijn dan de beelden van stakeholders, zeker als ze volgen uit een rijke onderzoekstraditie met empirisch rigoureuze methoden. Een nadeel is dat de raamwerken niet altijd één op één passen bij een concrete casus. Om die reden is het belangrijk

om expliciet te benoemen tot op welke hoogte een theoretisch kader bruikbaar is voor de evaluatie van een bepaald beleidsaspect of hypothese. Het kan nodig zijn om meerdere raamwerken te integreren, of anderszins te vertalen naar de context in kwestie, conform de methode van lerend evalueren. Ook hier kan eventueel een beroep worden gedaan op stakeholders die zicht hebben op redenen om het Nederlandse beleid op een bepaalde manier in te steken.

Stap 7: Meten - Verricht de afzonderlijke deelanalyses

Wanneer alle relevante conceptuele kaders geselecteerd en geoperationaliseerd zijn kan het feitelijke meten plaatsvinden. Hiervoor kunnen onder andere de in Hoofdstuk 5 besproken kwantitatieve en kwalitatieve methoden gebruikt worden. Er bestaat een beperkt aantal methoden om de omvang van effecten

te meten, terwijl er diverse alternatieven zijn als het gaat om onderzoek naar werkingsmechanismen en verbanden.

Welke methode het meest geschikt is voor een evaluatie is onder andere afhankelijk van de precieze evaluatievraag en de databeschikbaarheid. Tabel 6.1 hieronder vat samen welke relaties er zijn tussen de zes evaluatie-invalshoeken en de methoden die daarvoor benut kunnen worden. Een gevulde cel betekent dat de methode waarschijnlijk op een manier (e.g. verkennend, toetsend) bruikbaar kan zijn om uitspraken te doen over die invalshoek. Bij het selecteren van een methode is het essentieel om de vraag te stellen wat er al bekend is over de werking van het beleid, aangezien sommige opties zich vooral lenen voor het blootleggen en begrijpen van werkingsmechanismen (wat bij S/T-beleid zelden zomaar afgeleid kan worden uit beleidsdocumenten) en andere opties geschikter zijn om empirisch vast te stellen of en wanneer er effecten zijn, hoe dit komt en hoe groot die effecten zijn.

Tabel 6.1 suggereert dat er maar weinig methodologische handvatten zijn voor het onderzoeken van de doelmatigheid en doeltreffendheid van beleidsdoelen die zich richten op governance en het bijstellen van de beleidsmix, terwijl deze aspecten cruciaal zijn voor experimenteel S/T-beleid dat voortkomt uit de gedachte “het is belangrijk om iets aan maatschappelijk doel X te doen, ook al weten we nog niet goed wat werkt”.²⁰ Dit komt doordat het beoordelen van governance meer te maken heeft met evaluatiecriteria als coherentie/consistentie, kwalitatief goede uitvoering en draagvlak voor beleid - en daarmee dus niet direct doeltreffendheid of doelmatigheid beschrijft. Wel is het natuurlijk zo dat het goed organiseren van beleid een opmaat kan zijn naar het genereren van effecten; vandaar ook de plek van invalshoek B in het bredere evaluatiekader.

Stap 8: Beleidskeuzes

In S/T-beleid is de manier waarop beleid wordt vormgegeven en uitgevoerd van belang. Bij een onzeker beeld over de toekomst zouden drie mogelijke vormen van uitvoering zichtbaar moeten zijn.

De eerste is de optie om te wachten met het invoeren van beleid als ex ante kosten en baten erop wijzen dat er in de toekomst hogere baten met dezelfde inspanning te behalen zijn. Afwegingen die zichtbaar moeten zijn bestaan uit vastgelegde keuzes om beleid in te zetten vanaf een bepaald moment, een afweging over de intensiteit en omvang van beleid en een inschatting van de kosten en baten.

De tweede vorm betreft het opknippen van beleid. Hierbij speelt een rol in hoeverre sprake is van een doelmatige keuze van verschillende opties met expliciet aandacht voor kosten, de optie

om beleid te kunnen stoppen en het meewegen van onzekerheid en risico's.

De laatste vorm betreft experimenteren. Onzekerheid ontstaat vaak door kennislacunes. Als dat het geval is, zou de optie om deze lacunes te dichten door te experimenteren zichtbaar moeten zijn. Een oordeel over in hoeverre sprake is van de juiste experimenten en in hoeverre deze doelmatig zijn uitgevoerd past hierbij, evenals een oordeel over de lessen die zijn getrokken voor beleid.

Stap 9: Combineer de inzichten uit de complementaire deelanalyses

Gegeven de relatieve voor- en nadelen van verschillende soorten analyses schrijft deze laatste stap voor om een synthese te maken waarbij de deelluitkomsten (mogelijk uit verschillende evaluaties) naast elkaar worden gelegd. Inzicht in vermeende sterktes en zwaktes in de governance-structuur kan op zichzelf al waardevol zijn, maar het is nog bruikbaar als er ook indicaties zijn voor de effecten die er van het beleid uitgaan – en vice versa. Bij het combineren en rapporteren van resultaten uit deelanalyses is het essentieel om scherp te blijven op het soort uitspraken dat er gedaan kan worden op basis van de ingezette methoden. Hierbij is het vooral zaak om te vermijden dat er gesproken wordt over algehele beleidseffecten op het moment dat die niet betrouwbaar gemeten zijn.

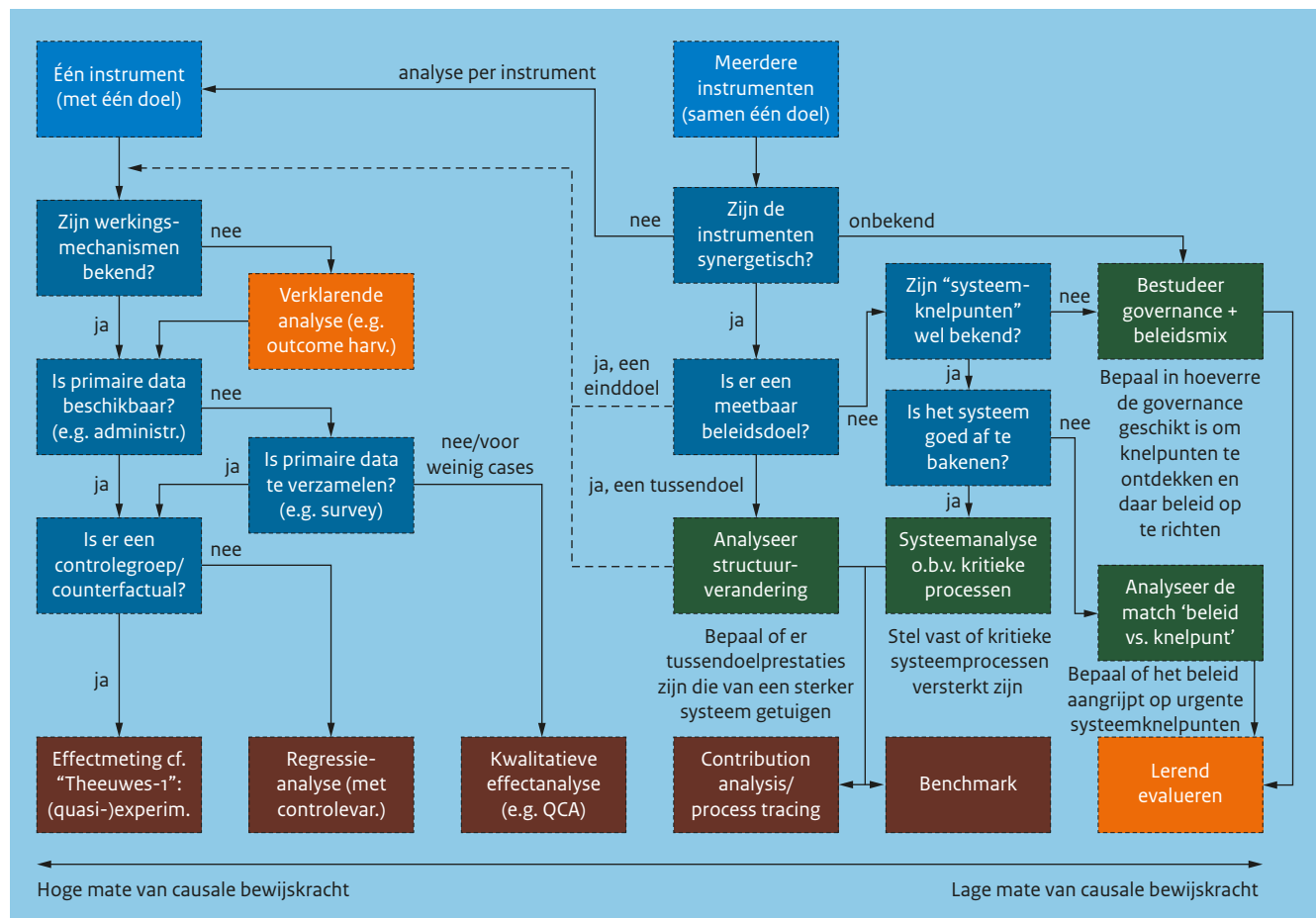
²⁰ Dit vertrekpunt (we weten niet of ons beleid werkt, maar het probleem is urgent en bij niets doen wordt het probleem groter) wordt aangeduid als input-legitimiteit. Dit verschilt van output-legitimiteit; die vorm stelt dat beleid legitiem is als aannemelijk gemaakt kan worden dat het doeltreffend en doelmatig is.

7. Beslissinghulp voor het selecteren van methoden

Om evaluatoren op weg te helpen bij het selecteren van methoden presenteert dit hoofdstuk een beslissinghulp in de vorm van een stroomschema (Figuur 7.1). Dit is niet een model waar op mechanische wijze een antwoord uitrolt, maar een schema om op systematische wijze af te wegen welke analysekaders en -methoden het best passen bij de eigenschappen van een concrete casus van S/T-beleid. Het dwingt dus ook tot nadenken over die eigenschappen die bij S/T-beleid zelden evident zijn en besproken in de eerdere hoofdstukken.

Door dit schema van boven naar beneden te doorlopen worden evaluatoren gewezen op de empirische methoden die het meest geschikt lijken voor een te evalueren casus. Deze methoden zijn grofweg gerangschikt op de mate van causale bewijskracht die ze kunnen bieden, conform de 'effectladder'; zie Sectie 4.1. Links in de figuur staan de methoden waar die causale bewijskracht het hoogst is, met de aanname dat daar ook een voorkeur voor is bij effectmetingen. In het rapport van de Expertwerkgroep Effectmeting is reeds beschreven welke concrete technieken overwogen kunnen worden bij het inzetten van (quasi-)

Figuur 7.1: Beslissinghulp voor het selecteren van empirische methoden t.b.v. evaluatie S/T-beleid



experimentele methoden. Hoewel de as niet volstrekt ordinaal is, neemt de causale bewijskracht af voor methoden naarmate die meer rechts in de figuur staan.

De te overwegen methoden zijn getoond in de donkeroranje blokjes. De meest rechtse methode, lerend evalueren, leent zich zoals eerder besproken niet voor uitspraken over causale effecten. Dit methode-blokje heeft daarom dezelfde lichtere kleur als het blokje ‘verklarende analyse’. De groene blokjes verwijzen naar de analysekaders die in beeld komen afhankelijk van de eigenschappen van het te evalueren beleid, en de informatie die daarover beschikbaar is. Ze zijn gekoppeld aan de invalshoeken uit het evaluatiekader van Hoofdstuk 4. Uit Tabel 6.1 bleek al dat er een samenhang is tussen die invalshoeken (en bijbehorende analysekaders) en geschikte empirische methoden.

7.1. Beslissing voor beleid met één instrument en één doel

Voor evaluatiestudies die gericht zijn op beleid dat berust op slechts één instrument met maar één beleidsdoel zijn er relatief goede mogelijkheden voor het toepassen van de meest wenselijke evaluatiestandaard, de (quasi-)experimentele methode. Het toepassen daarvan is haalbaar op het moment dat er aan drie voorwaarden kan worden voldaan:

- er moet een duidelijk beeld bestaan van de werkingsmechanismen van het beleid, zodat bekend is op welke indicatoren getoetst kan worden, bij wie, wanneer in de tijd (afhankelijk van de ‘incubatietijd’ van beleidspraktijken en -uitkomsten) etc.;
- er dienen data beschikbaar te zijn over o.a. de verstrekte praktijken en prestaties, bij voorkeur primaire objectieve data zoals beschikbaar in administratieve overzichten of bij het CBS; en
- het is noodzakelijk dat er een *counterfactual* (relevante vergelijkingsbasis) zoals een controlegroep is, of op zijn minst geconstrueerd kan worden (e.g., met behulp van econometrische methoden als *propensity score matching*).

Indien er niet aan de laatste voorwaarde kan worden voldaan kunnen evaluatoren terugvallen op regressieanalyses, waarbij bij de behandelgroep bepaald kan worden of er een statistisch significante samenhang is tussen beleidspraktijken en prestaties. Het is zaak om dan met behulp van controlevariabelen zoveel mogelijk te corrigeren voor factoren die ook op de geanalyseerde prestatie van invloed zijn. De causale bewijskracht zal echter minder sterk zijn dan wanneer er ook een valide vergelijking gemaakt kan worden met soortgelijke partijen die géén behandeling gehad hebben.

Als er ook geen relevante primaire data beschikbaar zijn in administratieve bronnen (die eventueel nog aan elkaar gekoppeld moeten worden), zal dat probleem eerst verholpen moeten worden. De vraag is dan of dergelijke data verzameld kunnen worden. Is het verzamelen van primaire data niet mogelijk, of slechts voor een handvol respondenten, dan kan er teruggerepen worden op gestructureerde vormen van kwalitatieve effectanalyse.

Mocht er überhaupt geen goed beeld bestaan van de mechanismen waarlangs beleid uiteindelijk geacht wordt bij te dragen aan beleidsdoelen, dan dient dit eerst achterhaald te worden. Hiervoor kan een beroep worden gedaan op verklarende analyses zoals *outcome harvesting* en casestudies. Door met dergelijke methoden in kaart te brengen op welke manier gewenste tussen- en einduitkomsten tot stand komen wordt het mogelijk op het stroomschema te vervolgen, en zo te bepalen welke methode ingezet kan worden om ook causaal te toetsen of dergelijke uitkomsten aan beleid zijn toe te schrijven.

7.2. Beslissing voor beleid met meerdere instrumenten en één doel

De eerste vraag is of de instrumenten uit een S/T-beleidsprogramma synergetisch zijn. Hiervan is sprake als die meervoudige instrumenten hetzelfde systeem of dezelfde transitie beïnvloeden, zonder dat de afzonderlijke invloed per instrument gedestilleerd kan worden. Het stroomschema biedt drie opties:

- De instrumenten zijn niet of beperkt synergetisch. In dat geval kunnen onderzoekers per instrument de linkerkant van de figuur doorlopen;
- Het is niet bekend of de instrumenten synergetisch zijn. Deze situatie doet zich voor als er onvoldoende zicht bestaat op hoe het S/T-beleidsprogramma precies in elkaar steekt. Dit speelt bijvoorbeeld als beleid zich vooral richt op coördinatie en het stroomlijnen van (de inhoud) van allerlei bestaande en nieuwe instrumenten. Nagaan of dergelijk beleid verschil maakt vergt een precies begrip van de organisatie van het beleid en de verhouding tot bestaande instrumenten. De *counterfactual* is namelijk niet dat al die instrumenten er niet zouden zijn, maar dat alleen de sturing vanuit het S/T-programma er niet zou zijn. Het ophelderend van de situatie kan middels analysekaders behorende bij **invalshoek B** uit het evaluatiekader: ‘**Governance en beleidsmix**’. Zoals beschreven in Sectie 3.4 draait om het in kaart brengen van de governance, en het bepalen van de mate waarin die geschikt is om beleid te richten op de meest urgente knelpunten in een systeem of transitie. Normatieve evaluatieve uitspraken vergen dat er principes geïdentificeerd moeten worden waaraan de governance moet voldoen. De methode van ‘lerend evalueren’ lijkt hier daarom bij uitstek relevant, aangezien participatie van beleidsmakers en -uitvoerders essentieel is bij het vaststellen van dergelijke principes;
- De derde mogelijkheid is dat ervan uitgegaan kan worden dat het beleid synergetisch is. Dit leidt tot de vraag of er ook een meetbaar beleidsdoel aan gekoppeld is.

De vraag naar een meetbaar doel gaat over de beschikbaarheid van een indicator die helpt aflezen of er ontwikkelingen zijn op de variabele die het beleid probeert te beïnvloeden. Ook hier zijn weer drie opties, waaronder twee varianten van ‘ja, er is een doel’:

- Er is een meetbaar einddoel; de variabele waar het uiteindelijk ook om te doen is kan direct gemeten worden. Indien dit het geval is kan gepoogd worden de gehele synergetische beleidsmix

als één gebundelde beleidsprikkels te beschouwen, en daar een evaluatie-ontwerp voor te ontwikkelen conform de linkerkant van het stroomschema. Idealiter zijn de werkingsmechanismen op hoofdlijnen bekend, zijn er data en kan er een *counterfactual* gevonden worden zoals de situatie vóór en na het invoeren van de S/T-beleidsmix. De lijn naar de linkerkant van het stroomschema is gestippeld, omdat het niet zeker is dat er uiteindelijk een geschikte methode geselecteerd kan worden. Als er de verschillende onderdelen uit de S/T-beleidsmix bijvoorbeeld heel geleidelijk zijn geïmplementeerd, of als er exogene factoren zijn die duidelijk ook een grote invloed hebben gehad op het beleidsdoel, dan wordt het moeilijk om een gedegen causale analyse op te zetten;

- Soms zijn er ook één of meerdere tussendoelen bekend, die eerst behaald moeten worden om de beoogde impact te kunnen realiseren. Het gaat dus niet om temporele tussendoelen, maar doelen die in de *theory of change* vóór het uiteindelijke beleidsdoel komen. Om beleidseffecten op dat tussendoel te evalueren kan gebruik worden gemaakt van **invalshoek E** uit het evaluatiekader, voor het analyseren van '**structuurveranderingen**'. Dit betreft veranderingen in (in dit geval) de economische structuur die van belang is voor het bereiken van het uiteindelijke beleidsdoel. Er kunnen diverse methoden overwogen worden:
 - Wederom is het verstandig om eerst te bepalen of de structuurverandering via de linkerkant van het stroomschema geëvalueerd kan worden. Bij het uitwerken van de MTIB-casus in Hoofdstuk 8 zien we bijvoorbeeld dat er mogelijkheden lijken te zijn om te bepalen of individuele bedrijven zich in hun eigen innovatieprojecten (zoals bekend in de WBSO-administratie) inderdaad meer zijn gaan richten op de thema's waar de MTIB-coördinatie en -beleidsmix een impuls aan hebben gegeven. Dit zou vergeleken kunnen worden met innovatieactiviteiten op gelijkende economische activiteiten of technologieën die net buiten de scope van de KIAs en MMIPs vallen;
 - Indien analyses met een hoge causale bewijslast niet haalbaar zijn, kan er teruggevallen worden op methoden als *contribution analysis* en *process tracing*. Deze bieden een basis om systematisch inzichtelijk te krijgen welke veranderingen er hebben plaatsgevonden op alle ordes van de effectketen tot aan het te onderzoeken tussendoel. Als er op al die punten positieve ontwikkelingen zijn levert dat nog geen causaal bewijs, maar zo'n reconstructie geeft wel bruikbare informatie over of er überhaupt wenselijke ontwikkelingen plaatsvinden. Is dat niet het geval, dan is dat mogelijk een indicatie voor een gebrek aan beleidseffectiviteit;
 - Een alternatieve methode is het benchmarken van meetbare structuurveranderingen, bijvoorbeeld door te kijken naar andere regio's of landen. Bij het ontwerp en de interpretatie van zo'n benchmark moet sterk rekening worden gehouden met de mate van vergelijkbaarheid: is er elders géén beleid (en is het dus qua beleidseffectiviteit in Nederland een slecht teken als de structuurverandering daar even snel gaat), of is

er juist hetzelfde soort beleid (en is het pas zorgwekkend als Nederland sterk achterblijft)?

Een derde mogelijkheid in het stroomschema is dat het S/T-beleid geen meetbaar einddoel kent. Dit is typisch aan de orde bij systeembeleid dat het functioneren van een (innovatie)systeem als geheel wil bevorderen, wat verschilt van transitiebeleid gericht op een concreet en meetbaar maatschappelijk doel inzake bijvoorbeeld CO₂-reductie. Indien het S/T-beleid gericht is op het versterken van systemen, zoals bij het oorspronkelijke Topsectorenbeleid het geval was, dan is de volgende vraag of bekend is wat de voornaamste knelpunten in die systemen zijn. Het stroomschema is op twee manieren te vervolgen:

- Nee, de systeemknelpunten zijn niet bekend. In dat geval resteert er weer niets anders dan het onderzoeken van de kwaliteit van '**Governance en de beleidsmix**', conform **invalshoek B**. Hierbij kan ook aandacht worden besteed aan de vraag of de governance en het beleid zodanig georganiseerd zijn dat er in ieder geval geleerd kan worden wat de systeemknelpunten zijn. Zowel in het domein van innovatiebeleid als klimaatbeleid zijn er beleidsstrategieën die niet vertrekken vanuit een reeds bekende set knelpunten, maar juist middels experimenteren proberen te achterhalen waarom bepaalde wenselijke veranderingen niet tot stand komen. Een evaluatie zou zich dus kunnen toespitsen op het uitwerken van het analysekader door passende principes te formuleren/identificeren, die van belang zijn bij het leren over knelpunten. Hierbij kan gebruik worden gemaakt van de aanpak van lerend evalueren, om zo ook enige mate van 'reflexiviteit' te waarborgen (reflectie op beleidsdoelen, kaders, organisatie etc.);
- Ja, de systeemknelpunten zijn bekend. Duidelijk is dat bijvoorbeeld specifieke processen als kennisontwikkeling of scholing niet goed op orde zijn voor het optreden van wenselijke economische of maatschappelijke veranderingen. Om te kijken of beleid daar verandering in heeft gebracht is het zaak om het systeem goed af te kunnen bakenen; in welke sector/technologiegebied/regio zouden de zwakke processen nu versterkt moeten zijn? Wederom volgt een tweesprong:
 - Het systeem is goed af te bakenen, op een manier waardoor er gegevens verzameld en geanalyseerd kunnen worden voor dat systeem. In dat geval dient een evaluatie uit te wijzen of het systeem na verloop van tijd beter is gaan functioneren. Hiervoor kan gebruik worden gemaakt van de analysekaders behorende bij **invalshoek D**, '**Systeemversterking/-transformatie**'. Een voorbeeld is het TIS-raamwerk met zeven sleutelprocessen, zoals dat ook gehanteerd is in de evaluatie van de Topsectoren-aanpak. Qua empirische methoden zijn de voornaamste opties hetzelfde als bij het analyseren van structuurverandering;
 - Ten slotte is het ook denkbaar dat het systeem niet goed af te bakenen is. Dit speelt als het systeem heel algemeen omschreven is, of als de grenzen niet goed te vertalen zijn naar variabelen waarmee geanalyseerd kan worden. Een 'regionaal innovatiesysteem' valt bijvoorbeeld niet altijd

binnen administratieve regiogrenzen, en verbindt vaak een set van actoren die niet goed te omschrijven is met behulp van sector-codes of technologiegebieden. Nagaan of het relatief ongrijpbare systeem beter functioneert is dan ondoenlijk. Wat wel nog een optie is, is vaststellen of het beleid ten minste aangrijpt op de bekende knelpunten. Zie hiervoor **invalshoek C, de 'Match tussen beleid en knelpunt'**. Onderzoek op basis van de in Sectie 4.4 aangereikte analysekaders zegt niets over de omvang van een causaal effect, maar het is wel waarschijnlijk dat beleid effectiever is als het urgente knelpunten aanpakt dan wanneer het de facto vooral gericht is op processen die geen probleem opleveren.²¹ Bij dit type analyse is de methode 'lerend evalueren' te overwegen, aangezien dat een manier is om met o.a. de beleidsmakers en -uitvoerders zelf na te lopen hoe de verstrekte beleidsimpulsen er in de praktijk precies uitzien. Zoals meermaals aangegeven is dit vaak onduidelijk bij S/T-programma's die vooral het karakter hebben van agenda's of coördinatiestructuren i.p.v. subsidies of regels.

²¹ De hier voorgestelde analyse gaat ervan uit dat reeds bekend is wat de systeem-knelpunten zijn. In de praktijk kan het zo zijn dat er wel bewijs is dat bijvoorbeeld een proces als 'kennisverspreiding' of 'marktcreatie' het knelpunt vormt, maar dat er geen goed beeld is van de oorzaak van dit knelpunt. In dat geval kunnen evaluatoren of andere onderzoekers eerst nog de probleemdiagnose aanscherpen, door bijvoorbeeld te kijken of relevante actoren een gebrek hebben aan benodigde kennis om gewenst bedrag te vertonen (*weten*), of dat ze dit gedrag niet *willen* of *kunnen* vertonen. Des te scherper de probleemanalyse, des te beter kunnen evaluatoren onderzoeken of het ingezette beleid goed gericht is.

8. Eerste lessen op basis van verkennende casussen

De bruikbaarheid van de in dit rapport besproken kaders en methoden, en specifiek de beslishulp uit Hoofdstuk 7, blijkt het best uit een concrete toepassing daarvan. Het ministerie van EZK heeft daarom een inventarisatie gemaakt van casussen die potentieel baat hebben bij de inhoud van dit rapport. Hieruit zijn vier casussen gekozen die verschillen in de mate en manier waarop ze voorbeelden zijn van systeem- en/of transitiebeleid. De variatie op dit punt maakt het mogelijk om te kunnen verkennen wat de kansen en beperkingen zijn van de aangereikte handvatten. De vier geselecteerde casussen zijn:

- Het Missiegedreven Topsectoren- en Innovatiebeleid;
- Het CO₂-reductiebeleid / Klimaatbeleid;
- De Nederlandse Digitaliseringsstrategie;
- Het Techniekpact.

De gedetailleerde casusuitwerkingen, die zijn opgesteld in samenwerking met dossierhouders van EZK, zijn op navraag beschikbaar. Voor dit rapport zijn ze instrumenteel geweest bij het in kaart brengen van overwegingen die evaluatoren als startpunt kunnen nemen bij het opzetten en uitvoeren van evaluaties. In dit hoofdstuk rapporteren we eerst hoe de beslishulp voor ieder van de casussen helpt bij het afbakenen en invullen van evaluatieontwerpen. Vervolgens reflecteren we op enkele lessen die over de vier casussen heen strekken.

8.1. Toepassing beslishulp per casus

Het Missiegedreven Topsectoren- en Innovatiebeleid

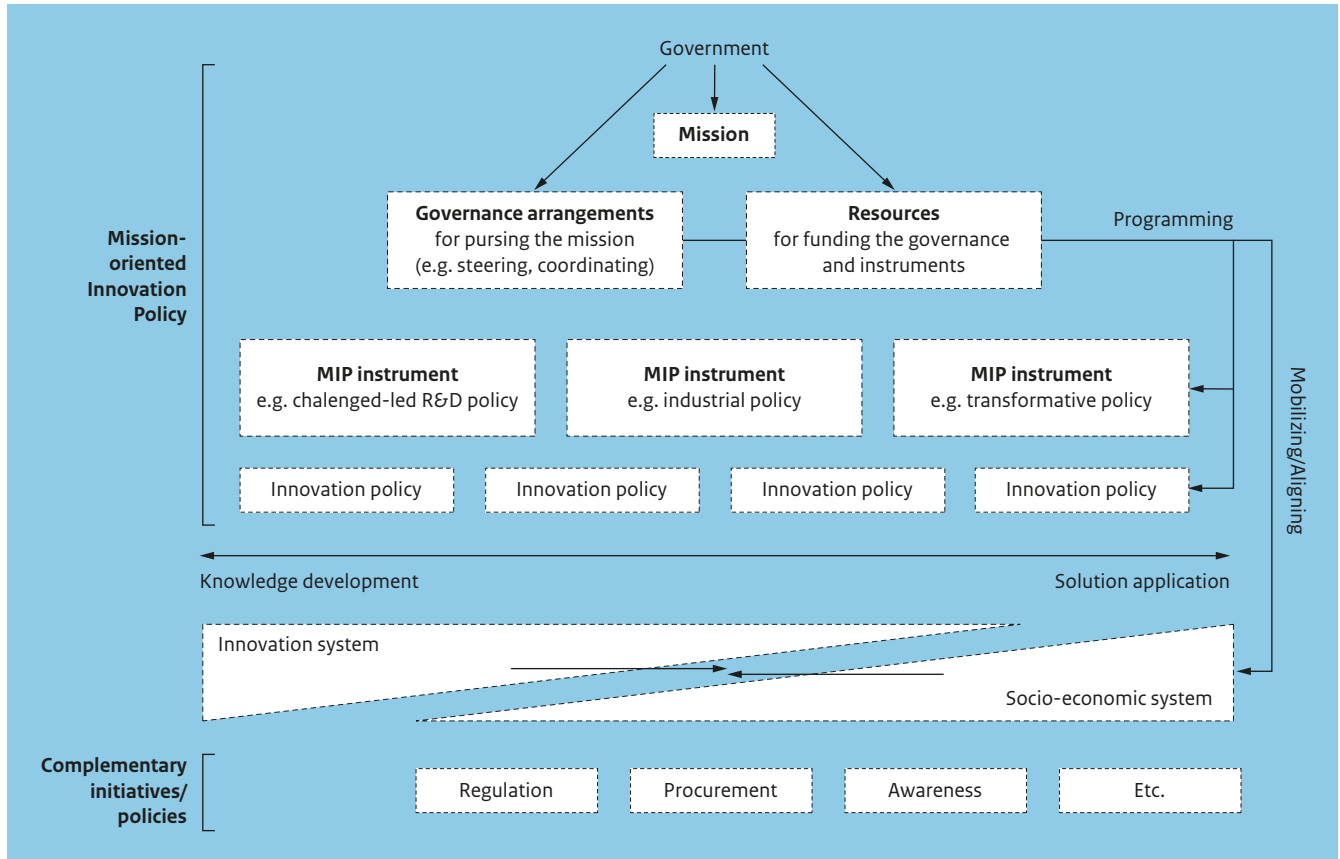
Beschrijving

Het Missiegedreven Topsectoren- en Innovatiebeleid (MTIB) borduurt voort op het in 2012 door de ministeries van EZK en OCW gelanceerde Topsectorenbeleid. Het hart van dat beleid wordt gevormd door de negen Topsectoren; techno-economische domeinen die gekenmerkt worden door een sterk onderzoeks- en innovatieprofiel. Iedere Topsector kent een Topteam met daarin vertegenwoordigers uit de wetenschap, het bedrijfsleven (inclusief een MKB-vertegenwoordiger) en relevante vakdepartementen. Deze Topteams hebben de opdracht om innovatiekansen voor hun domein te vergroten. Hiervoor kunnen zij leunen op de

grotendeels door EZK gefinancierde Topconsortia voor Kennis en Innovatie (TKIs), die fungeren als secretariaten. De TKIs houden zich bezig met zaken als netwerkactiviteiten en roadmap-ontwikkeling, om zo ecosystemen te creëren en boven water te krijgen wat de innovatieambities zijn en welke knelpunten daarbij in de weg staan. Hieruit vloeien ook de Kennis- en Innovatieagenda's (KIAs) van de Topsectoren voort. De Topteams gebruiken de KIAs en onderliggende signalen om invloed uit te oefenen op o.a. onderzoeksfinanciering, regelgeving, exportbevordering, en curriculumontwikkeling. Om onderzoeksprogrammering in het Nederlandse innovatiesysteem te laten aansluiten bij de KIAs worden er eens in de vier jaar Kennis- en Innovatiecontracten of -convenanten (KICs) ondertekend door de Topsectoren, betrokken ministeries, en kennispartners als de TO2-instituten. Daarnaast zijn er enkele regelingen die expliciet ten dienste staan van de Topsectoren en het Topsectorenbeleid. De grootste is de PPS-toeslag, het instrument waarbij TKIs publiek-private R&D-projecten kunnen aanjagen doordat ze 30 procent toeslag ontvangen op de private bijdragen aan dergelijke projecten. Voorwaarde is wel dat zowel de toeslag-genererende 'grondslag-projecten' als de 'inzet-projecten' (waar de toeslag voor gebruikt wordt) passen binnen de KIAs. Om ook projecten van het MKB beter te laten aansluiten bij de KIAs is er een aparte regeling; de MKB-innovatiestimulering Regio en Topsectoren (MIT). Activiteiten die hiermee bekostigd worden zijn o.a. kennisvouchers, haalbaarheidsprojecten, R&D-samenwerking, en TKI-Innovatiemakelaars.

Het uiteindelijke doel van de oorspronkelijke Topsectorenaanpak was om private bijdragen aan publiek-privaat onderzoek te vergroten, en op termijn de concurrentiekracht van Nederland te versterken. In de loop der jaren is er echter gaandeweg meer nadruk te komen liggen op mogelijkheden om de collectieve innovatie-inspanningen sterker te richten op maatschappelijke uitdagingen. Hierbij speelde onder andere mee dat het Europese R&D-beleid thema's als 'grand societal challenges' en 'societal missions' centraal is gaan stellen, en dat enkele vakdepartementen meer te zeggen wilden krijgen over innovatierichtingen in de voor hen relevante Topsectoren. Bij het ontwikkelen van de KIAs voor 2018-2021, halverwege 2017, zijn de Topsectoren gevraagd deze op te stellen langs maatschappelijke uitdagingen en

Figuur 8.1: Opzet van het MTIB, op hoofdlijnen (bron: 'Post-commencement assessment MTIP'. Janssen, 2020)



sleuteltechnologieën.²² Halverwege 2018 kondigde EZK aan dat deze thema's leidend zouden zijn voor een nieuwe generatie Topsectorenbeleid.²³ Vanaf 2019 is deze beweging doorgezet met de kanteling naar het Missiegedreven Topsectoren- en Innovatiebeleid.²⁴ Leidend hierbij zijn de 25 missies behorende bij vier overkoepelende missie-thema's (zie Appendix), in combinatie met een beleidsspoor ter bevordering van sleuteltechnologieën.

Zowel qua governance als beleidsinstrumentarium is het MTIB stevig in het Topsectorenbeleid verankerd. De kanteling naar het MTIB impliceert dat er in die bestaande structuren wijzigingen worden doorgevoerd, zoals het instellen van 'Missie-teams'. Bij missies die sterk leunen op één Topsector zijn deze teams naast de Topteams komen te staan, terwijl ze bij bijvoorbeeld de missie rondom Circulaire Economie (die voor innovaties bij meerdere Topsectoren terechtkan) een relatief nieuw orgaan vormen. Het inrichten van het MTIB is per 2020 in gang gezet, waarbij de snelheid van de ontwikkelingen varieert tussen de diverse missies.

Bovenstaande figuur geeft op hoofdlijnen de beleidsopzet van het MTIB weer:

De essentie van het MTIB is dat er, net als bij de Topsectorenaanpak, governance-structuren ingericht zijn die innovatie-inspanningen programmeren. Hiervoor maken die structuren primair gebruik van bestaande instrumenten van EZK en andere overheden. De ambitie is om een betere afstemming te bewerkstelligen tussen investeringen in onderzoek en innovatie enerzijds, en inspanningen om maatschappelijke problemen te verhelpen anderzijds. Door de twee bijbehorende systemen meer met elkaar in verband te brengen ontstaat er naar verwachting een beter perspectief voor innovatieve oplossingen waar missiedoelen bij gebaat zijn.

Afwegingen o.b.v. de beslihsulp

Hieronder lopen we stap voor stap de vragen uit het stroomschema af. Het vertrekpunt is dat we te maken hebben met beleid waarmee meerdere instrumenten samen één doel nastreven.

Zijn de instrumenten synergetisch?

De in het MTIB beoogde synergiën betreffen de versteviging van systeemfactoren die een voorwaarde zijn voor het succes van specifieke innovatiepaden (namelijk de paden waarvan de

²² Kennis- en Innovatieagenda 2018-2021: Maatschappelijke uitdagingen en sleuteltechnologieën.

²³ Ministerie van EZK (13-07-2018). Kamerbrief over innovatiebeleid en de bevordering van innovatie: naar missiegedreven innovatiebeleid met impact.

²⁴ Ministerie van EZK (26-04-2019). Kamerbrief over Missies voor het topsectoren- en innovatiebeleid.

coördinerende gremia menen dat die kunnen bijdragen aan het behalen van missiedoelen). Het is mogelijk om alleen kennisontwikkeling of marktvorming te versterken voor een dergelijk pad, maar zeker bij 'wicked' maatschappelijke uitdagingen en transities is het uitgangspunt dat nieuwe bewegingen pas loskomen als vele lichten tegelijkertijd op groen worden gezet. Dit is ook de centrale gedachte bij het nastreven van een continuüm in de verstrekte beleidsondersteuning; beleidsprykkels moeten elkaar zodanig complementeren dat hindernissen voor innovatiepaden worden weggenomen. Daarbij kunnen het verhelpen van diffusie-obstakels, zoals belemmerende regelgeving, een *demand-pull* uitoefenen op innovatieactiviteiten die zich nog in een vroeg ontwikkelingsstadium bevinden. Het is dus niet zomaar mogelijk is om projecten uit één aan het MTIB gelieerde regeling te vergelijken met thematisch gelijksoortige innovatieprojecten die geen steun hebben gehad vanuit die regeling. Immers, die laatste innovatieprojecten zouden ook voordeel gehad kunnen hebben uit de gunstigere omstandigheden voor innovatie op dat thema.

Is er een meetbaar beleidsdoel?

Het MTIB als zodanig heeft niet een concreet einddoel, aangezien het vooral voorwaarden schept waarmee een diverse set van publieke en private partners samen kan werken aan het behalen van de 25 missies. Veel van die missies hebben ieder voor zich wel een concreet einddoel, met zelfs een specifiek moment waarop dat doel bereikt moet zijn. Voorbeelden zijn "Een CO₂-vrije gebouwde omgeving in 2050" en "In 2030 wordt zorg 50 procent meer (of vaker) in de eigen leefomgeving georganiseerd, in plaats van in zorginstellingen". Bij andere missies is de doelformulering relatief ambigu, zoals "In 2035 beschikt Nederland over de marine van de toekomst, die in staat is om flexibel te reageren op onvoorspelbare ontwikkelingen".

Het is hoe dan ook aan te bevelen om voor iedere afzonderlijke maatschappelijke missie te onderzoeken of het mogelijk is de linkerkant van het beslissings-stroomschema te volgen. Voor het MTIB zelf, dat dient ter ondersteuning van de ontwikkeling en adoptie van innovaties waar de missiedoelen bij gebaat kunnen zijn, valt deze optie af bij gebrek aan een concreet meetbaar doel. Daarmee resteren er twee opties. Er kan gesteld worden dat er géén beleidsdoel is (hetgeen direct hieronder verder uitgewerkt wordt), maar het is ook mogelijk om te beweren dat het MTIB gericht is op het aanjagen van 'het verschuiven van mensen en middelen naar gebieden waar ze het meeste impact hebben (= thematische convergentie in innovatie-inspanningen)'. Dergelijke convergentie is een uniform tussendoel als opstap naar het bereiken van missie-specifieke doelen. Welke methoden zich lenen om een het tussendoel van thematische convergentie te evalueren bespreken we aan het eind van deze casus.

Zijn de systeemkelpunten bekend?

De vraag die volgt is of systeemkelpunten bekend zijn. Het antwoord hierop hangt wederom af van hoe het MTIB, als

overkoepelend kader voor het prioriteren van de missies en het organiseren van structuren voor het najagen van die missies, precies geïnterpreteerd wordt. In een brede interpretatie heeft het MTIB betrekking op het gehele instrumentarium om missiedoelen te bereiken (inclusief het beleid van departementen die zich aan die doelen geëngageerd hebben), en dient het dus ook daarop afgerekend te worden. Een smalle interpretatie van het MTIB is dat het vakdepartementen in positie heeft gebracht om missies te formuleren en na te jagen, maar dat het MTIB vervolgens 'zelf' enkel gaat over het aanleveren van potentieel geschikte oplossingsrichtingen. In dat perspectief is het slechts een onderdeel van het Nederlandse missiebeleid, waarbij dit laatste ook (beleids) inspanningen omvat die niet rechtstreeks met het Topsectoren en innovatiebeleid van EZK te maken hebben – maar bijvoorbeeld wel met wetgeving en diffusiebeleid vanuit de vakdepartementen. Het feit dat het voorlopige monitoringskader voor het MTIB zich op techno-economische indicatoren richt suggereert dat er daar van de smalle interpretatie wordt uitgegaan.

Als we de smalle interpretatie van het MTIB nemen is het voornaamste knelpunt het tekort aan coherentie in de onderzoeks- en innovatie-inspanningen in Nederland. In de literatuur over rationales voor (transformatief) innovatiebeleid wordt dit ook wel '*directionality failure*' genoemd. Het MTIB, de Nederlandse variant op *mission-oriented innovation policy*, is een poging dit probleem te verhelpen door urgente maatschappelijke uitdagingen aan te grijpen als richtingen voor het bundelen van innovatie-investeringen en capaciteiten en het uitlokken van multidisciplinaire cross-sectorale innovaties.²⁵ De in stelling gebrachte structuur dient zowel voor het concretiseren van de richtingen (inclusief de vertaalslag van probleem naar oplossingsrichtingen) als voor het continu bijstellen van beleidsimpulsen. Dit laatste is extra relevant op het moment dat oplossingsrichtingen in hun ontwikkeling tegen steeds weer andere problemen aanlopen, hetgeen kenmerkend is voor innovaties die transities teweeg moeten brengen.

Hoewel het in abstracte zin dus mogelijk is om te benoemen welke failure ten grondslag ligt aan het MTIB, is het niet eenvoudig om concreter te worden in termen van systeemkelpunten. Waarschijnlijk bestaan er tussen de diverse missies grote verschillen in welke processen (e.g. kennisontwikkeling, human capital opleiden, wettelijke weerstanden wegnemen) de ontwikkeling en toepassing van innovatieve oplossingen vooral tegenhouden. De kamerbrieven over het MTIB benoemen in ieder geval geen algemene terugkerende problemen die de voortgang daarin belemmeren. Als gevolg leidt het stroomschema dus naar het advies om eerst de governance en beleidsmix beter te onderzoeken. Voor de volledigheid vermelden we weer dat deze redenering uitgaat van de vraag om het MTIB als geheel te evalueren. Indien voor afzonderlijke missies de systeemkelpunten bekend zijn, dan is de volgende vraag of het mogelijk is om de bijbehorende systemen goed af te bakenen.

²⁵ Mazzucato, M. (2018). Mission-oriented innovation policies: challenges and opportunities. *Industrial and Corporate Change*, 27(5), 803-815.

Dit lijkt realistisch voor concrete missies rondom bijvoorbeeld de Nederlandse marine of het verduurzamen van de gebouwde omgeving. In die gevallen leidt de beschrijving naar ‘Systeemanalyse op basis van kritieke processen’, waarvoor gebruik kan worden gemaakt van analytische kaders als dat voor Technologische Innovatiesystemen (zie Sectie 4.4) of de op missies aangepaste variant van Missiegedreven Innovatiesystemen.²⁶ Deze analyses laten we hier echter buiten beschouwing, mede omdat de evaluatie van de Topsectorenaanpak hier reeds een voorbeeld van is; zie box 2 in Sectie 3.5. In die evaluatie is impliciet gebruikgemaakt van contribution analysis, aangezien er kwalitatieve en kwantitatieve indicaties verzameld zijn om vast te stellen in welke mate TIS-processen versterkt zijn.

> ‘Bestudeer governance en beleidsmix’

Bij het onderzoeken van de governance van het MTIB is de vraag of de geïntroduceerde coördinatiestructuren (themateteams, missieteams, TKIs etc.) geschikt zijn om imperfecties te verhelpen in de mix van beleids- en andersoortige prikkels, en zo de ontwikkeling en uitrol faciliteren van innovaties waarmee ambitieuze missiedoelen behaald kunnen worden.

De toegevoegde waarde van het onderzoeken van governance-aspecten is hier om een wat rijker inzicht te creëren in de kwaliteit van de beleidsuitvoering en de belemmeringen die daarin spelen. Dergelijke resultaten geven geen uitsluitel over het effect van het beleid, maar helpen om te begrijpen wat er goed en fout gaat in de sturing en controle – en wat er beter kan. Bij casussen als het MTIB geldt bovendien dat onderzoek naar hoe er gecoördineerd en gestuurd wordt een basis vormt

om vervolgens te analyseren wat er met die sturing gebeurt (zie verderop). Immers, als niet duidelijk is welke accenten er met het beleid gelegd worden is het ook niet mogelijk om te toetsen welke bewegingen dit in gang zet.

Het onderzoeken van de governance vraagt niet zozeer om het vinden van de juiste empirische methoden (het komt veelal neer op interviews en deskresearch), maar vooral om het stellen van de juiste vragen. Zoals uitgelegd in Sectie 4.4 is evaluatie langs de invalshoek van governance (en de daaruit volgende beïnvloeding van de beleidsmix) gebaat bij een ‘toetsingskader’ dat uiteenzet aan welke beginselen het beleid moet voldoen. De crux is in dit geval dus het identificeren van beleidsprincipes die kunnen dienen als basis voor uitspraken over de mate waarin het MTIB qua governance goed ontworpen is.

Er zijn twee routes om die principes helder te krijgen: via coproductie, door met het veld (inclusief de ontwerpende en uitvoerende beleidsmakers) in gesprek te gaan, en via het

raadplegen van relevante wetenschappelijke literatuur. De methode van **lerend evalueren** combineert die twee routes, aangezien literatuuranalyse de basis kan vormen voor het ontwikkelen van een theoretisch kader (met principes) waarmee beleid getoetst kan worden. In het geval van missiebeleid als het MTIB lijkt het zinnig om in ieder geval een beroep te doen op de academische discussie rondom ‘mission-oriented innovation policies’ die de afgelopen jaren weer flink op gang is gekomen.²⁷ Op basis daarvan ligt het voor de hand om in de evaluatie van de governance-processen aandacht te besteden aan elementen die van groot belang zijn als actoren uit het innovatiesysteem en uit maatschappelijke sectoren samen een impuls moeten geven aan de ontwikkeling en toepassing van innovatieve oplossingen voor de problemen waar de missiedoelen van het MTIB op gebaseerd zijn. Hieronder volgen enkele vragen die daarbij van pas kunnen komen. Merk op dat sommige vragen uitgaan van de veronderstelde noodzaak van een bepaald governance-aspect (e.g. er moeten *checks & balances* zijn), terwijl andere vragen op dit moment niet verder kunnen gaan dan blootleggen wat de voor- en nadelen zijn van een zeker governance-aspect (e.g. is het wenselijk of onwenselijk als overheden bestaand i.p.v. nieuw beleid op missies richten?).

Voorbeeldvragen (indicatoren) voor onderzoek naar de governance van het MTIB:

- “Heeft het MTIP geleid tot het ontstaan van governance-structuren die geschikt zijn om innovatie- en invoeringsinitiatieven op elkaar af te stemmen?” Geschikt verwijst daarbij naar principes zoals vastgesteld o.b.v. de theorie en/of in interactie met het veld (zie hierboven).
 - Bevat de structuur elementen die zijn toegerust voor het verzamelen van informatie met betrekking tot opkomende innovatiekansen en missie-gerelateerde problemen?
 - Omvat de structuur elementen die in staat zijn deze informatie te combineren en beslissingen te nemen over de te volgen koers? Hoe selectief zijn de programma’s (b.v. MMIP’s) die ontstaan? En hoe gedragen zijn ze?
 - Bevat de structuur goed functionerende “checks and balances” ter voorkoming van een al te sterke vastlegging van belangen?
 - Zijn er indicaties dat de door het MTIB verbonden partijen zulke sterke gezamenlijke plannen ontwikkelen dat ze hier aanvullende steun voor denken te kunnen vinden (bijv. bij het Groeifonds)?
- Zijn er aanwijzingen dat het MTIP aanslaat wat betreft het mobiliseren van partners? Dit betreft de daadwerkelijke *buy-in* van lijnministeries, regionale overheden, kennispartners en cetera. Door het ondertekenen van de KIC hebben veel partijen al toegezegd middelen in te zetten voor de missie.

²⁶ Hekkert, M. P., Janssen, M. J., Wesseling, J. H., & Negro, S. O. (2020). Mission-oriented innovation systems. *Environmental Innovation and Societal Transitions*, 34, 76-79.

²⁷ Voor een beknopte review, zie: Janssen, M., Torrens, J., Wesseling, J., & Wanzenböck, I. (2021). The promises and premises of mission-oriented innovation policy: A reflection and ways forward. *Science and Public Policy*, 48 (3), 438-444.

- In hoeverre hebben die partijen bestaande budgetten opgesomd die relatief los van het MTIP en MMIPs worden ingezet? Zijn er aanwijzingen dat de verschillende partners werkelijk bereid zijn hun budgetten te bundelen tot een alomvattende en consistente MIP-aanpak (bijvoorbeeld zoals bij gezamenlijk gefinancierde beleidsregelingen als het MOOI?). Stijgen de jaarlijkse budgetten?
- Volgen de vakministeries de ontwikkelingen op het gebied van het MMIP actief, als onderdeel van een geformaliseerde beleidscyclus?
- “Berust de ‘uitgebreide’ MTIP-strategie (waarbij ook initiatieven van vakministeries worden betrokken) op een beleidsmix en financieringsstromen die geschikt zijn om het gehele spectrum van innovatieontwikkeling en -toepassing te ondersteunen?”.
 - Is de programmering van subsidies voor wetenschappelijk onderzoek afgestemd op de kennis die nodig is in (PPP-) O&O-projecten die betrekking hebben op bepaalde innovaties? Is het innovatiebeleid voor de middelste TRL’s gesynchroniseerd met initiatieven ter ondersteuning van de invoering?
 - In hoeverre is het problematisch als er discrepanties bestaan; belemmert het de continuïteit, en/of zorgt het voor een kritische herijking van de innovatietrajecten die worden nagestreefd?
 - Zijn er aanwijzingen (bv. uit gegevens over programmering en projectportefeuilles) dat de beleidsmix het ook mogelijk maakt om instrumenten met een laag TRL aan te passen als reactie op knelpunten die zich voordoen bij innovatieactiviteiten met een hoog TRL?

Zoals hierboven aangegeven is het ook mogelijk een andere afslag te nemen in het stroomschema, uitgaande van de interpretatie dat het MTIB met haar focus op ‘thematische convergentie in innovatie-inspanningen’ wel degelijk een duidelijk tussendoel heeft. Dit zou leiden naar een ander deel van de beslishulp:

> Analyseer structuurverandering

Als het doel is om meer focus en versnelling aan te brengen in innovatiepaden met potentie voor missies, dan dient effectmeting uit te wijzen of er inderdaad meer momentum ontstaat voor specifieke innovatierichtingen. De hamvraag is dan niet of er *binnen* de gemobiliseerde regelingen meer aandacht komt voor die innovatierichtingen, maar of al die beleidsprykkels en innovatieprojecten zoveel in gang zetten (“systeemversterking” bewerkstelligen) dat individuele organisaties van *buiten* de direct bereikte populatie ook aanhaken en investeren in die innovatierichtingen.

De meest veelbelovende manier om dergelijke ‘crowding in’ te observeren is door te kijken naar de mate waarin innovatieprojecten in de WBSO-regeling zich naar die richtingen voegen. Aangezien de WBSO een relatief generieke innovatieregeling is, zonder inhoudelijke sturing, zou het effect van het MTIB kunnen blijken uit een geobserveerde toegenomen bereidheid

van bedrijven om R&D te verrichten op het vlak van producten en processen die passen bij de beleidsmatig gesteunde innovatiepaden. Is een dergelijke toename niet waarneembaar, dan is het niet waarschijnlijk dat het MTIB veel impact heeft gehad. Wanneer de toename wél waarneembaar is dient er uiteraard zo goed mogelijk gecorrigeerd te worden voor een autonome trend, bijvoorbeeld door een vergelijking te maken met de ontwikkeling van WBSO-projecten op thema’s die óók raken aan de maatschappelijke uitdagingen van de missies, maar die juist niet uitverkoren zijn in de agenda’s en programma’s (KIAs en MMIPs) om de missiedoelen te behalen. In het verlengde van bovenstaande meting, i.e. een analyse van toenames van private en toegepaste innovatieactiviteit op MTIB-innovatierichtingen, kan er ook gekeken worden naar de bedrijfseconomische prestaties van de partijen met WBSO-projecten in het verlengde van de geprioriteerde innovatierichtingen.

Kort gesteld lijkt het dus mogelijk om het onderzoek naar structuurveranderingen uit te voeren op basis van empirische methoden links in de beslishulp (quasi-experimenteel of middels regressies). De werkingsmechanismen zijn voldoende bekend, er zijn data beschikbaar die aangeven op welke thema’s innovatieprojecten binnen en buiten het MTIB betrekking hebben, en daardoor kan er ook een counterfactual aangewezen worden. Dat laatste vereist dus wel dat heel duidelijk moet zijn wat de in het MTIB geprioriteerde oplossingsrichtingen zijn, en hoe innovatieprojecten binnen de daaraan gelieerde regelingen (NWO KIA calls, PPS-toeslag, MIT, MOOI) en buiten die regelingen (WBSO) zich tot die richtingen verhouden. Voor de ultieme vergelijking dient het labelen van projecten heel precies te gebeuren, omdat dit het mogelijk maakt om te kijken of wél gekozen innovatierichtingen op het gebied van bijvoorbeeld duurzame industrie zich nu sneller en beter ontwikkelen dan de niet gekozen richtingen. Veelbelovend op dit punt is het fundamentele uitgangspunt van de MTIB om duidelijke richtingen leidend te laten zijn voor het bundelen van innovatie- en diffusie-inspanningen. De missies zelf zijn enkel de ambitieuze doelen, maar de governance-structuur en beleidsinspanningen en -instrumenten zijn erop gericht momentum te geven aan de meest veelbelovende oplossingsrichtingen. Die richtingen worden expliciet gedocumenteerd in de KIAs en MMIPs en dienen als basis voor de programmering van diverse regelingen.

Als bij de evaluatie niet duidelijk is om welke richtingen het gaat, is dat bij voorbaat al een zeer slecht teken voor de effectiviteit van het beleid. De implicatie voor het selecteren van empirische methoden zou bovendien zijn dat er teruggevallen moet worden op alternatieven die een lagere causale bewijskracht hebben. Het stroomschema verwijst in dit geval naar contribution analysis of process tracing. De toepassing daarvan zou neerkomen op het gestructureerd nalopen van indicaties die aangeven dat er sprake is van een serie van gewenste ontwikkelingen, bestaande uit in ieder geval: meer thematische en organisatorische/administratieve samenhang tussen potentieel complementaire beleidsinstrumenten en -inspanningen; thematische convergentie van de

door diverse regelingen gesteunde innovatieprojecten (in relatie tot de MMIPs) en een versnelde en meer succesvolle ontwikkeling van de aan de MMIPs gerelateerde projecten (in termen van patenten, publicaties, prototypes), idealiter ook ten opzichte van andere innovatieprojecten.

Het Klimaatbeleid

Omschrijving van het beleid

Het doel van het Klimaatbeleid is het reduceren van de uitstoot van broeikasgassen. In het recente coalitieakkoord zijn aangescherpte doelen geformuleerd, en wordt in 2050 klimaatneutraliteit beoogd. Er is een tussendoel gedefinieerd van 55 procent reductie in 2030. Deze doelen zijn vastgelegd in de Klimaatwet. Op basis van deze wet is het Klimaatplan 2021-2030 opgesteld. Dit Klimaatplan wordt voor een belangrijk deel bepaald door de hoofdlijnen uit het Klimaatakkoord. Het Klimaatplan geeft een integraal beeld van het klimaatbeleid van het kabinet in Nederland.

Het einddoel vraagt niet om één transitie, maar eigenlijk om een veelvoud aan transities tegelijkertijd in de diverse sectoren en productie-consumptiesystemen. Deze transities verlopen ook niet allemaal gelijk, en kennen verschillende ‘knoppen’ waaraan gedraaid moet worden. Wat werkt in de ene transitie, werkt niet in de andere.

Om deze transities te realiseren zijn er geen *quick fixes*. Fossiele energie zit diep verweven in de socio-technische en productie-consumptiesystemen waar de Nederlandse maatschappij op leunt. Klimaatneutraliteit vraagt om het transformeren van deze systemen, waar technische, sociale en organisatorische innovatie voor nodig is. Het innovatiebeleid is erop gericht deze maatschappelijk wenselijke innovaties te realiseren en zo transformaties (langdurende processen) te ‘voeden’.

Het Klimaatbeleid is belegd bij een aantal departementen, onder uiteindelijke verantwoordelijkheid van het ministerie van EZK. Het Klimaatbeleid betreft een veelvoud aan regelingen, instrumenten, wetten en activiteiten. Hoewel in deze illustratie het innovatie-instrumentarium van EZK centraal staat, zijn er in (soms sterke) raakvlakken met andere instrumenten. Enkele voorbeelden van instrumenten zijn:

- Energie-innovatieregelingen, TSE, MOOI, DEI+, HER en VEKI;
- IKIA Klimaat en Energie met bijbehorende MMIP's;
- Energie Investeringsaftrek;
- SDE++ als subsidie voor hernieuwbare elektriciteitsprojecten;
- Opzet en uitvoering van de Regionale Energiestrategieën (RES);
- Programma ‘Kas als Energiebron’;
- Uitvoering van de Nationale Agenda Laadinfrastructuur;
- Green Deal verduurzaming vervoer over water;

- Aanpassing Wet Milieubeheer (bedrijven moeten energiebesparingsmaatregelen nemen die zichzelf binnen vijf jaar terugverdienen);
- Programma Aardgasvrije Wijken (PAW);
- Verbod op het verbruik van kolen bij de productie van elektriciteit vanaf 2030.

Ook Europese beleidsinstrumenten spelen een rol in het Nederlandse Klimaatbeleid, zoals het ETS (waar Nederland nog aanvullend beleid op voert, zoals via de Subsidieregeling Indirecte Emissiekosten ETS en de CO₂-heffing). Daarnaast raakt het Klimaatbeleid aan een aantal andere overkoepelende beleidsdossiers: ofwel omdat vanuit de dossiers voor een deel elkaars beleid wordt ingevuld, ofwel omdat er spanning zit tussen dossiers.

Er is een aantal redenen om als overheid te interveniëren in de klimaattransitie. Deze redenen leggen op hoofdlijnen de knelpunten bloot die opgelost dienen te worden. Vanuit het perspectief van systeem- en transitiebeleid benoemen we de belangrijkste:

- *Richting* – De doorvertaling van het klimaatdoel binnen sectoren is essentieel om te voorkomen dat er veranderingsprocessen in gang gezet worden die niet op elkaar aansluiten – of dat de investeringen in die veranderingen juist uitblijven omdat niet duidelijk is op welk nieuw systeem partijen dienen in te spelen (*directionality failure*). Het stellen van duidelijke doelen en definiëren van oplossingsrichtingen dient dat op te lossen. Geformuleerde subdoelen dienen op hun beurt geconcretiseerd worden in bij specifieke partijen belegde acties die meetbaar bijdragen aan het gestelde doel;
- *Snelheid* – Het ingrijpend veranderen van productie-consumptiesystemen gaat te traag zonder overheidsingrijpen. Het belang van snelheid wordt duidelijker als we bedenken dat er ‘tipping points’ bereikt kunnen worden in temperatuurstijging, waarbij gevolgen onomkeerbaar zijn geworden;
- *Beleidscoördinatie* – Veranderingen in productie-consumptiesystemen grijpen op zoveel onderling gerelateerde vlakken aan (e.g. in met elkaar verweven waardeketens waarin gesynchroniseerde aanpassingen dienen plaats te vinden in productie, transport en gebruik van goederen en diensten), dat coördinatie van beleid nodig is.²⁸ Dit vergt nauwe en continue aansluiting tussen beleid en praktijk, en beleidsterreinen onderling, om de praktische knelpunten die zich voordoen te adresseren;
- *Ontbrekende systeemfuncties* – Ook geldt dat innovatiesystemen die ten dienste staan van klimaat-neutrale productie-consumptiesystemen niet altijd goed functioneren. Bijvoorbeeld

²⁸ Dit hangt samen met de gekozen veelkleurige instrumentenmix en staatsinrichting. Naarmate er minder instrumenten ingezet worden, en er meer centrale sturing c.q. invloed is, is minder beleidscoördinatie nodig. Overigens hangt dit samen met potentieel *overheidsfalen*, wat zich voordoet als de overheid niet de capaciteit en kennis heeft om (complex) beleid goed te coördineren.

vanwege beperkende 'fossiele' regelgeving of een gebrek aan ondernemerschap. In het veranderende systeem spelen op onderdelen ook markt-falen, b.v. omdat negatieve effecten niet worden beprijsd of zelfs omdat niet-duurzame goederen (zoals fossiele energie) gesubsidieerd zijn.

Naast het uitvoeren van allerlei instrumenten ter bevordering van (de adoptie van) innovatie, behelst het Klimaatbeleid sturingsmechanismen om coördinatie en afstemming in de transitie in te brengen.

Afwegingen o.b.v. de beslijsulphulp

Voor het Klimaatbeleid doorlopen we de stappen van de beslijsulphulp uit hoofdstuk 7. Bij elk van de stappen c.q. vragen geven we enkele afwegingen mee. Het startpunt is dat we te maken hebben met beleid waarin een veelvoud aan instrumenten en activiteiten één beleidsdoel nastreven.

Zijn de instrumenten synergetisch?

Eén van de doelen van het overkoepelende klimaatbeleid is het introduceren van coördinatie van beleid, instrumenten en activiteiten. Bestaande en nieuwe instrumenten die innovatie stimuleren, zijn onder de noemer klimaatbeleid gebracht. Verschillende van deze instrumenten waren al synergetisch in opzet (b.v. de HER en SDE+). Daar zijn andere instrumenten, ook van andere departementen, aan toegevoegd. Zo ontstaat er een palet aan instrumenten en activiteiten dat tot doel heeft om innovaties van idee tot implementatie te ondersteunen, waarbij ruime aandacht is voor zgn. co-innovatie (b.v. organisatorische of sociale innovatie die óók nodig is).

De instrumenten lijken qua intentie op hetzelfde doel gericht te zijn. Voor de meeste instrumenten is echter nog onduidelijk óf en hoe synergie (in beleidsopzet) gerealiseerd is. Daarom is aan te bevelen om te kijken hoe dit precies uitpakt. Voor het klimaatdossier ligt het dus voor de hand om allereerst te onderzoeken hoe processen werken om synergie te bereiken (governance) en óf instrumenten in opzet ook synergetisch zijn (beleidsmix-analyse); zie volgende pagina.

Bovenstaande laat onverlet dat een aantal van de instrumenten ook individueel geëvalueerd zal moeten worden op doeltreffendheid en doelmatigheid. Vanuit het ministerie van EZK zijn er stappen ondernomen om individuele instrumentevaluaties op zo'n manier uit te voeren dat de onderlinge vergelijkbaarheid vergroot wordt. Zo kunnen individuele evaluaties bijdragen aan een evaluatie op systeemniveau.

Als bekend is of de governance bijdraagt aan synergie in instrumenten en activiteiten, en instrumenten daadwerkelijk synergetisch zijn in opzet, dan kan de beslijsulphulp opnieuw doorlopen worden. In dit voorbeeld lopen we daarom ook een aantal vervolgvragen langs.

Is er een meetbaar beleidsdoel?

Het Klimaatbeleid kent meetbare en kwantificeerbare beleidsdoelen voor 2030 en 2050. Ook het innovatie-instrumentarium heeft hier een aandeel in, zij het vooral in de fases van de transitie waarin innovatie een belangrijke driver is om zo'n fase door te komen. Het is aan te bevelen deze beleidsdoelen nauwgezet te monitoren.

Daarnaast zijn er ook tussendoelen voor de verschillende productie- en consumptiesystemen, bijvoorbeeld dat alle nieuwe personenauto's in 2030 voor 100 procent elektrisch moeten zijn, dat het elektriciteitsnetwerk flexibel moet worden, of dat de industrie moet kunnen fungeren als een buffer voor energienetwerken. Voor dergelijke productie- en consumptiesystemen kan bezien worden of het innovatiesysteem optimaal functioneert, en of er daadwerkelijk in de structuur van productie- en consumptiesystemen veranderingen optreden.

Vaststellen of die veranderingen veroorzaakt worden door beleid vergt dat er steeds geïdentificeerd wordt welke van de vele mogelijke beleidsinstrumenten van invloed zijn geweest op het systeem in kwestie. Hierbij kan het zo zijn dat sommige instrumenten aangrijpen op meerdere systemen. Dat wil niet zeggen dat het evalueren van die afzonderlijke instrumenten ook volstaat; de vraag is immers of de interactie van al die instrumenten (en bijkomende coördinatie) een systeem werkelijk versterkt of juist verzwakt. Hoe een evaluatie-ontwerp er precies uit dient te zien, en welke methoden daarbij benut kunnen worden, verschilt waarschijnlijk van systeem tot systeem. We werken dat in deze beknopte analyse niet verder uit.

Zijn de systeemknelpunten bekend? Is het systeem goed af te bakenen?

> Systeemanalyse

Wanneer er géén heldere tussendoelen zijn kan een evaluatie zich toeleggen op het onderzoeken van de relatie tussen beleid en het verhelpen van knelpunten die transitiestappen hinderen. De vraag is allereerst of die knelpunten reeds bekend zijn. In het kader van de monitoring van het Klimaatbeleid werkt EZK aan studies waarbij er voor diverse 'sectoren' (met hun productie-consumptie-systemen) onderzocht wordt hoe die ervoor staan. Daarbij is een werkbare afbakening van sectoren gevonden die het mogelijk maakt om te onderzoeken of er na verloop van tijd ook verbeteringen plaatsvinden. De beslijsulphulp suggereert om dit te doen door middel van systeemanalyses, bijvoorbeeld op basis van het analytisch kader waarin de verschillende sleutelprocessen van technologische innovatiesystemen naast elkaar gezet zijn.

Qua methoden ligt het voor de hand om hier te kiezen voor *contribution analysis* of (op het moment dat er a.d.h.v. de beleidstheorie ook concrete hypothesen gevormd kunnen worden) *process tracing* op basis van empirisch materiaal dat inzage kan geven in alle activiteiten die er in Nederland worden ingezet op de geselecteerde sectoren / systemen. Hierbij kan gekeken worden naar de gegevens die RVO bijhoudt bij het uitvoeren van veel van

de eerdergenoemde regelingen. Specifiek interessant is daarbij de samenhang door die regelingen heen, aangezien sommige regelingen meer zicht geven op kennisontwikkeling, en andere juist op samenwerking, het vermarkten van duurzame innovatie, het integreren van deeloplossingen etc. Door de administratie van verschillende regelingen samen te brengen kan er een integraal beeld ontstaan van welke actoren/netwerken samen het systeem vormen, en welke accenten er gelegd worden in het versterken van die systemen. De essentie van contribution analysis en process tracing is vervolgens om, ook op basis van aanvullend materiaal, stapsgewijs na te gaan welke data iets zeggen over de sequentie van veranderingen die zich zou moeten voordoen. Als er voor bijvoorbeeld het systeem rondom aanpassing van het elektriciteitsnetwerk een knelpunt is in de vorm van gebrek aan kennis, dan moeten we zien dat er extra middelen zijn gegaan naar het specifieke onderwerp waarover kennis ontbreekt, dat er ook extra projecten op zijn ontstaan, en dat dit weer geleid heeft tot concrete kennisresultaten (artikelen, rapporten, patenten etc.) op dat onderwerp. Voor een knelpunt rondom bijvoorbeeld gedrag is de causale keten waarschijnlijk ingewikkelder, hetgeen benadrukt hoe belangrijk het is om eerst te reconstrueren via welke stappen interventies en resultaten uiteindelijk dienen op te tellen naar impact.

> 'Bestudeer governance en beleidsmix'

Op het moment dat de beleidstheorie niet scherp in beeld is, en dus niet bekend is hoe beleid moet bijdragen aan het verhelpen van knelpunten, dan helpt het wellicht om eerst beter te doorgronden hoe de governance en het vormgeven van de beleidsmix überhaupt georganiseerd zijn – inclusief welke overwegingen daarbij leidend waren, en of die ook geïmplementeerd zijn. In de monitoringsopzet van het Klimaatbeleid is er ook een onderdeel dat vraagt om inzicht in de vraag of het onderzoeken van (en reageren op) knelpunten wel goed georganiseerd is. Hierin inzicht verschaffen vraagt om een analyse van de governance van de klimaattransitie en van de mix van beleidsinstrumenten die hier een impuls aan beogen te geven. Vragen die hierbij aan de orde komen, zijn: Verlopen transitieprocessen adequaat (governance, samenwerking, doorzettingsmacht, afstemming beleid etc.)? Zijn de processen in lijn met principes? Zijn de plannen en instrumenten coherent en consistent?

Bij het evalueren van de **governance** is de vraag of er sturingskaders en -principes zijn opgesteld en ingevoerd die maken dat de klimaattransitie (dwars door allerlei verbanden in de samenleving) vorm krijgt in specifieke transitieprocessen. De waarde van het onderzoeken van deze hypothese is tweeledig. Allereerst is momenteel onduidelijk of wát in stelling is gebracht aan sturingsprincipes en sturingsactiviteiten, ook daadwerkelijk functioneert zoals gedacht; sturing op transities is überhaupt nog een vrij onbekend terrein voor overheden en onderzoekers. Daarnaast geeft een dergelijke analyse inzicht in de manier waarop doelen zoals de versnelling van en richting geven in de transitie

worden bereikt. Beide type inzichten kunnen van waarde zijn bij het verbeteren van het overkoepelende transitiebeleid.

Beleidsmixen zijn vaak historisch gegroeid. Instrumenten worden, ook bij het Klimaatbeleid, meer in lijn gebracht met de doelen uit het Klimaatbeleid. Daarnaast blijft 'oud' beleid gewoon bestaan. Ook worden nieuwe instrumenten en acties ontwikkeld om aanvullende knelpunten op te lossen. Deze beleidsinitiatieven worden ondernomen op regionaal én landelijk niveau. Al met al zorgt dit op een omvangrijk dossier als het Klimaatbeleid tot een veelvoud aan beleidspraktijken, waarbij niet noodzakelijkerwijs op voorhand analyses worden uitgevoerd op de beleidsmix als geheel. Dit leidt tot de vraag of de beleidsmix rond het klimaatdossier optimaal is ingericht. Dit is te onderzoeken door te kijken naar de feitelijke scope/afbakening van beleidsdocumenten en regelingen (e.g. de tenders), naar de visies van stakeholders hierop, en naar het portfolio van consortia/projecten dat zich 'door de beleidsmix' beweegt.

Het onderzoeken van de governance en beleidsmix vraagt niet zozeer om het vinden van de juiste empirische methoden (het komt veelal neer op interviews en deskresearch), maar vooral om het stellen van de juiste vragen. Het gaat daarbij om het vinden van 'toetsingskaders' die als spiegel kunnen dienen voor de daadwerkelijke governance en beleidsmix. Dergelijke toetsingskaders zijn af te leiden uit zowel de (wetenschappelijke) literatuur als gesprekken met betrokkenen die het beleid vormgeven en uitvoeren. Daarom stellen we voor beide perspectieven een stappenplan voor:

1. Identificeren van de initiële opzet en uitgangspunten van de governance en beleidsmix;
2. Nagaan van wetenschappelijke principes rond de governance en een optimale beleidsmix;
3. Beschrijving van de praktijk, zowel van sturingsprocessen als van de beleidsmix.
4. Een analyse van de praktijk versus de theorie.

Lerend evalueren kan bij deze analyses als onderzoeksmethode van dienst zijn. Daarbij geven evaluatoren en betrokkenen uit de praktijk betekenis aan zowel de evaluatie-'meetlat' als hoe er op die meetlat gepresteerd wordt.

Box 8.1: Lerend evalueren als manier om het governance-proces te evalueren.

We suggereren om de evaluatie van transitie-governance in te steken volgens de principes van 'lerend evalueren'. De achtergrond van deze insteek staat onder meer beschreven in de Toolbox Beleidsevaluaties²⁹. Deze insteek past goed om verschillende redenen:

- Het is niet eenduidig wat dé manier is om governance in te richten voor een transitiedossier als het Klimaatdossier.

²⁹ Toolboxbeleidsevaluaties – onderzoeksmethoden – O11. Lerend evalueren <https://www.toolboxbeleidsevaluaties.nl/onderzoeksmethoden>

Wetenschappelijke principes geven houvast, maar de evaluatieve zoeklens moet verder ingevuld en geoperationaliseerd worden. Gezamenlijk met praticioners wordt vastgesteld wat 'goede' governance is;

- Voor het Klimaatdossier zelf is nog niet helder hoe transitie governance in de praktijk precies uitwerkt en verder uitgewerkt moet worden. Een lerende evaluatie biedt ondersteuning bij dit proces van professionalisering;
- De transitie is een proces dat in ontwikkeling blijft. Deze dynamiek leent zich het best voor een ex durante insteek van evaluaties, waarbij regelmatig wordt gezien hoe het proces zich ontwikkelt en hoe bijgestuurd kan worden.

In het evaluatieproces rond lerend evalueren kunnen vijf (algemene) fases onderscheiden worden (mede geïnspireerd op het werk van PBL):

1. Het in samenspraak formuleren van evaluatievragen of sub-hypotheses. Dit kan met betrokken partijen zoals de departementen en de uitvoeringstafels van het Klimaatakkoord;
2. Het gezamenlijk ontwikkelen van een evaluatiekader – dit komt overeen met de hierboven gepresenteerde stappen 1 en 2, waarbij samen met betrokkenen wordt nagedacht welke kaders en principes fundamenteel zijn voor de sturing op innovatie in de Klimaattransitie. Hierbij past het verder ontwikkelen van de beleidstheorie;
3. Dataverzameling (mixed methods) – dit komt overeen met stap 3, en bestaat voornamelijk uit deskstudie en (groeps)interviews. Daarnaast kunnen ook kwantitatieve data als input dienen, bijvoorbeeld die opgedaan is in de monitoring;
4. Gezamenlijke interpretatie – In een workshop-setting kunnen betrokkenen met de onderzoekers betekenis gaan geven aan het gevonden feitenmateriaal, waarbij duiding wordt gegeven aan de voortgang, werking en kwetsbaarheden t.a.v. de governance;
5. Het formuleren van handelingsperspectieven – dit is waar het gehele lerende proces vooral op gericht is, namelijk het inbedden van lessen in de beleidscontext, al gedurende het onderzoek. Dit maakt dat er handelingsperspectieven ontwikkeld worden waar beleidsmakers ook daadwerkelijk mee aan de slag zullen gaan.

Lerend evalueren is relatief nieuw. Betrokkenen hebben dan ook vaak nog niet scherp wat mogelijke valkuilen of knelpunten zijn bij deze methode. Mogelijke knelpunten en mitigatie-maatregelen in een lerende evaluatie zijn:

1. De benodigde tijd van betrokkenen is substantieel – lerend evalueren is relevant voor de uitvoeringspraktijk, dus ook voor de praktijk van de betrokkenen;
2. Openheid van betrokkenen is nodig – een professionele én persoonlijke houding van de onderzoeker is van belang. Een setting van vertrouwen dient gecreëerd te worden;
3. De lobby van partijen is sterk – dit vraagt om goede onderzoekers die het onderzoeksproces kunnen begeleiden en alle geluiden in gelijke mate aan bod laten komen, en daarnaast deze lobby uit te

dagen met kritische vragen. Mogelijk kan dit óók betekenen dat de onderzoekers, los van de betrokkenen, hun evaluatieve conclusies trekken;

4. De onafhankelijkheid van onderzoekers (en validiteit van conclusies) wordt in twijfel getrokken – een duidelijke rolverdeling en rolscheiding is van belang. Het proces moet daarbij transparant zijn vormgegeven, en duidelijk moet zijn hoe conclusies en handelingsperspectieven tot stand zijn gekomen.

De Nederlandse Digitaliseringsstrategie

Beschrijving

De Nederlandse Digitaliseringsstrategie (NDS) betreft een kabinetsstrategie (met als trekker het ministerie van EZK) die ambities en doelstellingen op verschillende beleidsterreinen voor een succesvolle digitale transitie in Nederland bundelt. Het beleid bestaat uit een verzameling van al het beleid dat door deze ministeries gedaan wordt, en dat thematisch onder de agenda valt te scharen.

Hoofddoel van de NDS is één strategie voor Nederland. Het eerste-orde-effect van één strategie, in plaats van een losstaande verzameling van beleidsactiviteiten, is om de uitwisseling van kennis en *best-practises te versterken*, en te voorkomen dat op verschillende plekken het wiel opnieuw wordt uitgevonden of conflicterende regelgeving wordt ingevoerd.³⁰ Het tweede-orde-doel in termen van outcome zoals geformuleerd in de brief is driedelig: we worden digitale koploper van Europa; iedereen doet mee en profiteert; en vertrouwen dat digitale technologie zorgvuldig gebruikt wordt. De NDS wordt als volgt concreet gemaakt: het benutten van maatschappelijke en economische kansen en het versterken van grensverleggend onderzoek innovatie, nieuwe vaardigheden en een leven lang leren, een dynamische digitale economie, digitale weerbaarheid van burgers en organisaties, en het vastleggen van grondrechten en ethiek in digitale tijd. Het uiteindelijke doel van de NDS in termen van impact is een hogere welvaart door economische groei en borging van publieke belangen.

De strategie wordt jaarlijks geüpdatet, waarbij de subdoelen soms verschuiven of nader gepreciseerd worden. Onder het beleid vallen acties door en activiteiten van de volgende ministeries: EZK, BZK, IenW, VWS, LNV, OCW, SZW, DEF en JenV. Ook diverse actieplannen worden onder het beleid geschaard: De Nederlandse Cyber Security Agenda (NCSA); het Actieplan Digitale Connectiviteit; NL DIGIbeter, Agenda Digitale Overheid; NL DIGITAAL, Data Agenda Overheid; het actieplan mkb en het Actieplan Digitale Inclusie, Strategisch Actieplan voor Artificiële Intelligentie. Daarnaast omvat de digitaliseringsstrategie een aantal coalities: Dutch Blockchain Coalition, samenwerkingsplatform voor cybersecurity kennis en innovatie, Nederlandse AI Coalitie, Online Trust Coalitie Data Sharing Coalition, alliantie digitaal samenleven.

³⁰ <https://www.nederlanddigitaal.nl/nederlandse-digitaliseringsstrategie/documenten/publicaties/2018/6/1/publicatie-nederlandse-digitaliseringsstrategie>

In het overzicht financiële middelen voor de digitale economie worden vrijwel alle generieke instrumenten van het ministerie van EZK genoemd (de PPS toeslag, internationaal innoveren, innovatiekrediet, eurostart, MIT, en de WBS, SBIR) aangevuld met een aantal specifieke instrumenten (Smart Industry Fieldlabs en MKB-actieplan, Thematische AI calls en projecten, AI Bestedingen van andere departementen dan EZK en OCW, AI investeringen van Europese fondsen, en AI investeringen vanuit het Nationaal Groeifonds).

De meest recente rapportage over de digitaliseringsstrategie poogt een overzicht van resultaten te geven (en dus impliciet de nagestreefde doelen) door te proberen een aantal stellingen te onderbouwen: (1) Nederland staat hoog in internationale ranglijsten op het terrein van innovatie en digitalisering; (2) digitalisering draagt bij aan het oplossen van maatschappelijke uitdagingen; (3) digitalisering draagt bij aan economische groei; (4) er wordt steeds meer samengewerkt op het gebied van digitalisering; en (5) er is meer aandacht voor de governance van digitalisering.

Los van de vraag of de beschikbare data deze stellingen onderbouwt, is het de vraag of dit het beoogde effect van het beleid ('de digitaliseringsstrategie') is. De link tussen het overzicht en de beoogde doelen is onduidelijk. Ook kunnen op basis van dit overzicht beleidsinspanningen worden toegerekend aan het beleid en is het de vraag wat de digitaliseringsstrategie nu precies bereikt op deze terreinen. Een relevante vraag bij een dergelijke samenvatting is uiteraard de mate waarin de prestaties dankzij de strategie tot stand zijn gekomen: prestaties kunnen mogelijk ook zonder beleid tot stand gekomen zijn of door marktfactoren worden aangewakkerd.

De digitaliseringsstrategie is de facto het onder één noemer brengen van diverse beleidsactiviteiten die al eerder plaatsvonden en ingezet zijn. Niet duidelijk is hoe deze activiteiten als gevolg van de digitaliseringsagenda zijn bijgesteld of versterkt. Complicerende factor hierbij is dat de beoogde doelen van de NDS zijn geformuleerd in termen van de oorspronkelijke doelen van de onderdelen waaruit de agenda is opgebouwd. Daarmee verdwijnt het doel van de strategie zelf als beleidsinstrument uit zicht.

Afwegingen o.b.v. de beslisshulp

Voor de digitaliseringsstrategie geldt dat deze vooral bestaat uit een aantal samengevoegde beleidsinitiatieven en een coördinatiemechanisme. Een aantal van deze initiatieven is afzonderlijk toetsbaar op doeltreffendheid en doelmatigheid, omdat het instrumenten betreft met een duidelijk doel en budget. Op het eerste gezicht is dit aspect van de NDS helder geformuleerd en toetsbaar. Er moet wel worden voorzien in data om de doelen te toetsen. Veelal is er geen controlegroep beschikbaar, maar er zou bezien kunnen worden of beleid regionaal of voor verschillende bedrijven/sectoren op verschillende momenten is ingezet. Als dat niet het geval is, kan met behulp van bijvoorbeeld meer kwalitatieve methoden of regressieanalyse een effect aannemelijk worden gemaakt.

Er zijn echter ook onderdelen die niet helder geformuleerd en toetsbaar zijn. Een inspanning om de instrumenten binnen de NDS helder te formuleren aan de hand van doelen en middelen is noodzakelijk voor het evalueren ervan. Op basis van regressietechnieken is dan inzichtelijk te maken in hoeverre er aanwijzingen zijn voor doeltreffendheid van de instrumenten. Waarschijnlijk is op dit moment slechts een verklarende analyse mogelijk.

Ten slotte is de coördinatie van de activiteiten in één strategie ook te evalueren. Er kan worden geëvalueerd of sinds de coördinatie meer is samengewerkt en de synergie tussen instrumenten is verbeterd. Met behulp van regressieanalyse is hierop te toetsen. Tevens kan doelmatigheid worden getoetst door te analyseren in hoeverre dezelfde doelen met minder middelen tot stand komen na het opstellen van de NDS.

Het Techniepact

Beschrijving

Het Techniepact heeft als doel het aantal technici te vergroten en bèta, technische en technologische vaardigheden op de Nederlandse arbeidsmarkt te versterken. Dit is vormgegeven langs de lijnen kiezen, leren en werken. Kiezen gaat om vroeger belangstelling voor en verankering van techniek in het onderwijs, van funderend tot beroepsonderwijs. Leren gaat om samenwerking tussen bedrijfsleven en onderwijs om de kennis van docenten up-to-date te houden, voldoende stage- en leerwerkplekken te creëren en het hoger onderwijs meer te laten samenwerken met het bedrijfsleven. Werken legt de nadruk op permanent leren, het behouden van werkenden voor de techniek en betere samenwerking tussen het landelijke en regionale niveau.

De stichting Platform Bèta en Techniek (opgericht op 9 juli 1998) heeft ten doel de bèta-techniek problematiek te verkleinen door de aantrekkelijkheid van bèta-techniek in onderwijs en beroep te vergroten en daarbij de samenwerking tussen intermediairs en actoren uit het onderwijs en het bedrijfsleven te bevorderen. Het Techniepact is daarvan een onderdeel dat bestaat uit een samenwerkingsverband van onderwijsinstellingen, werkgevers- en werknemersorganisaties, regio's en het Rijk. Het Techniepact functioneert als een netwerk waarbij verschillende partijen zich kunnen aansluiten en waarin initiatieven op landelijk en regionaal niveau worden ontwikkeld en uitgevoerd. Doel is om knelpunten in kiezen, leren en werken op te lossen. In 2013 zijn concrete afspraken gemaakt door bedrijfsleven, onderwijs en overheid voor een periode van zeven jaar. In 2018 zijn die afspraken opnieuw vastgelegd en bijgesteld. Het belangrijkste doel van het vernieuwde pact was om het Techniepact te verbinden aan nieuwe maatschappelijke opgaven, zoals de energietransitie, klimaatverandering en circulaire economie. In 2021 is een nieuwe strategie opgesteld. De ambitie van het Techniepact is door een structurele aanpak bij te dragen aan een goed opgeleide beroepsbevolking met voldoende technici voor de banen van nu en morgen. De context van deze ambitie wordt periodiek bijgesteld, omdat het doel wel

duidelijk is maar de vraag naar bèta, technische en technologische vaardigheden steeds verandert.

Binnen vijf landsdelen vindt de uitvoering van het Techniekpact plaats; Noordwest, Noord, Oost, Zuidoost en Zuidwest. Er zijn drie ministeries betrokken (OCW, SZW en EZK). Het Techniekpact functioneert als een netwerk waar partijen zich bij kunnen aansluiten. De doelen en opbrengsten zijn niet door het Rijk bepaald, maar vooral afhankelijk van het aanpakken van knelpunten op het gebied van vraag en aanbod in de regio.

De Landelijke Regiegroep bestaat uit een aantal partijen, zoals vertegenwoordigers van de overheid (ministeries EZK, OCW, SZW), van de vijf landsdelen, van Platform Talent voor Technologie, van onderwijs- & opleidingsraden/-verenigingen, sector-/werkgevers- en werknemersorganisaties, de Topsectoren, het UWW en samenwerkingsverbanden/platforms zoals Samenwerkingsorganisatie Beroepsonderwijs Bedrijfsleven (SSB) en het Platform voor de Informatiesamenleving (ECP). De Landelijke Regiegroep wordt aangestuurd door de voorzitter, die ook een boegbeeldfunctie voor het Techniekpact vervult.

Het landelijke niveau is vervolgens opgedeeld in vijf landsdelen: Noordwest (Noord-Holland, Utrecht, Flevoland), Noord (Friesland, Groningen, Drenthe), Oost (Overijssel, Gelderland), Zuidoost (Noord-Brabant (deels), Limburg) en Zuidwest (Zuid-Holland, Zeeland, West-Brabant). Elk landsdeel heeft een coördinator die deelneemt in de Landelijke Regiegroep. Veelal betreft de coördinator het niveau van een gedeputeerde, wethouder of portefeuillehouder. Ook is er per landsdeel een centrale contactpersoon op een meer operationeel niveau die ook gezamenlijk aan een overleg deelnemen, het liaisonoverleg.

Het Platform Talent voor Technologie ondersteunt de coördinatie tussen de landsdelen en activiteiten in de landsdelen. Het Platform Talent voor Technologie is een publiek-private samenwerking waarin de ministeries van EZK, OCW en SWZ samenwerken met werkgevers- en werknemersorganisaties en elf sector-/werkgevers- en werknemersorganisaties zoals FME, BOVAG, De Unie en FNV. Daarbij bouwt het Platform Talent voor Technologie voort op o.a. het oude Platform Bèta Techniek, TechniekTalent.nu en TecWijzer. De financiering van het werk van het Platform Talent voor Technologie is, naast de middelen voor de jaarlijkse conferentie, één van de weinige directe financiële investeringen vanuit het Rijk in het Techniekpact. Het Platform Talent voor Technologie werkt met één centrale contactpersoon per landsdeel.

De actielijnen bestaan uit doelen met daarin per landsdeel een eigen invulling per doel. De Landelijke Regiegroep Techniekpact (LRT) coördineert, volgt en bewaakt de uitvoering van de landelijke strategie, de doelen en de gemaakte afspraken in het Techniekpact. De regiegroep is samengesteld uit vertegenwoordigers uit de vijf landsdelen, de Rijksoverheid, werkgevers, werknemers, topsectoren en onderwijs. De ministeries van EZK; OCW en SZW zijn verantwoordelijk voor de landelijke acties die binnen

hun terreinen vallen. Het projectteam van het Techniekpact is ondergebracht bij het ministerie van EZK.

Een aantal organisaties is betrokken.

- *Sterk Techniekonderwijs*. Het ministerie van OCW, Stichting Platforms vmbo (SPV), Platform Talent voor Technologie (PTVT), zetten zich de komende jaren samen in voor Sterk Techniekonderwijs. Zij werken hierbij nauw samen met VNO-NCW, de Federatie Techniek, MKB-Nederland, Technische branches, Platform TL, MBO Raad, en de VO-raad. Het landelijke ondersteuningsteam Sterk Techniekonderwijs ondersteunt scholen en regio's bij het maken en uitvoeren van de regionale plannen. Ze bieden ondersteuning, data, inspiratie, tools en informatie om regio's te helpen bij het vormen van het samenwerkingsverband, het planvormingsproces en de uitvoering.
- *Regionale VO/HO Netwerken*. Negen netwerken van VO en HO-instellingen met als doel een doorlopende vak- en curriculumontwikkeling in het VO, verbetering van de aansluiting tussen VO en HO en doorlopende professionele ontwikkeling van docenten, onderwijsassistenten en schoolleiders in het VO en HO.
- *Katapult* is een netwerk van meer dan 300 samenwerkingsverbanden tussen onderwijs en bedrijfsleven en groeit continu. Doelstelling is om de samenwerking tussen onderwijs, bedrijfsleven en maatschappij te verbeteren. Dit gebeurt bijvoorbeeld door professionals uit het bedrijfsleven die lessen verzorgen. Of door studenten die tijdens hun opleiding onderzoek doen voor een MKB-bedrijf. Er participeren inmiddels 84.000 studenten, 9.800 bedrijven en 5.000 docenten in deze samenwerkingsverbanden, ook wel Centers of expertise, Centra voor innovatief vakmanschap en andere vormen van samenwerking in het beroepsonderwijs. Er bestaat een jaarlijkse subsidie van ongeveer 0,3 mln. euro voor deze organisatie.
- *Jet-Net & TechNet*. Deze organisatie geeft uitvoering aan de gezamenlijke ambities van de technische werkgevers- en werknemersorganisaties, de overheid en het bedrijfsleven om samen gastlessen, bedrijfsbezoeken en projecten te organiseren. Er is een aanbod voor het PO, VO en voor bedrijven. Partners zijn Bouwend Nederland, BOVAG, FME, Koninklijke Metaalunie, Techniek Nederland, CNV Vakmensen, De Unie, FNV Metaal, VHP2 en het ministerie van OCW. Er bestaat een jaarlijkse subsidie van ongeveer 0,8 mln. euro voor deze organisatie.

Er zijn in 2021 vier thema's geïdentificeerd. De totale subsidie van het Techniekpact bedraagt ongeveer 0,8 mln. euro per jaar. De totale subsidie van de stichting Platform Bèta en Techniek bedraagt zo'n 3,1 mln. euro. Onderstaande thema's bevatten de aanpak van het kabinet in de komende periode, waarvoor de ministeries verantwoordelijk zijn.

- **Instroom**
 - meisjes/vrouwen kiezen vaker voor een technische opleiding en baan
 - jongeren met een migratieachtergrond kiezen vaker voor een technische opleiding en baan
 - jongeren kiezen vaker voor een technische mbo-opleiding
 - het inbedden van wetenschap & digitalisering in het programma van primair en voortgezet onderwijs
 - toename van het aantal leerwerkbedrijven
 - nauwe samenwerking tussen onderwijs en bedrijfsleven op gebied van het hbo

De uitvoering, ondersteuning en coördinatie van initiatieven, activiteiten en programma's vindt plaats vanuit de regio, onder andere via Jet-Net & TechNet (primair onderwijs en voortgezet onderwijs), Sterk Techniekonderwijs (vmbo); Regionale VO-HO netwerken (havo /vwo) en Katapult (mbo/hbo).

- **Publiek-private samenwerking**
 - versterking van praktijkonderwijs en samenwerking in het technisch hbo
 - het opschalen van goede initiatieven
 - het onder de aandacht brengen van financieringsregelingen en toepassingsmogelijkheden

Alle Nederlandse publiek-private samenwerkingen komen samen in Katapult.

- **Lerarentekort en professionalisering**
 - hybride docentschap een plek te geven op regionale agenda's en in sectorplannen
 - lerarenopleidingen meer aan te laten sluiten bij de kennis en ervaring van zijinstromers

De uitvoering, ondersteuning en coördinatie van initiatieven, activiteiten en programma's vindt plaats vanuit de regio, onder andere via Sterk Techniekonderwijs (vmbo); Regionale VO-HO netwerken (havo/vwo); Katapult (mbo/hbo).

- **Leven lang ontwikkelen**
 - toename van het aantal omscholingen naar en bijscholing in de technische beroepen
 - het onder de aandacht brengen van financieringsregelingen en toepassingsmogelijkheden
 - het borgen van het perspectief in de technische beroepen in de digitale provinciale LLO-portalen

Monitoring

- Er is de Techniepact monitor waarin relevante data over technische opleidingen, arbeidsmarkttekorten etc. continu worden bijgewerkt. De informatie is daarnaast regionaal uitgesplitst waarmee de monitor gebruiksvriendelijk is voor beleidsmakers in de regio. Zie www.techniepactmonitor.nl.

- Platform Talent voor Technologie (PTVT) krijgt jaarlijks een opdracht van de drie ministeries om de verbinding met en tussen regio's te organiseren binnen het Techniepact. Daarnaast maken communicatie en evenementen onderdeel uit van de subsidieopdracht. PTVT rapporteert jaarlijks aan de departementen over de voortgang van de activiteiten.
- De drie ministeries rapporteren jaarlijks aan de Tweede Kamer over de activiteiten en voortgang op de doelen binnen het Techniepact. Tot 2019 gebeurde dat met een Voortgangsrapportage waarin een veelheid aan activiteiten (ook regionaal) werd genoemd. Eind 2020 is gekozen om voor het kalenderjaar 2021 juist op voorhand een agenda met toekomstige acties naar de Kamer te sturen. Dit is in lijn met de Techniepactevaluatie van SEO Economisch Onderzoek en Technopolis waarin is geadviseerd meer programmatisch te werk te gaan. Een eerste verantwoording over deze kabinetsinzet in de nieuwe vorm heeft nog niet plaatsgevonden en staat gepland voor eind 2021/begin 2022.
- Zoals gezegd kan het Techniepact gezien worden als een netwerk-aanpak; een soort paraplu waaronder verschillende organisaties verschillende activiteiten organiseren in lijn met de Techniepactdoelen. Zo moeten bijvoorbeeld regelingen als MKB!dee (40 mln. euro) en het RIF (100 mln. euro) worden beschouwd als instrumenten die vallen onder de paraplu van het Techniepactbeleid. Deze regelingen worden echter zelfstandig geëvalueerd.

Afwegingen o.b.v. de beslisthulp

Het Techniepact pakt een aantal knelpunten aan in het onderwijs en op de arbeidsmarkt voor technisch geschoold personeel. De visie (uit 2013) is dat een structurele en overkoepelende aanpak bijdraagt aan een goed opgeleide beroepsbevolking met voldoende slimme en vakbekwame technici voor de banen van nu en morgen. Het falen dat centraal staat is capaciteitsfalen van het onderwijs. Vanaf het primair onderwijs tot de arbeidsmarkt is onvoldoende aandacht voor en kennis van techniek. Dit leidt ertoe dat kinderen te weinig in aanraking komen met techniek, er te weinig kinderen kiezen voor een technische opleiding en de kinderen die kiezen voor een technische opleiding vaak switchen. Als gevolg hiervan ontstaan knelpunten op de arbeidsmarkt en stopt de innovatie in het bedrijfsleven. Op de arbeidsmarkt bestaat ook capaciteitsfalen doordat er te weinig bekwame docenten zijn die kinderen opleiden en er onvoldoende wordt geïnvesteerd in nieuwe kennis en vaardigheden om werkenden in technische beroepen te behouden of hun inzetbaarheid te verbeteren.

De lage instroom in het technische onderwijs is het gevolg van verkeerde beelden over technische opleiding en beroepen aan de vraagkant en onvoldoende aanbod van effectief onderwijs. Problemen aan de aanbodzijde vallen verder uiteen in gebrek aan voldoende gekwalificeerde docenten in het primair en voortgezet onderwijs en later in het techniekonderwijs en onvoldoende betrokkenheid van het bedrijfsleven waardoor de match tussen opleiding en arbeidsmarkt suboptimaal is.

Er vindt onvoldoende scholing plaats in de techniek als mensen eenmaal aan het werk zijn. Aan de vraagzijde is sprake van de volgende knelpunten: (i) restricties op de onderwijs- of financiële markten kunnen een verklaring bieden voor onvoldoende scholing, (ii) externe effecten doordat werkenden die het meest baat hebben bij scholing het onderwijsaanbod niet (kunnen) benutten of geen prikkel ervaren, (iii) vanwege het holdup-probleem (de werkgever heeft geen garantie dat de werknemer die een opleiding volgt, ook in de toekomst voor het bedrijf blijft werken en de werknemer die zich verder ontwikkelt, is onzeker of de baten leiden tot een hoger loon) en (iv) gedragseconomische verklaringen, zoals kortzichtigheid, verliesaversie, defaults en normen, die leiden tot een suboptimale leercultuur. Aan de aanbodzijde bestaan beperkingen van het aanbod van opleidingen op de onderwijsmarkt door bijvoorbeeld hoge investeringskosten om opleidingen te starten en marktverstoringen door kruissubsidiëring van onderwijsinstellingen.

Er is dit jaar (2021) een actieagenda opgesteld. Met deze actieagenda geeft het kabinet richting. In de praktijk gebeurt er veel meer dan dat. Op provinciaal, arbeidsmarkt en gemeentelijk niveau werken onderwijs, bedrijfsleven en overheden samen aan technisch onderwijs en banen in de techniek, bouw en ICT. Deze regionale samenwerking is het fundament van de Techniepactaanpak en blijft doorgaan. Ook landelijk zijn er veel initiatieven en instrumenten die belangrijk zijn om de doelen van het Techniepact te bereiken. De acties die in 2021 op de agenda staan kunnen worden geëvalueerd.

Als voorbeeld kan de instroom in de techniek worden genomen. Er is een aantal concrete doelen en instrumenten gedefinieerd. Samen moeten deze ervoor zorgen dat de instroom wordt vergroot (output), wat leidt tot meer economische groei en grotere concurrentiekracht (outcome).

Instrumenten en activiteiten die hierbij horen richten zich op meer meisjes en vrouwen in de techniek:

- Ontwikkelen van het Beta&TechMentality model. Om meisjes te kunnen inspireren voor de mogelijkheden van techniek is het van belang om allereerst te weten wat hen drijft en motiveert;
- Diversiteitscheck. Bewust omgaan met verschillen tussen jongens en meisjes;
- Communicatiecheck. Communicatie beter laten aansluiten bij de vrouwelijke doelgroep;
- 30 procent vrouwen in de techniek in 2030. FME heeft de ambitie om in 2030 het aantal vrouwen in de technologische industrie ruimschoots te verdubbelen naar 30 procent door het aantrekkelijker te maken om meer uren te werken en de genderstereotypering aan te pakken.

Meer jongeren met een migratieachtergrond in de techniek:

- Vergroot de zichtbaarheid van techniek;
- Zorg voor een brede kennismaking met techniek;
- Vergroot technisch zelfvertrouwen;
- Benadruk de potentie van een toekomst met techniek;

- Ontwikkel een duidelijk beleid op discriminatie in de techniek.

Primair en voortgezet onderwijs beter richten op techniek:

- Structurele inzet en aanpak door alle onderwijslagen, -fases en -keten heen. In de herziening van het curriculum wordt technologisch gerelateerd onderwijs opgenomen;
- Op korte termijn kunnen instellingen gebruik maken van Sterk Techniekonderwijs en aan de slag met wetenschap & digitalisering vanuit Jet-Net & TechNet.

Meer BBL-instroom: Meer leerlingen moeten gaan kiezen voor technische opleidingen, omdat de behoefte van de arbeidsmarkt aan mensen met het diploma technisch mbo (o.a. BBL) toeneemt.

Versterken praktijkonderwijs in het hbo: De praktijkcomponent in het hbo moet meer aandacht krijgen. Het programma 'mkb-route in het hbo' is gestart met zeven pilots

Aantrekken van internationaal talent: In de Talent Coalition wordt gewerkt aan de positionering van Nederland als aantrekkelijk vestigingsland voor internationaal talent in loondienst.

Het is een flinke uitdaging om het Techniepact als geheel te evalueren (zie SEO/Technopolis, 2020). Wel kan de netwerkaanpak en de output die daaruit voorkomt worden gemonitord en op consistentie worden beoordeeld. Specifieke subdoelen (zoals de instroom in technische opleidingen in de gehele keten) kunnen geëvalueerd worden. Ten slotte is het mogelijk om op instrumentniveau (zoals de instroom van meer meisjes in de techniek of programma's zoals het Regionaal Investeringsfonds) te evalueren.

Het is op systeemniveau onduidelijk wat het optimale niveau van het aantal mensen in de techniek is. Hierdoor is het onduidelijk hoe groot de impuls in het onderwijs zou moeten zijn waar het gaat om aanpassingen in het curriculum (als techniek meer aandacht krijgt, gaat dat ten koste van een ander vak) en waar het gaat om het beïnvloeden van de keuzes van jongeren. Wel kan de mate waarin zich knelpunten voordoen worden gemeten en kunnen instrumenten worden ingezet en geëvalueerd.

Jongeren maken keuzes op basis van voorkeuren, hun netwerk en eerdere ervaringen in het onderwijs. Op die manier ontdekken ze hun optimale opleiding. Aanpassen van de onderwijspaden is alleen nuttig als vaststaat dat jongeren suboptimale opleidingskeuzes hebben gemaakt. Als jongeren niet willen werken in de techniek is stimuleren niet effectief.

Aanpassingen van het curriculum gaan waarschijnlijk langer duren dan met interventies mensen stimuleren om voor techniek te kiezen. Voor effectieve keuzes is het van belang dat beide in de tijd niet te ver uit elkaar liggen.

Er kan worden gemeten hoeveel meisjes en jongeren met een migratieachtergrond kiezen voor techniek. Hetzelfde geldt voor de

BBL-instroom, het praktijkonderwijs in het hbo en het aantrekken van talent uit het buitenland. Er moet wel worden gedefinieerd wat het exacte doel is en op welke termijn dit moet worden bereikt. Een wijziging van het curriculum is vast te stellen en de effecten ervan op de instroom zijn te meten.

De verschillende initiatieven kunnen worden geconcretiseerd. Inspanningen rondom meer meisjes in de techniek lijken het meest concreet. Door in een aantal regio's concrete programma's uit te voeren is het mogelijk om te meten in hoeverre deze inspanningen effectief en efficiënt zijn en kan worden geleerd in hoeverre programma's in andere regio's ook zouden kunnen werken.

De samenhang in de onderwijsketen kan in de tijd worden gemeten. Als in het primair onderwijs programma's worden ingevoerd om keuzes te beïnvloeden, kan in het voortgezet onderwijs worden gemeten in hoeverre meer meisjes en jongeren met een migratieachtergrond kiezen voor een technisch profiel en vervolgens een beroepsopleiding in de techniek kiezen. Ook is het mogelijk om hier interventies te doen en het versterkende effect ervan in kaart te brengen.

Het is goed mogelijk om de effecten van de programma's en instrumenten die vallen onder de paraplu van het Techniepact te meten op effectiviteit en efficiency. Voorwaarde is dat concrete doelen worden gesteld en programma's en instrumenten niet in iedere regio tegelijk worden ingevoerd zodat een effectmeting mogelijk is (causaliteit kan worden vastgesteld). De data zijn beschikbaar, zeker waar het gaat om de in- en doorstroom in het onderwijs. De maatschappelijke impact van het Techniepact als geheel is niet hard vast te stellen. De meting beperkt zich tot consistentie van de aanpak.

8.2. Lessen over het afbakenen van (en aanmerken als) S/T-beleid

Kijken we naar de overeenkomsten en verschillen tussen de vier onderzochte casussen, dan is een eerste observatie dat het afbakenen van de te evalueren beleidsterreinen een delicate aangelegenheid is. Een beleidsdomein als CO₂-reductiebeleid of Klimaatbeleid correspondeert bijvoorbeeld niet met een concreet beleidsdossier (met instrumenten en bijbehorende begrotingen), en het Missiegedreven Topsectoren- en Innovatiebeleid komt niet overeen met hoe sommige documenten en stakeholders naar missiebeleid kijken. De Nationale digitaliseringsstrategie is erg breed en bestaat uit een bundeling van veelal reeds bestaande instrumenten die door verschillende ministeries worden uitgevoerd. Ten slotte bestaat het Techniepact al wat langer en zijn de accenten door de jaren heen veranderd. Doordat S/T-beleid niet om een instrument draait, maar om een verzameling beleidsimpulsen die onder een al dan niet geformaliseerd programma vallen, kan het lastig zijn om grip te krijgen op de beleidsstrategie en het beleidspakket dat geëvalueerd dient te worden; zie ook stap 3 in het stappenplan. Alternatieve

afbakeningen kennen vaak ook andere doelen, hetgeen betekent dat er al snel vragen rijzen over waarop evaluaties zich eigenlijk zouden moeten toespitsen. Dergelijke onduidelijkheden dienen in een vroeg stadium door de opdrachtgever van een evaluatie weggenomen te worden, aangezien ambigue uitgangspunten een barrière kunnen vormen voor het ontwikkelen van evaluatie-ontwerpen die de meest prangende evaluatievragen kunnen beantwoorden en tot discussie kunnen leiden over wat er nu eigenlijk geëvalueerd dient te worden. Dit roept wel gelijk de vraag op hoe voorkomen kan worden dat de opdrachtgever middels de vraagsturing teveel bepaalt wat er wel en niet geëvalueerd mag worden. Door slim zelf te kiezen wat en hoe er geëvalueerd moet worden, kan een opdrachtgever immers al te kritische beoordelingen voorkomen. Dit is een vorm van overheidsfalen die raakt aan de vormgeving van evaluatieprocessen binnen de overheid zelf, hetgeen buiten de scope van dit onderzoek valt.

Een tweede observatie betreft de variëteit van de onderzochte casussen, specifiek als het gaat om de kenmerken die onderscheidend zijn voor S/T-beleid. Het MTIB en het CO₂-reductiebeleid komen waarschijnlijk het dichtst in de buurt van de in dit rapport gehanteerde interpretatie van S/T-beleid, aangezien het overkoepelende strategieën zijn die zelf weer leunen op (innovatie)interventies die op complementaire wijzen dienen aan te grijpen op hetzelfde systeem. Er zijn echter ook casussen die niet al deze kenmerken hebben. We onderscheiden drie gevallen:

- Eén categorie bestaat uit beleid waar de overkoepelende strategie minder centraal staat, en waar het zwaartepunt ligt bij een verzameling concrete beleidsinterventies die ook niet pretenderen heel synergetisch met elkaar samen te hangen. Iets dergelijks speelt in zekere mate bij de Nationale Digitaliseringsagenda: die heeft meer weg van een thematisch overzicht van beleidsimpulsen, dan van een geïntegreerde strategie met synergie tussen de verschillende onderdelen. Consequentie van deze beleidsopzet is dat het in zo'n geval beter is elk afzonderlijk instrument mogelijk te evalueren is op basis van de standaarden uit de Expertwerkgroep Effectmeting (2012), er is immers weinig samenhang;
- Daarnaast is er ook ogenschijnlijk S/T-beleid dat juist veel nadruk legt op de overkoepelende strategie of het programma, waarbij weliswaar een verzameling concrete beleidsinterventies is, maar de relatie met een concrete set beleidsinterventies relatief zwak is. Men kan denken aan het Smart Industry programma of (de huidige staat van) de beleidsstrategie over de Circulaire Maakindustrie. In beide gevallen lijkt het beleidsmatige zwaartepunt (qua middelen) eerder bij het programmeren en coördineren te liggen dan bij het inzetten of significant bijsturen van concrete vormen van overheidsingrijpen. De implicatie voor beleidsevaluatie is in dit geval dat die zich waarschijnlijk het best kan toespitsen op enkel de invalshoeken A of B uit het evaluatiekader, aangezien de andere invalshoeken vergen dat er ook substantiële prikkels zijn waarlangs systeemversterking/-transformatie (en daaruit volgende uitkomsten) tot stand kunnen komen;

- Ten slotte is het ook mogelijk dat er bij S/T-beleid geen overkoepelende coherente strategie is, en eigenlijk ook geen pakket van heterogene beleidsinterventies. Dit lijkt nog het meest aan de orde bij de casus van het Techniepact, wat primair leunt op het organiseren van een netwerk waarin uiteenlopende partijen samenwerken aan het vergroten van bèta, technische en technologische vaardigheden op de Nederlandse arbeidsmarkt. Hierbij wordt weliswaar gebruikgemaakt van systeemdenken en worden ontwikkelingen en voortgang jaarlijks in beeld gebracht, maar de sterke focus op instrumenten of afgebakende subdoelen laat zien dat er waarschijnlijk ook per instrument kan worden geëvalueerd.

De algemene les is dat niet elke beleidsagenda met systeem- of transitieaspecten automatisch ook overeenkomt met het type S/T-beleid waar dit rapport zich op toelegt. Mogelijk zijn de hier behandelde evaluatie-opties op onderdelen relevant, bijvoorbeeld als het gaat om het evalueren van governance-processen en -structuren, maar het blijft verstandig om kritisch te kijken of een benadering zoals door de Expertwerkgroep Effectmeting (2012) is voorgesteld niet (ook) tot de mogelijkheden behoort. Dit geldt des te meer als het S/T-beleid afzonderlijke instrumenten kent die aangrijpen op helder afgebakende subgroepen van actoren.

8.3. Lessen over onderzoeksmethoden

Er zijn ook enkele lessen te trekken op basis van de onderzoeksmethoden waar de beschrijving naar verwijst na het doorlopen van het stroomschema. In de casussen waarbij kwantitatieve effectmeting mogelijk lijkt gaat het om het bestuderen van structuurveranderingen die bewerkstelligd worden door het Techniepact en NDS, omdat die casussen uit relatief onafhankelijke interventies bestaan waarvoor ook de methoden het rapport van de Expertwerkgroep Effectmeting (2012) zich lenen (zie Sectie 6.1). Ook in het geval van de voor S/T-beleid meer typische casus van het MTIB geldt echter ook dat er methoden ingezet kunnen worden die hoog op de effectladder scoren. Hierbij zijn de volgende observaties van belang:

- (Quasi-)experimentele methoden zijn toepasbaar voor deelaspecten van het S/T-beleid. In Hoofdstuk 3 is reeds besproken waarom er voor een S/T-programma als zodanig vaak moeilijk een *counterfactual* is aan te wijzen, waarvoor dan ook nog eens relevante data beschikbaar moeten zijn. Als gevolg hiervan is het lastig om te bepalen of MTIB-beleid voor bijvoorbeeld de circulaire economie verantwoordelijk is voor betere prestaties op dat laatste punt, of dat positieve prestaties toe te schrijven zijn aan andere ontwikkelingen. Wat haalbaarder lijkt, is het onderzoeken van beleidseffecten op specifieke intermediaire doelen, zoals structuurveranderingen in de zin van gewijzigde investerings- of productiepatronen (die het gevolg zouden moeten zijn van systeemversterking). Een evaluatie van MTIB-beleid zou gebruik kunnen maken van het feit dat het MTIB onderzoek naar specifieke oplossingsrichtingen probeert te versterken, en dat ontwikkelingen in bijvoorbeeld innovatie-inspanningen op de geselecteerde thema's - zoals geobserveerd in WBSO-data - afgezet kunnen worden tegen

ontwikkelingen in soortgelijke sectoren of technologiegebieden die *niet* geprioriteerd zijn vanuit het MTIB-beleid. Een dergelijke analyse wijst niet uit of die extra innovatie-inspanningen ook resulteren in meer circulaire economie, maar ze geven wel een indicatie voor eventuele *crowding-in* als gevolg van op zichzelf moeilijk meetbare systeemversterking;

- Bij S/T-beleid bestaat een databehoeft op een heel specifiek detailniveau over de richting van allerlei investeringen en inspanningen waar vaak geen kant-en-klare antwoorden op zijn. Bij enkelvoudige instrumenten gericht op individuele actoren is een behandelde groep en controlegroep te onderscheiden, maar S/T-beleid draait om ontwikkelingen op mesoniveau die niet goed naar individuele actoren te herleiden zijn. Dit betekent dat er gemeten moet worden op het niveau van abstracte begrippen als oplossingsrichtingen en innovatiepaden, en dat ontwikkelingen daarin afgezet moeten worden tegen ontwikkelingen in gelijkende paden die geen voordeel hebben ondervonden van het ingezette beleid. Dit kan theoretisch gezien, omdat de essentie van S/T-beleid is dat er op een zeker moment een heel duidelijke focus moet ontstaan in de verandering die het beleid probeert te ontketenen of versterken. In de praktijk kan het echter lastig zijn om hier grip op te krijgen. Op het moment dat er duidelijke visies zijn, zoals roadmaps, wil dat niet zeggen dat data over relevante investeringen en inspanningen al aan de richtingen in die roadmaps te koppelen zijn. Minimaal even belangrijk is om ook te kunnen noemen welk deel van de data gaat over richtingen die juist niet uitverkoren zijn;
- Een derde observatie is dat sommige (quasi-)experimentele methoden ter evaluatie van S/T-beleid mogelijk in generieke zin toe te passen zijn voor zeer uiteenlopende casussen. Net zoals er voor enkelvoudige instrumenten een aantal standaardtechnieken zijn, zoals het gebruik van (*propensity score*) *matching of regression discontinuities*, zo kan het ook voor S/T-beleid denkbaar zijn dat sommige aanpakken in een algemene toolbox belanden. Dit ligt bijvoorbeeld voor de hand voor netwerkanalyses, aangezien er vanuit beleidsadministratie vaak informatie is over welke partijen participeren in beleid en zelfs wie er samenwerken. Als S/T-beleid op actor-niveau tot een sterker of ander systeem moet leiden, zou dit moeten blijken uit wijzigingen in de omvang en/of structuur van betrokken netwerken. Hoewel de meest relevante indicatoren en type netwerken per casus kunnen verschillen, lijken netwerkanalyses dus wel een plek in de toolbox te verdienen. Hetzelfde kan gelden voor de hierboven beschreven aanpak om gelijkende richtingen te identificeren die wel en niet door beleid gesteund zijn, om zo te kijken of er (op hoge ontwikkelingsniveaus) verschillen bestaan in de mate van *crowding-in*. Samenvattend is de oproep hier dus om in toekomstige evaluaties van S/T-beleid niet alleen te kijken naar welke (quasi-)experimentele methoden daar het best passen, maar ook naar de bredere toepasbaarheid van die (quasi-)experimentele methoden.

Voor kwalitatieve methoden is er meer keuzevrijheid bij het ontwerpen in vergelijking met een (quasi) experimentele

evaluatieopzet die sterk door databeschikbaarheid beperkt wordt. Welke kwalitatieve methode wordt ingezet hangt daarbij sterk af van de keuze van de evaluator op welke aspecten van het beleid hij of zij zich wil richten. In de case-uitwerkingen is soms gekozen voor een bepaalde methode die het meest logisch of veelbelovend leek in het licht van de geselecteerde evaluatie-invalshoek en hypothesen. Indien er bij het doorlopen van stap 1-7 echter andere accenten waren gelegd had ook de keuze voor kwalitatieve methoden er anders uit kunnen zien.

9. Epiloog

Dit rapport beoogt uit te leggen wat S/T-beleid is; waarom het evalueren daarvan lastig is; welke ‘invalshoeken’ (en onderliggende analysekaders) er gekozen kunnen worden bij het evalueren van de keten van interventie-rationale naar impact; en welke empirische methoden hierbij naar verwachting het best benut kunnen worden.

Hoewel er zoveel mogelijk gebruik is gemaakt van wetenschappelijke literatuur en inzichten uit de evaluatiepraktijk, is er op diverse punten geconstateerd dat het evalueren van S/T-beleid nog in de kinderschoenen staat. Het doel van dit rapport is om nu de eerste stap te zetten in het evalueren van S/T-beleid met als uitgangspunt het advies van de Expertwerkgroep Effectmeting. Toekomstige evaluaties moeten uitwijzen in hoeverre het mogelijk is om S/T-beleid in het geheel of op onderdelen te onderwerpen aan evaluaties met een hoge causale bewijsstandaard. Tevens zal moeten blijken welke concrete analysekaders benut of ontwikkeld worden voor ieder van de in dit rapport benoemde ‘invalshoeken’ van S/T-beleidsevaluaties. Omdat het evalueren van S/T-beleid nog in zo’n pril stadium verkeert, en omdat het uiteindelijk ook steeds weer maatwerk vergt, verdient het aanbeveling om over enkele jaren terug te kijken en het net op te halen. Er zullen dan ongetwijfeld vorderingen gemaakt zijn waarmee de inhoud van dit rapport kan worden aangescherpt. Immers, net als bij S/T-beleid zelf is experimenteren soms de enige manier om te bepalen wat echt werkt. Door dit adagium ook toe te passen op beleidsevaluaties zelf, werken we toe naar het vestigen van een beproefde evaluatiesystematiek.

Bronmateriaal

Arnold, E. (2022). *Evaluating Missions*. Presentatie voor EC Research WP.

Arnold, E., Åström, T., Glass, C. and De Scalzi, M. (2018). *How should we evaluate complex programmes for innovation and socio-technical transitions?*

Åström, Arnold & Olsson (2022). *Evaluating the 17 Strategic Innovation Programmes (SIPs)*. Technopolis Sweden

Bergek, A., Jacobsson, S., Carlsson, B., Lindmark, S., and Rickne, A. (2008). Analyzing the functional dynamics of technological innovation systems: A scheme of analysis. *Research policy*, 37(3), 407-429.

Boon, W., & Edler, J. (2018). Demand, challenges, and innovation. Making sense of new trends in innovation policy. *Science and Public Policy*, 45(4), 435-447.

Barbrook-Johnson, P., Castellani, B., Hills, D., Penn, A., & Gilbert, N. (2021). Policy evaluation for a complex world: Practical methods and reflections from the UK Centre for the Evaluation of Complexity across the Nexus. *Evaluation*, 27(1), 4-17.

Dialogic (2017). *Evaluatie SBIR*.

Duflo, E. (2017). The economist as plumber. *American Economic Review*, 107(5), 1-26.

Elzinga, R., Negro, S. O., Janssen, M. J., Wesseling, J. H., & Hekkert, M. P. (2020). Het Missiegedreven Innovatiesysteem: Uitbreiding 'Technologisch Innovatie Systeem'-raamwerk ter monitoring van de Circulaire Economie. Universiteit Utrecht.

Expertwerkgroep Effectmeting / Commissie Theeuwes (2012). *Durf te meten*.

Folkert, R., Verwoerd, L., & Verwest, F. (2018). Lerend evalueren: navigeren tussen verantwoord en leren. *Beleidsonderzoek Online juli*.

Foray, D., *International Journal of Industrial Organization*, 1997, The dynamic implications of increasing returns: Technological change and path dependent inefficiency, 15(6), 733-752.

Grillitsch, M., Hansen, T., Coenen, L., Miörner, J., & Moodysson, J. (2019). Innovation policy for system-wide transformation: The case of strategic innovation programmes (SIPs) in Sweden. *Research Policy*, 48(4), 1048-1061.

Hekkert, M. P., Suurs, R. A., Negro, S. O., Kuhlmann, S., & Smits, R. E. (2007). Functions of innovation systems: A new approach for analysing technological change. *Technological forecasting and social change*, 74(4), 413-432.

Hekkert, M. P., Janssen, M. J., Wesseling, J. H., & Negro, S. O. (2020). Mission-oriented innovation systems. *Environmental Innovation and Societal Transitions*, 34, 76-79.

Hünermund, P., & Czarnitzki, D. (2019). Estimating the causal effect of R&D subsidies in a pan-European program. *Research Policy*, 48(1), 115-124.

Janssen, M.J. (2019). What bangs for your buck? Assessing the design and impact of Dutch transformative policy. *Technological Forecasting and Social Change*, 138, 78-94.

Janssen, M.J. (2022). Legitimation and effects of mission-oriented innovation policies: A spillover perspective. *Hacienda Pública Española / Review of Public Economics*.

Janssen, M.J., Bergek, A., Wesseling, J. (2022). Evaluating systemic innovation and transition programmes: Towards a culture of learning. *PLOS Sustainability and Transformations*.

Janssen, M.J., Bogers, M., & Wanzenböck, I. (2020). Do systemic innovation intermediaries broaden horizons? A proximity perspective on R&D partnership formation. *Industry and Innovation*, 27(6), 605-629.

Janssen, M.J. & Abbasiharofteh, M. (2022). Boundary spanning R&D collaboration: Key Enabling Technologies and Missions as alleviators of proximity effects? *Technological Forecasting and Social Change*, 108, 121689.

Janssen, M., Torrens, J., Wesseling, J., & Wanzenböck, I. (2021). The promises and premises of mission-oriented innovation policy: A reflection and ways forward. *Science and Public Policy*, 48 (3), 438-444.

- Kroll, H. (2019). How to evaluate innovation strategies with a transformative ambition? A proposal for a structured, process-based approach. *Science and Public Policy*, 46 (5), 635-47.
- Kuhlmann, S., and Rip, A. (2018). Next-Generation innovation policy and grand challenges. *Science and Public Policy*, 45(4), 448–454.
- Larrue, P. (2021). The design and implementation of mission-oriented innovation policies: A new systemic policy approach to address societal challenges. Paris: OECD.
- Loorbach, D., Frantzeskaki, N., & Avelino, F. (2017). Sustainability transitions research: transforming science and practice for societal change. *Annual Review of Environment and Resources*, 42.
- Lundvall, B. A. (1992). National systems of innovation: An analytical framework. London: Pinter.
- Magro, E., and Wilson, J.R., (2019). Policy-mix evaluation: Governance challenges from new place-based innovation policies. *Research Policy* 48, 103612.
- Ministerie van EZ. (2017). Sturen in een verweven dynamiek: Perspectieven op complexiteit en oriëntaties voor beleid. Den Haag: Ministerie van Economische Zaken.
- evaluatie van het Bedrijvenbeleid door het ministerie van EZK (2020)
- Molas-Gallart, J., Boni, A., Giachi, S., & Schot, J. (2021). A formative approach to the evaluation of Transformative Innovation Policies. *Research Evaluation*, 30(4), 431-442.
- Patton, M. Q. (2011), *Developmental evaluation: applying complexity concepts to enhance innovation and use*. New York and London: The Guilford Press. ISBN: 978-1-60623-87.
- PBL-rapport “Grote opgaven in een beperkte ruimte” (april 2021),
- Rodrik, D. (2008). *One economics, many recipes: globalization, institutions, and economic growth*. Princeton University Press.
- Rogge, K. S., & Reichardt, K. (2016). Policy mixes for sustainability transitions: An extended concept and framework for analysis. *Research Policy*, 45(8), 1620-1635.
- Schot, J. and Steinmueller, W.E. (2018). Three frames for innovation policy: R&D, systems of innovation and transformative change. *Research Policy* 47 (9), 1554-1567.
- Scotchmer, S. (1991), Standing on the Shoulders of Giants: Cumulative Research and the Patent Law. *Journal of Economic Perspectives* 5 (1), 29–41.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., and Befani, B. (2012). Broadening the range of designs and methods for impact evaluations. *DFID Working Paper*, 38.
- Turnheim, B., Berkhout, F., Geels, F., Hof, A., McMeekin, A., Nykvist, B., & van Vuuren, D. (2015). Evaluating sustainability transitions pathways: Bridging analytical approaches to address governance challenges. *Global Environmental Change*, 35, 239-253.
- Van Mierlo, B., Leeuwis, C., Smits, R., and Woolthuis, R. K. (2010). Learning towards system innovation: Evaluating a systemic instrument. *Technological Forecasting and Social Change*, 77(2), 318–334.
- Van der Knaap, P., Pattyn, V. & Hanemaayer, D. (2020). Beleidsevaluatie in theorie en praktijk.
- Van der Steen, M., Faber, A., Frankowski, A., & Norbruis, F. (2018). Opgavegericht evalueren. Beleidsevaluatie voor systeemverandering. Den Haag: NSOB.
- Verwoerd, L., Klaassen, P., van Veen, S. C., De Wildt-Liesveld, R., & Regeer, B. J. (2020). Combining the roles of evaluator and facilitator: Assessing societal impacts of transdisciplinary research while building capacities to improve its quality. *Environmental Science & Policy*, 103, 32-40.
- Weber, K.M. and Rohracher, H. (2012). Legitimizing research, technology and innovation policies for transformative change: combining insights from innovation systems and multi-level perspective in a comprehensive failures framework. *Research Policy*, 41, 1037–1047.
- Wesseling, J.H., and Meijerhof, N. (2021). Developing and applying the Mission-oriented Innovation Systems (MIS) approach. DOI: 10.31235/osf.io/xwg4e
- Wieczorek, A. J., & Hekkert, M. P. (2012). Systemic instruments for systemic innovation problems: A framework for policy makers and innovation scholars. *Science and public policy*, 39(1), 74-87.
- Wittmann, F., Roth, F., Hufnagl, M., Yorulmaz, M., Lindner, R. (November 2021). Second Mission Analysis Report of the Scientific Support Action to the German Hightech Strategy 2025 - Zooming in: Translating missions into policy instruments. Fraunhofer ISI. Karlsruhe.
- Woolthuis, R. K., Lankhuizen, M., & Gilsing, V. (2005). A system failure framework for innovation policy design. *Technovation*, 25(6), 609-619.

Uitgave

Ministerie van Economische Zaken en Klimaat
Postbus 20401, 2500 EK Den Haag
Bezuidenhoutseweg 73, 2594 AC Den Haag
070 379 8911

Meer informatie

www.rijksoverheid.nl/ezk
[@MinisterieEZK](https://twitter.com/MinisterieEZK)

Oktober 2022